

T. UES
1506
G172P
2001
F.2

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERIA Y ARQUITECTURA
ESCUELA DE INGENIERIA QUIMICA



**Propuesta para la enseñanza de Estadística
Aplicada a la Investigación en las carreras de
Ingeniería Química e Ingeniería de Alimentos
de la Universidad de El Salvador.**

PRESENTADO POR
MANUEL ANTONIO GALLO

PARA OPTAR AL TITULO DE
INGENIERO QUIMICO

CIUDAD UNIVERSITARIA, MAYO DE 2001

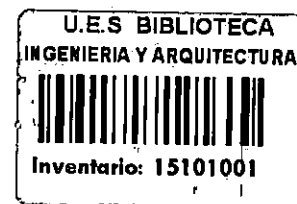
15/01001



15100001
5074

Recibido 16 de Mayo 2001

UNIVERSIDAD DE EL SALVADOR



RECTORA :
Dra. María Isabel Rodríguez

SECRETARIA GENERAL :
Licda. Lidia Margarita Muñoz Vela

FACULTAD DE INGENIERIA Y ARQUITECTURA

DECANO :
Ing. Alvaro Antonio Aguilar Orantes

SECRETARIO :
Ing. Saúl Alfonso Granados

ESCUELA DE INGENIERIA QUÍMICA

DIRECTORA :

Ing. Eugenia Salvadora Gamero Rodríguez



**UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERIA Y ARQUITECTURA
ESCUELA DE INGENIERIA QUIMICA**

**Trabajo de Graduación previo a la opción al grado de:
INGENIERO QUIMICO**

Título:

**Propuesta para la enseñanza de Estadística
Aplicada a la Investigación en las carreras de
Ingeniería Química e Ingeniería de Alimentos
de la Universidad de El Salvador.**

Presentado por:

MANUEL ANTONIO GALLO

Trabajo de Graduación aprobado por:

COORDINADORA :
Ing. Tania Torres Rivera

ASESOR :
Ing. Fernando Teodoro Ramírez Zelaya

San Salvador, mayo de 2001

Trabajo de Graduación aprobado por:

Coordinadora

Tania Torres

Ing. Tania Torres Rivera



Asesor

Fernando Teodoro Ramírez Zelaya

Ing. Fernando Teodoro Ramírez Zelaya

AGRADECIMIENTOS

Quiero agradecer a todas aquellas personas que de una manera u otra colaboraron para la realización de este trabajo.

Particularmente la colaboración especial de la coordinadora Ing. Tania Torres y el asesor Ing. Fernando Teodoro Ramírez. Además, la valiosa y apreciable colaboración del Ing. Edwin Antonio Alvarenga, Rene Eduardo Meléndez, Ing. Daniel Buenaventura Mejía y el Ing. Juan Rodolfo Ramírez; por su paciencia y tiempo, así como las orientaciones que me brindaron.

DEDICATORIA

A mi madre María Estebana.

Particularmente a mis tías Valentina; a mi padre, Manuel

A mis primos y primas

A mis hermanos: Juan Antonio y Daniel

A mis amigos: Juan, Edwin, Hector, Bayron, Rene, Gilberto, Don Vlady, Mego, Ulises, Emely y especialmente a mi coordinadora, Roberto y Juan José.

A Marvin (Q.D.D.G).

A todas aquellos (as) que hicieron de este más interesante este trabajo.

RESUMEN

La finalidad de este trabajo, es la elaboración de un documento que sirva de guía para la enseñanza y utilización de la estadística en las investigaciones que se realizan en las diversas áreas de la Ingeniería Química, Ingeniería de Alimentos y otras; tanto a nivel de pregrado como profesional.

El trabajo recopila una investigación bibliográfica, así como ejemplos de casos reales extraídos de trabajos de graduación realizados previamente por la Escuela de Ingeniería Química de la Universidad de El Salvador.

Las técnicas estadísticas consideradas en este documento y que son de mayor aplicación en las áreas afines a la Ingeniería Química e Ingeniería de Alimentos. Son las siguientes: *planteamientos y pruebas de hipótesis, inferencia estadística, diseños de experimentos, análisis de regresión y análisis de superficie de respuesta.*

A fin de desarrollar cada uno de los tópicos anteriormente mencionados el trabajo se estructuró de la siguiente manera: En el capítulo 1, se abarca las generalidades de la investigación y la importancia de estadística en la investigación; el capítulo 2, contiene la descripción de la variabilidad, considerando temas como: la descripción de la variabilidad por: medidas de tendencia central, medidas de dispersión, medidas de forma y por modelos matemáticos (distribución de probabilidad); el capítulo 3, trata sobre las muestras aleatorias y distribuciones de muestreo; el capítulo 4, trata sobre la inferencia estadística, conteniendo temas como: estimación de parámetros, pruebas de hipótesis simples y pruebas no paramétricas; el capítulo 5, trata sobre análisis de regresión; el capítulo 6, trata sobre diseños de experimentos, como: diseños unifactoriales y factoriales; y por último el capítulo 7, trata sobre análisis de superficies de respuesta.

ÍNDICE

CONTENIDO	Pág.
INTRODUCCIÓN	1
1. GENERALIDADES DE LA INVESTIGACIÓN Y LA IMPORTANCIA DE LA ESTADÍSTICA	3
1.1 HIPÓTESIS DE INVESTIGACIÓN	4
1.1.1 QUE SON LAS HIPÓTESIS.	4
1.1.2 IMPORTANCIA DEL PLANTEAMIENTO DE HIPÓTESIS	5
1.1.3 REQUISITOS DE LAS HIPÓTESIS	5
1.1.4 COMO SURGEN LAS HIPÓTESIS	6
1.1.5 CLASIFICACIÓN DE LAS HIPÓTESIS	7
1.1.5.1 Hipótesis de Investigación	7
1.1.5.2 Hipótesis Nulas	11
1.1.5.3 Hipótesis Estadísticas	11
1.2 DISEÑOS DE INVESTIGACIÓN	12
1.2.1 DISEÑOS EXPERIMENTALES DE INVESTIGACIÓN	13
1.2.1.1 Experimento	14
1.2.2 DISEÑOS NO EXPERIMENTALES DE INVESTIGACIÓN	16
1.2.2.1 Investigaciones transeccional o Transversal	18
1.2.2.2 Investigaciones longitudinales	19
1.2.3 CARACTERÍSTICAS DE LA INVESTIGACIÓN NO EXPERIMENTAL EN COMPARACIÓN CON LA INVESTIGACIÓN EXPERIMENTAL	20
1.3 EL MÉTODO CIENTÍFICO	22
1.3 IMPORTANCIA DE LA ESTADÍSTICA EN EL CAMPO DE LA INVESTIGACIÓN EN INGENIERÍA QUÍMICA E INGENIERÍA DE ALIMENTOS	23

2	DESCRIPCIÓN DE LA VARIABILIDAD	25
2.1	DESCRIPCIÓN DE UNA VARIABLE	25
2.2	DESCRIPCIÓN DE VARIABLE	26
2.3	DESCRIPCIÓN DE UN PARÁMERO	27
2.4	DESCRIPCIÓN DE UN ESTADÍSTICO	27
2.5	VARIABILIDAD EN LAS POBLACIONES Y CARACTERÍSTICAS ALEATORIAS	28
	2.5.1 DESCRIPCIÓN	28
2.6	DESCRIPCIÓN GRÁFICA DE LA VARIABILIDAD	28
2.7	DESCRIPCIÓN NUMÉRICA DE LA VARIABILIDAD	33
	2.7.1 MEDIDAS DESCRIPTIVAS O DE POSICIÓN	33
	2.7.2 MEDIDAS DE DISPERSIÓN	34
	2.7.3 MEDIDAS DE FORMA	36
2.8	DESCRIPCIÓN DE LA VARIABILIDAD POR MODELOS MATEMÁTICOS	39
	2.8.1 FUNCIÓN DE DISTRIBUCIÓN	40
	2.8.1.1. Valores Esperados, Media, Varianza y Desviación Típica	41
	2.8.1.2 Resumen	43
	2.8.2 FUNCIÓN DE DENSIDAD	43
	2.8.2.1 Valores Esperados, Media, Varianza y Desviación Típica	46
	2.8.2.2 Distribución Normal	46
	2.8.2.3 Distribución Normal Estándar	53
	2.8.2.4 Comprobación de Normalidad	54
	2.8.2.5 Aproximaciones Normales	54
	2.8.2.6 Transformaciones de Observaciones Hasta Casi la Normalidad	56
2.9	EJEMPLOS	56
2.10	EJERCICIOS PROPUESTOS	74

3	MUESTRAS ALEATORIA Y DISTRIBUCIÓN DE MUESTREO	81
3.1	MUESTRAS ALEATORIAS	82
3.2	ESTADÍSTICOS	83
3.3	DISTRIBUCIÓN DE MUESTREO	84
3.4	EJEMPLOS	90
3.5	EJERCICIOS PROPUESTOS	92
4	INFERENCIA ESTADÍSTICA	94
4.1	ESTIMACIÓN DE PARÁMETROS	94
	4.1.2 ESTIMACIÓN PUNTUAL	95
	4.1.2.1 Propiedades de los Estimadores	97
	4.1.2.2 Método de la Máxima Similitud.	101
	4.1.2 ESTIMACIÓN DE INTERVALO DE CONFIANZA	103
	4.1.2.1 Nivel de Confianza y Precisión de la Estimación	
	4.1.2.2 Elección del Tamaño de Muestra	105
	4.1.2.3 Intervalo de Confianza de un Lado	110
	4.1.2.4 Intervalos de Confianza Aproximados en la Estimación de la Máxima Similitud	111
	4.1.2.5 Intervalos de Confianza Simultáneas	111
4.2	PRUEBAS DE HIPÓTESIS	113
	4.2.1 HIPÓTESIS ESTADÍSTICA	113
	4.2.2 ERROR DE TIPO I Y II	115
	4.2.3 HIPÓTESIS UNILATERALES Y BILATERALES	116
	4.2.4 PRUEBA DE HIPÓTESIS EN MEDIA Y VARIANZA	118
4.3	PRUEBAS NO PARAMÉTRICAS	126
4.4	EJEMPLOS DEL CAPÍTULO	132
4.5	EJERCICIOS PROPUESTOS	154
5	ANÁLISIS DE REGRESIÓN	165
5.1	IMPORTANCIA DE LOS MODELOS DE REGRESIÓN	165
5.2	EN QUE CONSISTE UN ANÁLISIS DE REGRESIÓN	166
5.3	FASES DE UN ESTUDIO MEDIANTE MODELOS DE	

	REGRESIÓN	167
5.4	MODELOS DE REGRESIÓN	167
5.5	REGRESIÓN LINEAL SIMPLE	169
	5.5.1 TRANSFORMACIÓN A UNA LINEA RECTA	178
5.6	REGRESIÓN LINEAL MÚLTIPLE	180
	5.6.1 MODELOS DE REGRESIÓN MÚLTIPLE	180
	5.6.2 ESTIMACIÓN DE PARÁMETROS CON MODELOS DE REGRESIÓN	182
	5.6.3 ESTIMACIÓN GLOBAL DE AJUSTE	186
	5.6.4 INTERVALOS DE CONFIANZA EN REGRESIÓN LINEAL MÚLTIPLE.	191
	5.6.5 PREDICCIÓN DE NUEVAS OBSERVACIONES	192
	5.6.6 VALIDACIÓN POR ANÁLISIS DE RESIDUOS	192
5.7	REGRESIÓN POLINOMIAL	193
5.8	EJEMPLOS	194
5.9	EJERCICIOS PROPUESTOS	205
6	DISEÑOS DE EXPERIMENTOS	208
	6.1 APLICACIONES DE LOS DISEÑOS EXPERIMENTALES	209
	6.2 PRINCIPIOS BÁSICOS	210
	6.3 DISEÑO Y ANÁLISIS DE EXPERIMENTOS DE UN FACTOR	213
	6.3.1 EXPERIMENTOS DE UN FACTOR COMPLETAMENTE ALEATORIO (CASO DE EFECTOS FIJOS)	214
	6.3.1.1 Ventajas	215
	6.3.1.2 Modelos Estadístico Lineal	216
	6.3.1.3 Planteamiento de la Prueba de Hipótesis	216
	6.3.1.4 Tratamiento Estadístico	217
	6.3.1.5 Estimación de Parámetros	220
	6.3.1.6 Análisis de Residuos y Validación del Modelo	222
	6.3.1.7 Diseños Balanceado	224

6.3.1.8	Pruebas Estadísticas Para Igualdad de Varianza	225
6.3.2	MODELOS DE EFECTOS ALEATORIOS	226
6.3.2.1	Modelo Estadístico Lineal	227
6.3.2.2	Planteamiento de la Prueba de Hipótesis	227
6.3.2.3	Tratamiento Estadístico	228
6.3.2.4	Estimación de Parámetros	229
6.3.2.5	Selección de Tamaño de Muestra	230
6.3.2.5.1	<i>Curvas Características de Operación</i>	230
6.3.2.6	Método de Estimación de Intervalos de Confianza	232
6.3.3	DISEÑO DE BLOQUES ALEATORIOS	232
6.3.3.1	Ventajas de su Uso	233
6.3.3.2	Diseño Experimental	233
6.3.3.3	Modelo Estadístico Lineal	234
6.3.3.4	Planteamiento de la Prueba de Hipótesis	234
6.3.3.5	Tratamiento Estadístico	235
6.3.3.6	Pruebas Sobre de Tratamiento Individual	237
6.3.3.7	Análisis de Residuos y Validación del Modelo	237
6.3.4	ANÁLISIS DE COVARIANZA	239
6.3.4.1	Fundamento del Análisis	239
6.3.4.2	Finalidad del Análisis	240
6.3.4.3	Asunciones de la Covarianza	240
6.3.4.4	Método de Regresiones de Grupos	241
6.3.4.5	Diseño Unifactorial con una Covariable	242
6.3.4.5.1	<i>Planteamiento de la Prueba de Hipótesis</i>	244
6.3.4.5.2	<i>Análisis de Covarianza</i>	244
6.3.4.5.3	<i>Estadística de Prueba</i>	245
6.3.4.5.4	<i>Prueba de Idoneidad del Modelo</i>	246
6.3.4.5.5	<i>Determinación de Parámetros</i>	246
6.3.4.6	Otros Modelos de Covarianzas.	247

6.4	ANÁLISIS Y DISEÑOS DE EXPERIMENTOS CON VARIOS FACTORES	249
6.4.1	VENTAJAS	250
6.4.2	DESVENTAJAS	250
6.4.3	MODELOS PARA DISEÑOS FACTORIALES	251
6.4.3.1	Recolección de las Observaciones para los Diseños Factoriales	251
6.4.3.2	Análisis de Datos (Análisis de Varianza)	254
6.4.3.3	Estimación de Parámetros de los Modelos Factoriales	256
6.4.3.4	Elección de Tamaño de Muestra	258
6.4.3.5	Idoneidad del Modelo (Efectos Fijos)	261
6.4.4	DISEÑO FACTORIAL 2^k (cada factor con dos niveles)	261
6.4.4.1	Diseños 2^2	262
6.4.4.1.1	<i>Recolección de Datos</i>	263
6.4.4.1.2	<i>Análisis de Varianza</i>	264
6.4.4.1.3	<i>Análisis de Residuos y Idoneidad del Modelo</i>	265
6.4.4.2	Diseños 2^k Para Factores ≥ 3	265
6.4.4.2.1	<i>Recolección de Datos</i>	267
6.4.4.2.2	<i>Análisis Estadístico</i>	268
6.4.4.2.3	<i>Análisis de Varianza</i>	268
6.4.4.2.4	<i>Idoneidad del Modelo</i>	269
6.4.5	CONFUSIÓN EN EL DISEÑO 2^k	270
6.4.5.1	Construcción de Bloques	270
6.4.6	RÉPLICA FRACCIONAL DEL DISEÑO 2^k	273
6.4.6.1	Fracción Media del Diseño 2^k	273
6.4.6.2	Resolución del Diseño	276
6.4.6.3	Fracciones Menores	277
6.5	EJEMPLOS DEL CAPITULO	278
6.6	EJERCICIO PROPUESTOS	332

7	MÉTODOS Y DISEÑOS DE SUPERFICIES DE RESPUESTAS	344
7.1	MÉTODOS PARA DETERMINAR LA COMBINACIÓN ÓPTIMA DEL SISTEMA	346
7.2	MÉTODO DE LA MÁXIMA PENDIENTE	349
	7.2.1 DISEÑOS DE PRIMER ORDEN	350
	7.2.2 DISEÑOS DE SEGUNDO ORDEN	351
	7.2.2.1 Localización del Punto Estacionario	351
	7.2.2.2 Caracterización de la Superficie de Respuesta	353
7.3	SISTEMAS DE LOMAS	354
7.4	DISEÑOS EXPERIMENTALES PARA AJUSTAR SUPERFICIE DE RESPUESTA	356
	7.4.1 DISEÑOS PARA AJUSTAR EL MODELO DE PRIMER ORDEN	356
	7.4.2 DISEÑOS PARA AJUSTAR EL MODELO DE SEGUNDO ORDEN	357
7.5	FORMACIÓN DE BLOQUES EN DISEÑOS DE SUPERFICIE DE RESPUESTA	360
7.6	EJEMPLO DEL CAPÍTULO	364
7.7	EJERCICIOS PROPUESTOS	371
	CONCLUSIONES	373
	RECOMENDACIONES	375
	BIBLIOGRAFÍA	376
	ANEXOS	378
	ANEXO A	
	GLOSARIO DE TÉRMINOS	379
	ANEXO B	
	DIAGNÓSTICO DEL GRADO DE USO DE LA ESTADÍSTICA APLICADA A LA INVESTIGACIÓN	382
	ANEXO C	
	Distribución normal acumulativa estándar	393

ANEXO D		
Puntos porcentuales de la distribución X^2	395
ANEXO E		
Puntos porcentuales de la distribución t	397
ANEXO F		
Puntos porcentuales de la distribución F	398
ANEXO G		
Distribución acumulativa de Poisson	403
ANEXO H		
Distribución acumulativa Binomial	406
ANEXO I		
Números aleatorios	411
ANEXO J		
Valores críticos para la prueba Wilcoxon de dos muestras	412
ANEXO K		
Valores críticos para la prueba de signo	414
ANEXO L		
Valores críticos para la prueba Wilcoxon del rango con signo	415

ÍNDICE DE TABLAS

No.	TITULO DE TABLA	Pags.
1.1	Resumen para una elección de investigación experimental y no Experimental apropiada	22
2.1	Resumen para diferenciar la media aritmética entre una población infinita y una muestra	27
2.2	Descripción de las medidas de posición para poblaciones y Muestras	35
2.3	Descripción de la variabilidad a través de las medidas de dispersión para poblaciones y muestras	36
2.4	Resumen de distribuciones discretas	44
2.5	Resumen de las distribuciones de densidad de mayor aplicación en el área de Ingeniería	47
2.6	Algunas transformaciones útiles	56
2.7	Resistencia al rompimiento en libras por pulgadas cuadradas de 100 botellas de vidrio no retornables de refrescos de 1 litro	57
2.8	Distribución de frecuencia para los datos del ejemplo 2.1	58
2.9	Contenido en gramos de las bolsas de detergentes	62
2.10	Resumen de los datos del contenido en gramos de las bolsitas de detergentes	64
2.11	Comparación de los porcentajes de los datos de la serie estadística con los de una población normal	67
3.1	Descripción breve de las distribuciones de muestreo a partir de poblaciones normales (variables aleatorias continuas)	86
3.2	Descripción breve de las distribuciones de muestreo a partir de una función de distribución (variables aleatorias discretas)	89

4.1	Los parámetros que con frecuencia se necesita estimar con sus respectivos estimadores	96
4.2	Estimadores que maximiza las funciones de probabilidad normales y de las series de Bernoulli	102
4.3	Tipos de problemas para cálculos de intervalos de confianza de confianza para θ , como un estimador del parámetro θ	106
4.4	Decisiones en la pruebas de hipótesis	115
4.5	Ilustración del procedimiento de los problemas de prueba de hipótesis en medias y varianza	119
4.6	Resumen de procedimientos de pruebas de hipótesis en medias y varianzas para pruebas no paramétricas.	128
4.7	Resultados de prueba de resistencia a la presión en las tuberías.	134
4.8	Tiempo en segundos para la producción de jabón en forma paralela.	138
4.9	Datos del contenido de octanaje en diferentes mezclas de gasolinas.	150
4.10	Datos y rangos para el experimento de la prueba de la torsión.	153
5.1	Recolección de datos para el análisis de regresión.	169
5.2	Análisis de varianza para probar la significación de la regresión.	174
5.3	Análisis de varianza para la significación de la regresión múltiple.	189
6.1	Datos típicos para el análisis de varianza de clasificación en un sentido.	215
6.2	Análisis de varianza para el modelo de efectos fijos de clasificación en un sentido.	219
6.3	Análisis de varianza para el diseño de bloques completamente aleatorio.	236
6.4	Análisis de Covarianza como un análisis de varianza ajustado.	244
6.5	Análisis de covarianza para un experimento de un factor con una variable.	245

6.6	Análisis de covarianza para el diseño aleatorizado por bloques completos.	249
6.7	Modelos de experimentos factoriales.	252
6.8	Análisis de varianza del modelo bifactorial de efectos fijos y aleatorio.	255
6.9	Resumen de análisis de varianza para el diseño trifactorial.	256
6.10	Parámetros de la curva característica para el diagrama del anexo O. Modelo bifactorial de efectos fijos.	259
6.11	Parámetros de la curva característica para las tablas del anexo O y Q. Modelo bifactorial mixtos y aleatorio.	260
6.12	Signos para los efectos en el diseño 2^2 .	263
6.13	Resumen de las observación o recolección de datos para un diseño 2^3 .	264
6.14	Análisis de varianza para un modelo factorial 2^2 .	264
6.15	Recolección de datos para un diseño 2^3 .	267
6.16	Signos para efectos en un diseño 2^3 .	268
6.17	Análisis de varianza para un diseño 2^3 .	269
6.18	Signos positivos para el diseño factorial 2^3 .	274
7.1	Algunos diseños útiles factoriales 2^K para ajustar ecuaciones lineales.	350
7.2	Resumen de diseños compuestos centrales girables y de precisión uniforme.	360
7.3	Diseño de Box-Behnken de tres variables.	3.61

ÍNDICE DE FIGURAS

No.	TITULO DE FIGURA	Pags.
1.1	Metodología del proceso investigativo.	5
1.2	Relación entre variables en las hipótesis causales.	10
1.3	Relación entre conocimiento científico e investigación científica.	23
2.1	Histograma.	29
2.2	Polígono de frecuencia.	30
2.3	Distribución de frecuencia acumulada.	31
2.4	Diagrama de Box-Whisker.	32
2.5	Grafico de barras o rectángulos.	33
2.6	Ilustración de la distribución normal con sus respectivas asimetrías.	38
2.7	Comparación de datos normales con datos leptocúrtidos y planicúrtidos.	39
2.8	Distribución de probabilidad discreta.	41
2.9	Distribución de probabilidad continua.	43
2.10	Distribución normal estándar.	53
2.11	Representación grafica en papel probabilístico normal.	55
2.12	Histograma para el ejemplo 2.1.	59
2.13	Grafica de frecuencia acumulada vrs. intervalo de clases.	59
2.14	Diagrama de bigotes para el ejemplo 2.1.	61
2.15	Histograma de frecuencia para los datos del ejemplo 2.2.	64
2.16	Diagrama de polígono para los datos del ejemplo 2.2.	65
2.17	Grafica de probabilístico normal para los datos del ejemplo 2.2	66
4.1	Distribución de probabilidad de dos estimadores neutrales.	99
4.2	Estimador sesgado que tiene una varianza menor que un estimador neutral.	100
5.1	Uso de los modelos de regresión.	168

5.2	Las hipótesis $H_0: \beta_1=0$, no se rechaza.	174
5.3	La hipótesis $H_0: \beta_1=0$, se rechaza.	175
5.4	Patrones para las graficas de residuos.	176
6.1	Diseño de bloques completamente aleatorio.	234
6.2	Combinación de tratamiento en el diseño 2^2	262
6.3	Diagrama de combinación en un diseño trifactorial 2^3	266
6.4	Construcción de bloques para un diseño trifactorial.	271
6.5	Construcción de bloques para un diseño confundido 2^2	271
6.6	Construcción de bloques para un diseño 2^4	272
6.7	Construcción de cuatro bloques confundido 2^3	272
7.1	Superficie de respuesta tridimensional en la que se observa la variable respuesta.	347
7.2	Gráfica de contornos de una superficie de respuesta.	347
7.3	Superficie de respuesta de primer orden y trayectoria de máxima pendiente en ascenso.	349
7.4	Puntos estacionarios en una superficie de respuesta de segundo orden. ajustada.	352
7.5	Diseños compuestos centrales para $K=2$ y $K=3$	359
7.6	Diseño compuesto centrado en las caras $K=3$	361
7.7	Diseño de Box-Behnken para tres factores.	362

INTRODUCCIÓN

En la actualidad las complejas interrelaciones industriales del mundo globalizado, como por ejemplo: Industrias Químicas, de Alimentos, Agroindustrias, etc. Exigen que estas sean altamente competitivas para lograr subsistir en el medio. Es decir, desarrollar procesos eficientes que generen productos de alta calidad. Esto último implica que los profesionales encargados del análisis de un problema en particular deben ser capaces de identificar los factores que intervienen en el proceso a considerar. Lógicamente para poder abordar y resolver el problema se debe hacer una investigación de las posibles variables o factores involucrados, para identificar los efectos que éstos ocasionarán en el producto final. Con el objeto de conocer el proceso mediante la relación existente entre las variables y por ende optimizarlo. Para tal fin las empresas deben contar con personal calificado para lograr abordar y resolver problemas en forma planificada, segura y económica.

Esta investigación debe ser diseñada haciendo uso de ciertas herramientas en apoyo al método científico. Tales herramientas la proporciona la estadística aplicada a la investigación, con técnicas tales como: el diseño experimental, el análisis de regresión, análisis de superficie de respuesta, etc.

La estadística se han venido desarrollando en la medida en que las investigaciones más complejas han ido apareciendo. Desde que se comenzaron a utilizar en la Agricultura y Biología, su uso se ha generalizado en las Ciencias Físicas y Sociales; las áreas de Ingeniería y la industria en general.

Para que los resultados obtenidos en una investigación proporcionen conclusiones y toma de decisiones objetivas y confiables, está debe proceder de la siguiente manera: planteamiento del problema e identificación de las variables (plantear las hipótesis), planificar y evaluar los efectos de las variables (diseñar el experimento), analizar la interrelación entre las variables (análisis de regresión y

modelaje matemático) y optimizar los valores de las variables (análisis de superficie de respuesta).

El papel que la estadística aplicada ha desempeñado en los procesos de investigación y aprendizaje queda espectacularmente demostrado en el desarrollo de la revolución industrial, en cuanto a calidad y productividad.

Las diversas áreas de la Ingeniería Química e Ingeniería de Alimentos no son la excepción de este desarrollo; y por lo tanto no pueden dejar de verse involucradas en la realización de investigaciones científicas y tecnológicas, que demandan tanto de estudiantes como de profesionales, la necesidad de presentar resultados que se apeguen a la realidad del fenómeno investigado.

La Escuela de Ingeniería Química de la Universidad de El Salvador ante la necesidad de que los profesionales de Ingeniería Química y Alimentos posean las herramientas estadísticas necesarias para realizar investigaciones de una manera sistemática y objetiva, propone el presente trabajo de graduación con el fin de que sirva de guía tanto para la enseñanza de estas técnicas a estudiantes, así como a profesionales a la hora de realizar una investigación.

CAPITULO 1

GENERALIDADES DE LA INVESTIGACION Y LA IMPORTANCIA DE LA ESTADISTICA.

En las áreas de Ingeniería Química e Ingeniería de Alimentos se realizan investigaciones con mucha regularidad a nivel de pregrado, trabajos de graduación y en el ámbito profesional. Para realizarlas a cabalidad es necesario que el experimentador tenga adiestramiento, conocimiento y conciencia de la importancia de la formulación de hipótesis, diseños de experimentos, método científico y métodos estadísticos aplicados. No ignorando la aportación que da el conocimiento y fundamento sobre el tema investigado.

Para que en una investigación se establezcan relaciones entre hechos y explicar por que se producen es necesario que contenga planteamientos de hipótesis sobre el fenómeno de interés. Para que sea planificada correctamente, es necesario que se realice el diseño de la investigación. Según las características que se presenten en las variables independientes o causas, podemos asegurarnos si los diseños a realizarse son experimentales o no experimentales.

Cuando se realiza una investigación se sigue una secuencia de pasos lógicos de hechos observados, y para volverla objetiva y cristalina, es necesario cumplir con cada una de las etapas de la metodología científica.

Finalmente para que las investigaciones se realicen en forma exitosa, ordenada, y planificada deben ser sustentadas por aplicación correcta de técnicas estadísticas, con lo que se logra disminuir el error experimental y la desviación de la investigación de la realidad. A continuación se procede a profundizar en este tema.

1.1 HIPÓTESIS DE INVESTIGACIÓN.

1.1.1 QUE SON LAS HIPÓTESIS.

La hipótesis es una proposición que nos permite establecer relaciones entre hechos. Su valor reside en la capacidad para establecer dichas relaciones y explicar por que se producen.

Las hipótesis indican lo que estamos buscando o tratando de probar y pueden definirse como explicaciones tentativas del fenómeno investigado, formuladas a manera de proposiciones.

Las hipótesis no necesariamente son verdaderas; pueden o no serlo, pueden o no comprobarse con hechos. Son explicaciones tentativas, no los hechos en sí. Al formularlas, el investigador no puede asegurar que vayan a comprobarse.

Dentro de la investigación científica, las hipótesis son proposiciones tentativas acerca de las relaciones entre dos o más variables y se apoyan en conocimientos organizados y sistematizados.

Las hipótesis pueden ser más o menos generales o precisas, e involucrar dos o más variables, pero en cualquier caso son proposiciones sujetas a comprobación empírica y a verificación real por medio de resultados. Es decir, comprobar la teoría de la teoría de entrada o sobre la cual se han formulado las hipótesis y con esto, llegar a la teoría final o un nuevo conocimiento, producto de la investigación. En la Figura 1.1 se muestra en forma esquematizado toda la metodología del proceso investigativo.

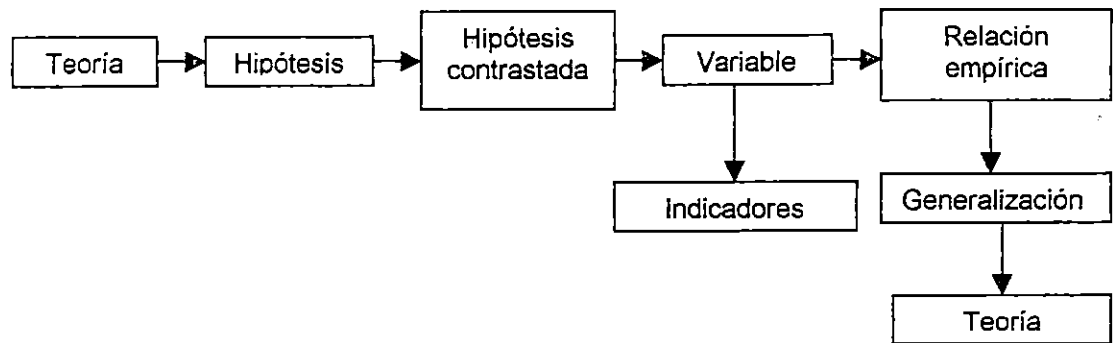


Figura 1.1. Metodología del proceso investigativo.

1.1.2 IMPORTANCIA DEL PLANTEAMIENTO DE HIPÓTESIS

La importancia de las hipótesis se deriva del nexo entre la teoría y la realidad empírica del sistema formalizado y la investigación. Son instrumentos de trabajo de la teoría y de la investigación en cuanto que introducen coordinación en el análisis y orientan a la elección de los datos; en este aspecto puede afirmarse que la hipótesis contribuye al desarrollo de la ciencia, así como la labor investigativa.

La hipótesis sirve para orientar y delimitar una investigación, dándole una dirección definida a la búsqueda de la solución de un problema.

1.1.3 REQUISITOS DE LAS HIPÓTESIS

Pueden formularse hipótesis a diestra y siniestra, siempre y cuando sirvan de cimientos a la investigación, de lo contrario se corre el peligro de que la investigación se derrumbe.

Por tal razón las hipótesis deben:

- Referirse a una situación real

- Establecer las variables a estudiar, es decir, especificar las variables y fijarles límites; de manera que sean comprensibles, precisas y lo más concretas posible.
- Establecer relaciones entre variables, las cuales deben ser claras y verosímiles (lógicas); es decir, la hipótesis debe ser específica de tal manera que sirva de base a inferencias que nos ayuden a decidir si explica o no los fenómenos observados. Se requiere que las hipótesis establezcan relaciones cuantitativas entre variables.
- Mantener la consistencia entre hechos e hipótesis, en la cual los términos de la hipótesis y la relaciones planteadas entre ellos deben ser observables y medibles; ya que estos se cimientan, al menos en parte, sobre hechos ya conocidos en el campo de estudio. Por lo tanto, las hipótesis no deben establecer implicaciones contradictorias o inconsistentes con lo ya verificado en forma objetiva. Debe establecerse un todo armónico entre hipótesis y hechos.
- Las hipótesis deben estar relacionadas con técnicas disponibles para probarlas. Este requisito está estrechamente relacionado con el anterior y se refiere a que al formular una hipótesis, tenemos que analizar si existen técnicas o herramientas de investigación para poder verificarla.

1.1.4 COMO SURGEN LAS HIPÓTESIS

Es natural que las hipótesis surjan del planteamiento del problema; es decir, provienen de la revisión misma de la teoría. Nuestras hipótesis pueden surgir del postulado de una teoría, del análisis de ésta, de generalizaciones empíricas pertinentes a nuestro problema de investigación y de estudios revisados o antecedentes consultados.

Existe pues, una relación muy estrecha entre el planteamiento del problema, la revisión de la literatura (teoría) y las hipótesis. La revisión inicial de la literatura

hecha para familiarizarnos con el problema de estudio nos lleva a plantearla, después revisamos la literatura y afinamos o precisamos el planteamiento, del cual derivamos la hipótesis. Al formular las hipótesis volvemos a evaluar nuestro planteamiento del problema.

1.1.5 CLASIFICACION DE LAS HIPÓTESIS

Existen diversas formas de clasificar las hipótesis, para nuestro caso se van a clasificar de la manera siguiente:

- a. *hipótesis de investigación,*
- b. *hipótesis nulas,*
- c. *hipótesis alternativas. e*
- d. *hipótesis estadísticas.*

En este trabajo se dará mayor importancia de estudio a las hipótesis de investigación, hipótesis nulas e hipótesis estadísticas.

1.1.5.1 Hipótesis de Investigación

Son proposiciones tentativas acerca de las posibles relaciones entre dos o más variables y que cumplan con los requisitos mencionados en 1.1.3. Se les suele simbolizar como H_i o H_1 , H_2 , H_3 , etc. (si son varias hipótesis) y también se les denomina hipótesis de trabajo. A su vez, las hipótesis de investigación pueden ser:

- a. *Hipótesis descriptivas del valor de las variables que se va observar en un contexto o en la manifestación de otra variable.*

Las hipótesis de este tipo se utilizan a veces en estudios descriptivos. Pero cabe mencionar que no en todas las investigaciones descriptivas se formulan hipótesis o que estas son afirmaciones generales. Algunos investigadores consideran estas hipótesis como afirmaciones *univariadas*, argumentando

que no relacionan variables. Más que relacionar variables, se esta planteando como se va a manifestar una variable en una *constante*.

Ejemplo 1.1

Hi: *En el invierno la humedad relativa presente en el aire es alta, razón por la cual los secadores de charolas por lotes son menos eficiente que en el verano, donde la humedad relativa disminuye.*

b. *Hipótesis correlacionales*

Especifican las relaciones entre dos o más variables. Corresponden a los estudios de correlaciones y pueden establecer la asociación entre dos variables. Por ejemplo: *"la eficiencia del secador de charolas por lotes está relacionada con el gradiente de humedad entre el aire de entrada y la sustancia que se pretende secar"*.

Sin embargo, las hipótesis correlacionales no sólo pueden establecer que dos o más variables se encuentran asociadas, sino cómo están asociadas. Alcanzan el nivel predictivo y parcialmente explicativo.

Ejemplo 1.2

Hi: *A mayor humedad relativa en el aire, mayor será el tiempo de secado por lotes en un secador de charolas; puesto que, a mayor cantidad de agua en el aire de entrada será menor el gradiente de humedad con respecto a la sustancia que se desea secar.*

En esta caso las variables están interrelacionadas de modo que ambas variables son independiente y dependiente a la vez. No importando donde se coloque las variables, si como independiente o dependiente, el resultados será el mismo.

Por otra parte, es común que cuando en la investigación se pretende correlacionar varias variables se tengan diversas hipótesis, y cada una de ellas relacione un par de variables.

Ejemplo 1.3

H₁: A mayor humedad en el aire, mayor calor (energía) en el medio.

H₂: A mayor humedad en el aire, mayor cantidad de agua evaporada del medio.

H₃: A mayor agua evaporada, mayor movimiento convectivo.

H₄: Mayor movimiento convectivo, disminución de la presión.

H₅: Menor presión, mayor humedad relativa.

c. Hipótesis de la diferencia entre grupos

Estas hipótesis se formulan en investigaciones cuyo fin es hacer una comparación, como se muestra en el ejemplo 1.4.

Ejemplo 1.4

La Saccharomyces cerevisiae produce mayor cantidad de alcohol que la Saccharomyces uvarum, considerando las mismas condiciones, como: cantidad de nutrientes/bacterias, sustrato, temperatura, presión atmosférica y tiempo de fermentación.

d. Hipótesis que establecen relaciones de causalidad

Este tipo de hipótesis no solamente afirman las relaciones entre dos o más variables y como se dan dichas relaciones, sino que además proponen un *sentido de entendimiento* de ellas. Este sentido puede ser más o menos completo dependiendo del número de variables que se incluyan, pero todas estas hipótesis establecen relaciones de causa-efecto. Para poder establecer causalidad antes debe haberse demostrado correlación, pero además la causa debe ocurrir antes que el efecto.

Las hipótesis correlacionales pueden simbolizarse de la siguiente manera:

$$X - Y$$

Las hipótesis causales pueden simbolizarse como se muestra en la Figura 1.2.

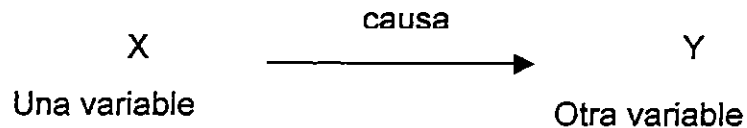


Figura. 1.2. Relación entre variables en las hipótesis casuales.

Los cambios en la causa deben provocar cambios en el efecto, como se muestra en el ejemplo 1.5.

Ejemplo 1.5

En un reactor de mezcla completa ideal, el tiempo espacial en que los reactantes estén en contacto en el reactor (CAUSA) definirá el grado de conversión (EFECTO).

Las hipótesis causales pueden ser:

- a. *Hipótesis causales bivariadas.* En estas hipótesis se plantea una relación entre una variable independiente y una variable dependiente.
- b. *Hipótesis causales multivariadas.* Plantean una relación entre varias variables independientes y una variable dependiente, o una variable independiente y varias variables dependientes, o varias variables independientes y varias variables dependientes.

Las hipótesis multivariadas pueden plantear otro tipo de relaciones causales, en donde ciertas variables intervienen modificando la relación (*hipótesis con presencia de variables intervinientes*).

Al hablar de hipótesis, a las supuestas causas se les conoce como *variables independientes* y a los efectos como *variables dependientes*. Solamente se puede hablar de variables independientes y dependientes cuando se formulan hipótesis causales o hipótesis de la diferencia de grupos, siempre y cuando en estas últimas se explique cuál es la causa de la diferencia hipotizada.

1.1.5.2 Hipótesis Nulas

Las hipótesis nulas, en cierto modo son el inverso de las hipótesis de investigación. También constituyen proposiciones acerca de la relación entre variables; solo que sirven para refutar o negar lo que afirma la hipótesis de investigación. Las hipótesis nulas se simbolizan por medio de: H_0

1.1.5.3 Hipótesis Estadísticas

Las hipótesis estadísticas son la transformación de las hipótesis de investigación y alternativas en símbolos estadísticos. Se pueden formular solo cuando los datos del estudio que se van a recolectar y analizar para probar o rechazar las hipótesis que son cuantitativas (números, porcentajes, promedios). Es decir, el investigador traduce su hipótesis de investigación y sus hipótesis nulas (y cuando se formulan hipótesis alternativas, también estas) en términos estadísticos. Básicamente hay tres tipos de hipótesis estadísticas que corresponden a las clasificaciones de las hipótesis de investigación y nula: *de estimación, de correlación y de diferencias de medias*.

a. *Hipótesis estadística de estimación.*

Corresponden a las que, el hablar de hipótesis de investigación, fueron denominadas "hipótesis descriptivas de una variable que se va a observar en un contexto". Sirven para evaluar la suposición de un investigador respecto al valor de alguna característica de una muestra dentro de la población o fenómeno de estudio.

b. *Hipótesis estadística de correlación.*

Estas hipótesis tienen por objeto traducir en términos estadísticos una correlación entre dos o más variables. El símbolo de una correlación entre dos variables es " r " (minúscula) y entre más de dos variables " R " (mayúscula).

c. *Hipótesis estadística de la diferencia de medias u otros valores.*

En estas hipótesis se compara una estadística entre dos grupos.

1.2 DISEÑO DE INVESTIGACIÓN

Es la planificación correcta de un experimento con la finalidad de obtener la máxima información a través de datos tomados del mismo y que permitan un análisis objetivo que conduzca a deducciones válidas con respecto al problema planteado.

El término *diseño* se refiere al plan o estrategia concebida para responder a las preguntas de la investigación. El diseño señala al investigador lo que debe hacer para alcanzar sus objetivos de estudio, contestar las interrogantes que se ha planteado y analizar la certeza de las hipótesis formuladas; tal como se ilustra el ejemplo 1.6.

Ejemplo 1.6

Si la pregunta de investigación fuera: ¿El tiempo de secado depende de la humedad relativa: por qué si y por qué no? y la hipótesis: "A mayor contenido de humedad en el aire, mayor tiempo de secado".

Si el diseño esta bien planteado, el producto final del estudio (sus resultados) tendrá mayores posibilidades de ser válido. No es lo mismo seleccionar un tipo de diseño que otro: cada uno tiene sus características propias. La precisión de la información obtenida puede variar según el diseño o la estrategia seguida.

Los diseños de investigación se clasifican en:

- a. *Investigación Experimental.*
- b. *Investigación no Experimental.*

Los diseños experimentales de investigación se dividen en:

- a. *Preexperimentos,*
- b. *Experimentos "puros" (verdaderos), y*
- c. *Cuasiexperimentos.*

Los diseños no experimentales de investigación se dividen en:

- a. *Transeccionales o transversales, y*
- b. *Longitudinales.*

1.2.1 DISEÑOS EXPERIMENTALES DE INVESTIGACIÓN.

Son los que se realizan cuando el experimentador requiere de la manipulación intencional o deliberada de las variables independientes (causas) para analizar sus posibles efectos (como son afectadas las variables dependientes).

1.2.1.1 Experimento

Es un proceso que permite comprobar o denegar mediante la práctica algún hecho. Este hecho constituye la hipótesis que el experimentador desea probar. Se caracteriza porque a través de él se observa únicamente los efectos y es de aplicación inmediata, ya sea para el experimentador o a la población a quien vayan dirigidos los resultados.

El término "*experimento*" puede tener al menos dos acepciones, una general y otra particular. La general se refiere a "tomar una acción" y después observar las consecuencias. Este uso del término es bastante coloquial; así hablamos de "experimentar" cuando mezclamos sustancias químicas y vemos la reacción de este hecho. La esencia de este concepto de "experimento" es que requiere la manipulación intencional de una acción para analizar sus posibles efectos.

La acepción particular, más armónica con un sentido científico del término, se refiere a "un estudio de investigación en el que se manipulan deliberadamente una o más variables independientes (supuestas causas) para analizar las consecuencias que la manipulación tiene sobre una o más variables dependientes (supuestos efectos), dentro de una situación de control para el investigador".

Si toma la acepción general del término "experimento", los preexperimentos, los experimentos "verdaderos" y los cuasiexperimentos podrían considerarse experimentos, ya que toman una acción y miden sus efectos. En cambio, si tomamos la segunda acepción (que hemos llamado "particular"), solo los experimentos "puros" serían "experimentos" y ambos conceptos se considerarían equiparables.

Los experimentos pueden ser:

- a. *Determinísticos, y*

b. *Aleatorios.*

Experimentos determinísticos: Es aquel cuyo resultado no está sujeto a variaciones no controlables por el experimentador, ejemplo: Experimentos en el área de matemáticas, físicas y química.

Experimentos aleatorios: Es aquel cuyos resultados si están sujetos a variaciones no controladas, ejemplo: Los experimentos biológicos. Los experimentos aleatorios se clasifican según el número de factores que se estudien simultáneamente, así se tienen:

- a. *Experimentos simples o de factor único, y*
- b. *Experimentos complejos o experimentos factoriales.*

Experimentos simples: Son aquellos en los cuales la estructura de entrada está compuesta por un solo factor, el cual puede ser:

- a. *Cualitativo, y*
- b. *Cuantitativo.*

Experimentos complejos: Los experimentos complejos o factoriales son utilizados para estudiar simultáneamente dos o más factores. En este tipo de experimentos, los tratamientos están constituidos por las combinaciones posibles de los niveles o modalidades de los factores en estudio. Cuando el número de factores en estudio es de dos se llama experimento bifactorial; cuando son tres, trifactorial; y cuando son cuatro o más, multifactorial. Por ejemplo, al comparar los métodos de extracción de ácido acético, por: extracción líquido – líquido o por destilación fraccionada, las variables son: los métodos de extracción, la concentración final, la calidad final, la capacidad de proceso y el costo de producción por unidad fijada.

1.2.2 DISEÑOS NO EXPERIMENTALES DE INVESTIGACIÓN.

Es la investigación que se realiza sin manipular deliberadamente una o más variables independientes. Lo que hace en la investigación no experimental es observar el fenómeno tal y como se dan en su contexto natural, para después analizarlo. La investigación no experimental o ex post-facto es cualquier investigación en la que resulta imposible manipular variables o asignar aleatoriamente el material experimental o las condiciones.

Las características que debe tener una investigación no experimental, son:

1. *No existe manipulación deliberada de las variables.* Se trata de investigaciones donde no se hace variar intencionalmente o deliberadamente las variables independientes (supuestas causas). *No hay asignación al azar del material experimental:* en el estudio no experimental el material de experimentación o factor ya pertenecía a un bloque o grupo determinado de la variable independiente por autoselección.
2. *No existe construcción alguna de la unidad experimental situación:* en la investigación no experimental se observaran situaciones ya existentes, no provocadas intencionalmente por el investigador
1. *Es una investigación sistemática y empírica:* Dado que las variables independientes no se manipulan porque ya han sucedido (solo se observan las consecuencias de las posibles causas). Las influencias sobre las relaciones entre variables se realizan sin intervención o influencia directa y dichas relaciones se observan tal y como se han dado en su contexto natural.

Ejemplo 1.7

"La contaminación del agua generada por la cantidad de aceite quemado producto de la actividad automotriz en la ciudad de San Salvador", dependerá de la cantidad de automotores, del mantenimiento de las unidades, del tratamiento y reciclaje del aceite, y de la disposición final del mismo. Pero el investigador que pretende evaluar la contaminación no tiene a su alcance manejar todas estas variables; por la cual no tiene control de las variables causales de la contaminación y se limita, nada más, a medir los efectos que estas producen en el agua.

Un ejemplo científico para ilustrar la diferencia entre un experimento y un no experimento serian los siguientes casos:

- | | |
|-------------------|---|
| Experimento 1 | Exponer una caja de cartón a la humedad, utilizando la cámara de humedad (condiciones controladas), para luego ver la influencia en el cartón. |
| No experimental 1 | Exponer la caja a la humedad ambiental o intemperie (condiciones no controladas). |
| Experimento 2 | En los procesos de fermentación industrial se controlan variables como: azúcares fermentables, concentración de inóculo, madurez del inóculo, pH, temperatura y nutrientes. |
| No experimental 2 | En forma artesanal se obtiene el producto, sin manipular ninguna de los factores que pueden afectar el proceso de fermentación. . |
| Experimento 3 | El secado de café por medio de secadores continuos de reflujo con reciclo, por medio de aire caliente. |
| No experimental 3 | Secarlos en patios a exposición a luz solar, que dependerá de las condiciones ambientales del momento. |

Los diseños no experimentales se clasifican por su dimensión temporal o el número de momentos o puntos en el tiempo en los cuales se recolectan datos, en:

- a. *Transversal o transeccional.* Las investigaciones se centra en analizar cuál es el nivel o estado de una o diversas variables en un momento dado, o bien en cuál es la relación entre un conjunto de variables en un punto en el tiempo.
- b. *Longitudinal.* Las investigaciones se centran en estudiar cómo evolucionan o cambian una o más variables, o las relaciones entre ellas.

1.2.2.1 Investigación Transeccional o Transversal

Los diseños de investigación transeccional o transversal recolectan datos en un solo momento (puntuales), en un tiempo único. Su propósito es describir variables, y analizar su incidencia e interrelación en un momento dado. Es como tomar una fotografía de algo que sucede. Pueden abarcar varios objetos o indicadores. Por ejemplo, estudiar el flujo de agua, demanda bioquímica de oxígeno (DBO), cantidad de sólidos solubles, y sólidos totales disueltos; etc., a las 12.00 M en un día cualquiera, asumiendo que es hora pico. Con la primicia de que es la hora de mayor gasto de agua en las diferentes actividades de una ciudad y con el propósito de calcular todas la variables mencionadas, y tomar en cuenta el diseño de una planta de tratamiento en un futuro inmediato, considerando todas las previsiones de seguridad, como lo es el tanque de distribución (función de manejar el flujo requerido o alimentar a la planta de tratamiento de agua).

A su vez, los diseños transeccionales pueden dividirse en dos tipos:

- a. *Diseños transeccionales descriptivos:* Tienen por objeto indagar la incidencia y los valores en que se manifiestan una o más variables, y proporcionar su descripción en un tiempo único.

- b. *Diseños transeccionales correlacionales*: Estos diseños describen relaciones entre dos o más variables. Se trata también de descripciones, pero no de variables individuales sino de sus relaciones, sean éstas puramente correlacionales o relaciones causales. En estos diseños lo que se mide es la relación entre variables en un tiempo determinado. Por lo tanto, los diseños correlacionales/causales pueden limitarse a establecer relaciones entre variables sin precisar sentido de causalidad, o pueden pretender analizar relaciones de causalidad.

1 2.2.2 Investigación longitudinal

En ocasiones el interés del investigador es analizar cambios a través del tiempo en determinadas variables o en las relaciones entre éstas, para cual se recolectan datos a través del tiempo en puntos o periodos especificados para hacer inferencias respecto al cambio, sus determinantes y consecuencias. Siguiendo el ejemplo anterior, un investigador busca analizar la cantidad (flujo) de agua, la demanda bioquímica de oxígeno (DBO), la demanda química de oxígeno (DQO), los sólidos totales, y los sólidos disueltos de las aguas negras de San Salvador todos los días a las 12.0 m (porque se tomarán muestra a través de un año (asumiendo esta como hora pico). En el ejemplo 1.7 se tenía una medida puntual. Para este caso es una medida puntual pero más sistematizada, ya que los flujo no se pueden suponer constantes dado que se tienen dos épocas en el año (seca e invierno), además las actividades en la semana son diferentes (días de laborales y fines de semana). Esta investigación longitudinal se realiza de manera descrita anteriormente con el objetivo de obtener datos más apegados a la realidad y con ello diseñar con mayor éxito, seguridad y confiabilidad el tanque de distribución o retención que alimentará el flujo de aguas residuales a la planta de tratamiento; y con lo cual se minimiza la contaminación del río y sus efluentes. A su vez los

diseños longitudinales suelen dividirse en tres tipos, estos son:

- a. *Diseños longitudinales de tendencia.* Los diseños de tendencia o trend son aquellos que analizan cambios a través del tiempo (en variables o sus relaciones), dentro de alguna población de interés en general. Su característica distintiva es que la atención se centra en una población de interés, como por ejemplo: *En un ingenio azucarero, como evoluciona en el tiempo de zafra el rendimiento de toneladas de caña molida/libras de azúcar producida.*
- b. *Diseños longitudinales de evolución de grupos.* Examinan cambios en subpoblaciones o grupos específicos, como por ejemplo: *La investigación sobre la contaminación del río Aseguashte de la ciudad de San Salvador, como contaminación por basura común, orgánica (biodegradable o no biodegradable), líquidos químicos, ferrosos y sólidos en general, observar como evoluciona en el transcurso de cinco años, tomando muestras al azar cada tres meses en puntos diferentes del río.*
- c. *Diseños longitudinales panel.* Los diseños panel son similares a las dos clases de diseños vistos anteriormente, sólo que la misma unidad de elementos es medido en todos los tiempos o momentos, como por ejemplo: *La investigación anterior descrita con la variante que se tomarán muestras al azar cada tres meses en los mismos puntos del río.*

1.2.3 CARACTERÍSTICAS DE LA INVESTIGACIÓN NO EXPERIMENTAL EN COMPARACIÓN CON LA INVESTIGACIÓN EXPERIMENTAL

Tanto la investigación experimental como la no experimental son herramientas muy valiosas de que dispone la ciencia y ningún tipo es mejor que el otro. El

diseño a seleccionar en una investigación depende más bien del problema a resolver y el contexto que rodea al estudio. Desde luego, ambos tipos de investigación poseen características propias que es necesario resaltar.

En lo que respecta a la posibilidad de replica, los diseños experimentales se pueden replicar más fácilmente, con o sin variaciones. Pueden replicarse en cualquier lugar siguiendo el mismo procedimiento.

En los experimentos (sobre todo en los laboratorios) las variables independientes pocas veces tienen tanta fuerza como en la realidad. En cambio, en la investigación no experimental estamos más cerca de las variables hipotetizadas como "reales" y consecuentemente tenemos mayor validez externa.

En resumen, ambas clases de investigación: experimental y no experimental, se utilizan para el avance del conocimiento y en ocasiones resulta más apropiado un tipo u otro dependiendo del problema de investigación a que nos enfrentamos. A continuación un Tabla 1.1. se presenta un resumen de las características para realizar una elección apropiada entre ambas tipos de investigaciones.

Tabla 1.1. Resumen para una elección de investigación experimental y no experimental apropiada.

ESTUDIO	HIPOTESIS	DISEÑO
Exploratorio	- No se establecen, lo que se puede formular son conjeturas iniciales.	- Transccional descriptivo - Preexperimental.
Descriptivo	- Descriptiva	- Transeccional descriptivo - Preexpciimental
Correlacional	-Diferencias de grupos sin atribuir causalidad	- Cuasiexperimental – - Transeccional correlacional - Longitudinal (no expelimental)
	- Correlacional	- Cuasiexperimental. - Transeccional correlaciona. - Longitudinal (no experimental)
Explicativo	- Diferencia de grupos atribuyendo	- Experimental - Cuasiexperimental, longitudinal y transeccional causal (cuando hay bases para inferir causalidad, un mínimo de control y análisis estadísticos apropiados para analizar relaciones causales)
	- Causales	• Experimental • Cuasiexperimental, longitudinal y transeccional causal (cuando hay bases para inferir causalidad, un mínimo de control y análisis estadísticos apropiados relaciones causales)

1.3 EL METODO CIENTÍFICO

La ciencia y la epistemología nos ponen de manifiesto el método científico, hasta no poder hablar de investigación sin tener que hablar de método científico. Podemos decir que entre la investigación científica y el conocimiento científico encontramos el método científico, que es el que nos asegura el primero y el segundo.

El método científico es un procedimiento para descubrir las condiciones en que se presentan sucesos específicos, caracterizados generalmente por ser tentativo, verificable, de razonamiento riguroso y observación empírica. Tal como se muestra en la Figura 1.3.

El método científico no es otra cosa que la aplicación de la lógica a las realidades o hechos observados. Nos lleva a eliminar el plano subjetivo en la interpretación de la realidad, permitiéndonos la objetividad del proceso investigativo.

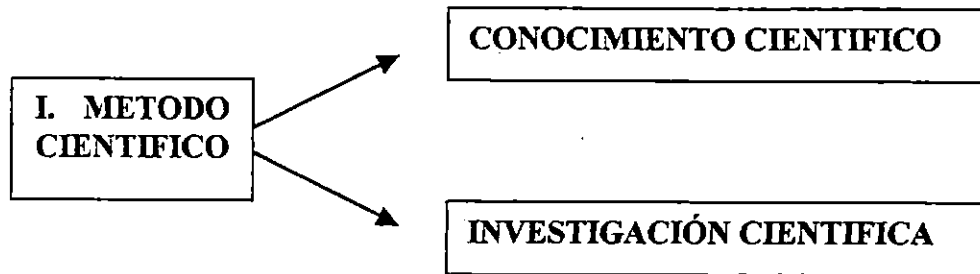


Figura 1.3. Relación entre el conocimiento científico e investigación científica.

Little y Jackson establecen que el experimento es un instrumento de investigación utilizado para descubrir algo desconocido o para probar un principio o una hipótesis; es un paso importante del método científico y las preguntas que este aspira a contestar serán fundamentales para el apoyo o rechazo de una hipótesis. De manera que toda investigación científica debe llevarnos a la búsqueda sistemática de algo nuevo.

La característica esencial del método científico es que pueden formularse hipótesis y comprobarse mediante una investigación, partiendo de una observación crítica.

1.5 IMPORTANCIA DE LA ESTADISTICA EN EL CAMPO DE LA INVESTIGACIÓN EN INGENIERIA QUÍMICA E INGENIERIA DE ALIMENTOS

La importancia de la estadística aplicada en las áreas de Ingeniería Química e Ingeniería de Alimentos es que puede incrementar grandemente la eficiencia de

las investigaciones y, a menudo reforzar las conclusiones obtenidas. En toda investigación es necesario aplicar un conjunto de conocimientos y disciplinas del tipo fisicoquímico aplicado con el fin de encontrar respuestas correctas a un problema específico.

La estadística aplicada puede ser considerarse un arte por permitir el uso de un conjunto de artificios necesarios para la recolección de información, y las cuales son ajustados en lo posible a la realidad, para permitir de esa forma, reducir al máximo todas las fuentes de error, se considera como una ciencia por la aplicación del método científico, las matemáticas, principios y leyes de las probabilidades.

El propósito de los métodos estadísticos aplicados a las investigaciones en los procesos químicos, agroindustriales y alimenticios, es hacerlas lo más objetivas, confiables y eficientes posible. Aliviando tres fuentes de dificultades mas generales a lo que se enfrenta el investigador, estas son:

1. *Error experimental (ruido)*: La variación producida por factores distorcionantes, tanto conocido como desconocidos.
2. *Confusión entre correlación y causalidad*: Diferenciar las relaciones entre variables (correlación) y la causa que ocasiona las variaciones en el fenómeno.
3. *Complejidad de los efectos*: La capacidad de analizar simultáneamente los efectos ocurridos en la investigación por varias causas.

Aliviando con estas tres fuentes de dificultades, darán resultados más objetivo, confiables, seguros, eficientes y serán más apegados a la realidad; además de garantizar la calidad la validez interna y externa de las investigaciones en las áreas de Ingeniería Química e Ingeniería de Alimentos.

CAPITULO 2

DESCRIPCION DE LA VARIABILIDAD

2.1 DESCRIPCION DE UNA POBLACION

Una Población: es un conjunto completo de mediciones ~~o~~ observaciones de interés que son objeto de estudio y sobre las que se desea obtener ciertas conclusiones. Generalmente las poblaciones se pueden clasificar según:

a. *El tipo de elementos de la población:*

a.1 *Población con existencia física real:* en estas poblaciones los elementos tienen existencia física real previa a la realización del estudio. Como ejemplo de estas poblaciones se tienen: se realiza un análisis para determinar la cantidad de sacarosa contenida en la caña de azúcar en un camión para la producción de azúcar, la población la constituye la caña contenida en el camión o en los camiones.

a.2 *Poblaciones abstractas:* en estas poblaciones los elementos de la misma se determina mediante la realización de un determinado proceso que se denomina experimento aleatorio (prueba o corrida experimental). Como ejemplo de estas poblaciones se tienen: en el práctica laboratorio de transferencia de masa para evaluar el coeficiente de difusión de líquidos y gases, se cuenta con aparatos para realizar las evaluaciones del coeficiente de difusión. Entonces, la población la constituyen todas las mediciones posibles hechas para evaluar el coeficiente de difusión de un solvente líquido en un medio gaseoso.

b *El número de elementos de la población:*

b.1 *Poblaciones infinitas:* son las que comprenden un número infinitamente grande de unidades elementales. Como ejemplo puede mencionarse: la cantidad de levadura residual (Scharomyces Cereviciea) que interviene en el proceso de cerveza en una industria cervecera.

b.2 *Poblaciones finitas:* es aquella que tiene un número limitado o finito de unidades elementales. Como ejemplo puede mencionarse: el número de mediciones de temperatura y humedad en un proceso industrial de secado.

2.2 DESCRIPCION DE VARIABLE

Una variable se caracteriza por el hecho que puede tomar diferentes valores. Generalmente se simbolizan con las ultimas letras del alfabeto X, Y, Z, ect. Si la variable toma un solo valor, se llama constante. Las variables se clasifican en:

- *Variable continua:* es una variable que teóricamente puede tomar cualquier valor entre dos valores enteros. Como por ejemplo: el cambio de presión, temperatura en una reacción química en estado gaseosa en un reactor químico.
- *Variable discretas o atributos:* es una variable que toma solamente valores específicos. Como por ejemplo: la cantidad de cajas de cervezas, bolsas de cementos, sacos de azúcar, ect.

2.2.1 VARIABLES ALEATORIAS

Una variable aleatoria se le llama a una función que asigna un número real a cada uno de los elementos de un espacio muestral S. Y se representa por medio de la letra x.

2.3 DESCRIPCIÓN DE UN PARÁMETRO

Un parámetro: es cualquier característica de una población que sea medible. Como ejemplo se tiene: se desea monitorear con cuanto alcohol, pH, glucosa residual, sólidos disueltos, ect., se esta produciendo cerveza en una industria cervecera.

2.4 DESCRIPCION DE UN ESTADÍSTICO

Un estadístico o estadígrafo: es una medida resultante del análisis de una muestra. Como por ejemplo: retomando el ejemplo anterior, el contenido de alcohol, pH, glucosa residual, sólidos totales disueltos promedios en la cerveza producida a partir de las muestras.

Para visualizar la diferencia entre un parámetro y un estadístico se puede apoyar del cuadro resumen de la Tabla 2.1 (Box. Et al, 1989), tomando para la definición de la media muestral y poblacional.

TABLA 2.1. Resumen para diferenciar la media aritmética entre una población infinita y una muestra.

<i>Población</i>	<i>Muestra</i>
Conjunto muy grande de N observaciones de la que podemos imaginar que proviene la muestra	Pequeño grupo de observaciones disponibles.
<i>Parámetro</i>	<i>Estadístico</i>
Media de la población: $\mu = \sum x / N$	Promedio muestral: $\bar{X} = \sum x / n$

2.5 VARIABILIDAD EN LAS POBLACIONES Y CARACTERÍSTICAS ALEATORIAS

2.5.1 DESCRIPCION

El concepto de variable aleatoria a una población, significa que es una característica expresable numéricamente cuyo valor se desvía de un elemento a otro de la misma. La variabilidad de una variable aleatoria determinada puede expresarse:

- a. Gráficamente
- b. Numéricamente
- c. Como forma sintética a un tipo de Modelo Matemático.

2.6 DESCRIPCION GRAFICA DE LA VARIABILIDAD

En forma general se trata de la presentación grafica de los datos de una serie estadística o descripción grafica de los mismos. En la actualidad se trata de dar el conocimiento en forma simplificada a través de pequeños volúmenes, donde se resume lo que se quiere dar a conocer. Los colaboradores de empresas presentan la información de tal manera que la cabeza de la empresa tenga una visión rápida de lo que esta ocurriendo en la misma, en cuanto a procesos y estados financieros. Así vemos como la información estadística que se les pasa va acompañada de graficas que informan en forma sintética, clara y practica el comportamiento del fenómeno que desean se tenga conocimiento.

La importancia de las representaciones gráficas de los cuadros, datos obtenidos y resultados estadísticos consiste en asimilar rápidamente y sin mucho esfuerzo las principales características de las series estadísticas. Decimos principalmente características porque para un examen exhaustivo del fenómeno que se quiere

conocer, se necesita de un análisis numérico de los resultados del fenómeno de interés. Estas gráficas se clasifican según la naturaleza de las variables, en:

- a. *Gráficas para variables continuas*
- b. *Gráficas para variables discretas*

a.1 *Histograma*

El histograma es una serie de rectángulos yuxtapuestos, de base (eje horizontal) igual al intervalo de clase de la característica medida de la población y altura (eje vertical) correspondiente a las frecuencias respectivas. Para que los rectángulos queden yuxtapuestos se escribirá, en el eje de las abscisas, los límites reales inferiores de cada clase, como se muestra en la figura 2.1. Constituyen una poderosa herramienta para el análisis descriptivo de datos, pues permiten muchas veces poner claramente de manifiesto problemas como:

- Existencia de datos anómalos.
- Mezclas de poblaciones distintas.
- Datos artificialmente modificados.
- No normalidad de los datos.

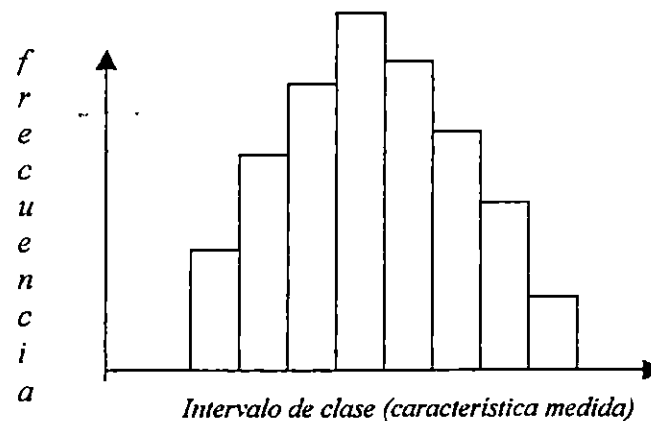


Figura. 2.1. Histograma: distribución de frecuencia versus intervalo de clase

Lo que el histograma representa en forma general son las medidas de tendencia central, la dispersión y de forma general la distribución de los datos. Con un poco de experiencia se aprende a detectar estas irregularidades o anomalías, de manera que para un ingeniero experimentado en ocasiones puede ser suficiente trazar un histograma para resolver un problema de ingeniería.

a.2 *Polígono de frecuencia*

El polígono de frecuencias resulta de unir, por segmentos de rectas, los puntos medios de la parte superior de los rectángulos del histograma. Para hacer la grafica, se escribe en las abcisas (eje horizontal) los puntos medios de cada clase (característica medida del fenómeno en estudio), luego para cada punto medio, se marca con un punto la frecuencia correspondiente a ese punto. Tal como se visualiza en la Figura 2.2.

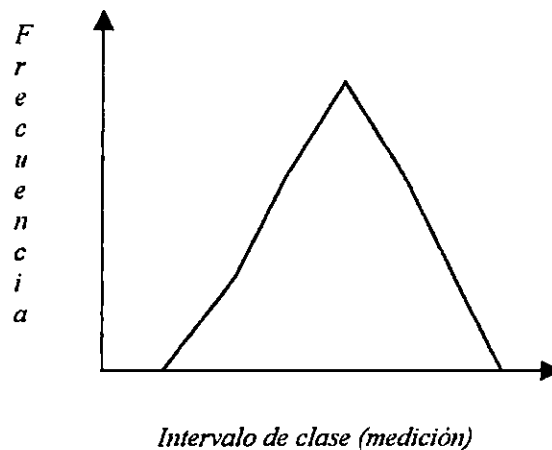


Figura. 2.2. Polígono de frecuencia: *distribución de frecuencias de las medidas versus las medidas de la característica (intervalo de clases) del fenómeno de interés.*

a.3 *Frecuencia acumulada*

Es una grafica para una distribución de frecuencias acumuladas suele tomar la forma de una ojiva, que viene a ser una curva en forma de ese (S). Nos sirve para realizar comparaciones entre un elemento de la población con respecto la población, como, ¿ que % de elementos de la población están por debajo de un elemento de la población en particular?. tal como se muestra en la Figura 2.3.

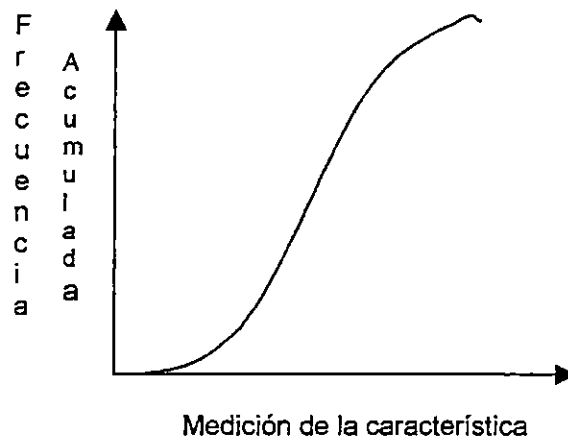


Figura 2.3. Distribución de frecuencia aculada del tipo " mas que"

a.4 *Diagramas de cajas o Box-Whisker*

En dichos diagramas, los valores máximo y mínimo, los cuartiles inferior y superior (percentiles 25 y 75 respectivamente) y la mediana (porcentaje 50) se presentan en una caja rectangular alineada ya sea horizontal o verticalmente. La caja se extiende del cuartil inferior al superior, y es atravesada de un lado a otro por la mediana. A partir de los extremos de la caja se extienden líneas "bigotes" hasta los valores máximo y mínimo. Tal como se muestra en la Figura 2.4.

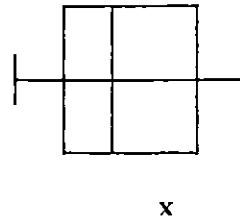


Figura 2.4. Diagrama Box-Whisker

b *Graficas para variables discretas*

Para este tipo de variables no necesita graficar los resultados obtenidos dentro de un sistema de coordenadas cartesianas, como se hizo con las variables continuas. Se vale de una serie de recursos gráficos a veces artísticos para presentar la información estadística en una forma sencilla y agradable a la vista. Entre los gráficos más comunes, se tiene:

1. Grafico de barras o rectángulos
2. Gráficos de sectores o diagrama circular
3. Gráfico polar
4. Pictogramas

b.1 *Grafico de barras o rectángulos*

Esta constituido por un conjunto de barras o rectángulos, separados por distancias iguales, que presenta las distintas categorías o clasificaciones que se hace del fenómeno de interés. Por ejemplo la producción de cemento Pórtland de cementos de El Salvador S.A (CESSA) los años 1995 al 1998, en forma hipotética, ver figura 2.5.

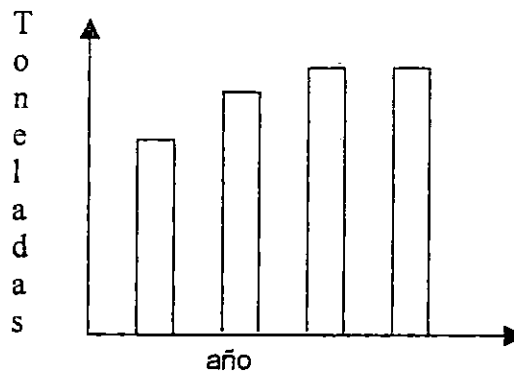


Figura 2.5. Grafico de barras o rectángulos para la producción de cemento portlan desde 1996 al 1998.

2.7 DESCRIPCION NUMÉRICA DE LA VARIABILIDAD

Del mismo modo que las gráficas pueden mejorar la presentación de datos, las descripciones numéricas son también valiosas. Se presentaran varias medidas numéricas para describir las características de los datos. En general se clasifican en:

- 2 *Medidas Descriptivas o De Posición*
- 3 *Medidas De Dispersión*
- 4 *Forma*

2.7.1 MEDIDAS DESCRIPTIVAS O DE POSICIÓN

Anteriormente se intentó describir algunas características de las series estadísticas; esto medio se logra con la representación de series estadísticas por medios de gráficos; ya sean variables continuas o discretas. Pero la descripción es incompleta, necesitamos de otros recursos que den una mejor descripción de las variables que intervienen en fenómeno; es decir, obtener medidas que sean capaces de resumir o representar el comportamiento que toma los valores del fenómeno que se desee analizar; esto se resolverá utilizando medidas de tendencia central, que como su nombre lo indica, son medidas alrededor de las

cuales tienden a girar los de más valore de la serie estadística; vale recordar, que las medidas de tendencia central llevan implícitas las características universales del fenómeno de interés. Las medidas de tendencia central que mas se utilizara en presente trabajo, son:

- a. Media aritmética
- b. Mediana
- c. Cuartiles

Estas medidas de tendencia central o de posición se describen para poblaciones en la tabla 2.2.

2.7.2 MEDIDAS DE DISPERSIÓN

Las medidas de tendencia central no necesariamente proporciona suficiente información para describir los datos en forma adecuada. Las medidas de dispersión describe la variabilidad de una población o muestras, considerando cuanto se desvía o se dispersa una medida de tendencia central de las medidas o observaciones tomadas, como por ejemplo cuanto se desvía la media con respecto a cada medida de la población o muestra, esto a través de la diferencia de la media con respectos a cada medida de la población o muestra elevada al cuadrado, así $(x_i - \bar{x})^2$. Dentro de las medidas de dispersión más importantes se puede mencionar:

1. La varianza
2. La desviación típica o estándar
3. Rango de muestra
4. Coeficiente de variación

Tabla 2.2. Descripción de las medidas de posición para poblaciones y muestras.

Medida de tendencia central	Población	Muestras	Descripción
Media aritmética	$\mu = \frac{\sum_{i=1}^N x_i}{N}$ <p>$\mu =$ media de la población</p> <p>$x_i =$ observaciones</p> <p>$N =$ número de observaciones</p>	$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$ <p>$\bar{X} =$ media muestral</p> <p>$x_i =$ observaciones en la muestra</p> <p>$n =$ número de muestras</p>	<p>La media aritmética es el punto de gravedad de los puntos diversos de una población y una muestra.</p> <p>La media de población se le denomina valor esperado de x o esperanza matemática de x. Y se escribe $E(x)$. Y la media para muestras se le conoce media de muestra</p>
Mediana	$\tilde{\mu} = x_{(\frac{n+1}{2})}$ <p>para n impar</p> $\tilde{\mu} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$ <p>para n par</p>	$\tilde{x} = x_{(\frac{n+1}{2})}$ <p>para n impar</p> $\tilde{x} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$ <p>para n par</p>	<p>La mediana tendencia central, la cual divide la muestra en dos mitades iguales</p> <p>Tiene la ventaja de que no es afectada de manera considerable por los valores extremos.</p>
Cuartiles	$Q_y = \frac{N}{4} y$ <p>donde:</p> <p>$Q_y =$ puede ser Q_1, Q_2 y Q_3</p> <p>$y =$ puede ser 1, 2 o 3</p>	$Q_y = \frac{N}{4} y$	<p>Al igual que la mediana, los cuartiles dividen la población o muestras en cuatro partes iguales.</p> <p>El primer cuartil, Q_1, es un valor con un cuarto, o 25%, de las observaciones por debajo de su valor, para Q_2 es un valor con un medio, o el 50% de las observaciones por debajo y arriba de su valor y para el tercer cuartil, Q_3, es un valor es el 75% de las observaciones por debajo de su valor</p>

Las medidas de dispersión más importantes se describen de forma general en la tabla 2.3, para muestras y para poblaciones.

TABLA 2.3. Descripción de la variabilidad a través de las medidas de dispersión para poblaciones y muestras.

Medidas	Población	Muestras	Observación
Varianza	$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{N}$ <p>donde: N denota el tamaño de la población, esta fórmula es válida para poblaciones finitas</p>	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ <p>donde: n denota el número de muestras.</p>	<p>La varianza para muestras se le denomina varianza muestral y para una población varianza poblacional. La varianza tiene la desventaja que expresa la dispersión de las medidas al cuadrado de las unidades originales.</p>
Desviación típica	$\sigma = \sqrt{\sigma^2}$	$s = \sqrt{s^2}$	<p>Ofrece la ventaja de proporcionar una medida de dispersión en las unidades originales de las medidas, tanto para una población como una muestra.</p>
Intervalo de muestra		$R = \max(x_i) - \min(x_i)$ <p>Es la diferencia entre los intervalos superior y menor</p>	<p>Tiene la inconveniencia de que se ignora toda información que existe entre las observaciones más pequeñas y más grande. Para tamaños de muestra pequeños, digamos $n \leq 10$, esta pérdida de información no es demasiado seria en algunas situaciones.</p>
Coefficiente de Variación	$CV = \frac{\sigma}{\mu}$	$CV = \frac{s}{\bar{x}}$	<p>El coeficiente de variación es útil cuando se compara la variabilidad de dos o más conjuntos de datos que difieren de modo considerable en la magnitud de las observaciones</p>

2.7.2 PARÁMETROS DE FORMA

Para estudiar este tipo de variabilidad ya se ha establecido un modelo matemático, de extraordinaria importancia en toda la Inferencia Estadística. Toda distribución viene completamente caracterizada por su media y su desviación típica, es decir por sus parámetros de posición y de dispersión. Y es útil para visualizar si la

distribución asumida que describirá el fenómeno de estudio se apega a los datos reales obtenidos en el (Romero. Et al, 1993)

Sin embargo un problema frecuente la verificar los datos reales es, si son precisamente los rigen una distribución normal, puestos que pautas de variabilidad que se alejen sensiblemente de la Normal pueden exigir el recurso a tratamientos estadísticos especiales o ser el síntomas de anomalías. Siendo las principales:

- a. Los coeficientes de asimetría
- b. Los coeficiente de curtosis

a. *Los coeficientes de asimetría*

Si los datos son simétricos lo son respecto a su media, y la suma de los cubos de las desviaciones de los datos respecto a dicha media $\sum(x_i - \bar{x})^3$ será nula. En caso contrario dicha suma será positiva si los datos presentan una cola alargada hacia la derecha y negativa si la presentan hacia la izquierda, tal como se ilustra en las figura 2.6 (Romero. Et al 1993). Se define el coeficiente de asimetría (CA) como el promedio (dividiendo por $N-1$ en vez de N) de los cubos de las desviaciones respecto a la media, dividido por el cubo de la desviación típica. La división por s^3 tiene por objeto obtener un coeficiente adimensional, que no dependa de la escala en provengan los datos o observaciones. Y su determinación está dada por la ecuación 2.1 (Romero. Et al, 1993).

$$CA = \frac{\sum (x_i - \bar{x})^3}{s^3 (N-1)} \quad 2.1$$

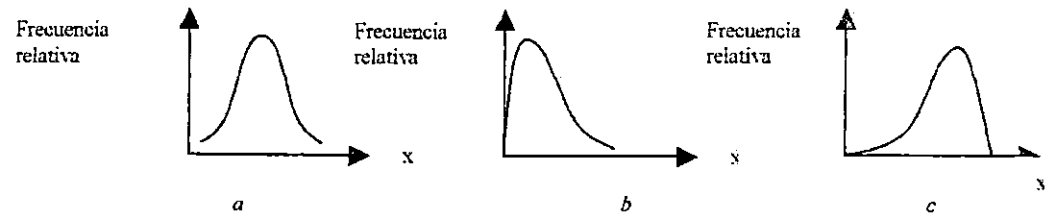


Figura 2.6. Ilustración de la distribución Normal con sus respectivas asimetrías; a. CA igual cero, b. CA positiva y, c. CA negativa.

b. *Coefficiente de Curtosis*

La curtosis puede ser:

1. Leptocúrtico
2. Planicúrtico

b.1 *Leptocúrtico*

Si los datos presentan valores muy alejados de la media con mayor frecuencia de lo que cabría esperar para unos datos normales que tuvieran la misma desviación típica. Frecuentemente valores elevados de la curtosis de un conjunto de datos suele ser síntoma que entre los mismos se incluyen observaciones anómalas (valores de transcripción o algún elemento perteneciente a una población distinta de estudio). Tal como se muestra en la figura 3.7 (Romero. Et al, 1993).

b.2 *Planicúrtico*

Si los datos presentan valores alejados de la media aparecen con una frecuencia menor que los que cabría esperar si los datos siguieran una distribución normal con la misma desviación típica. Una planicúrtosis excesiva puede revelar que los datos han sido modificados artificialmente o manejados en forma arbitraria. Tal como se muestra en la figura 2.7 (Romero. Et al, 1993).

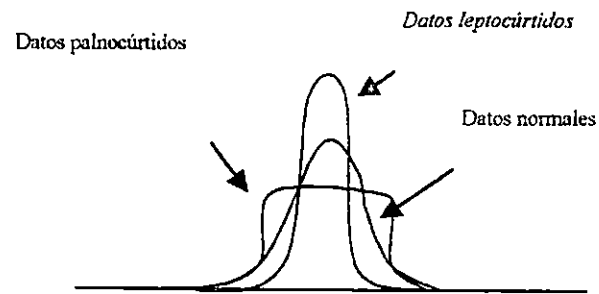


Figura 2.7. Comparación de datos normales con los datos leptocúrtidos y planocúrtidos

El grado de curtosis de un conjunto de datos se mide mediante el *coeficiente de curtosis* (CC), que es el cociente entre el promedio (dividiendo por $N-1$ en vez de N) de las cuartas potencias de las desviaciones respecto a la media y la desviación típica elevada a 4 (Romero. Et al, 1993). Y su determinación está dada por la ecuación 2.2.

$$CC = \frac{\sum (x_i - \bar{x})^4 / (N-1)}{s^4} \quad 2.2$$

2.8 DESCRIPCIÓN DE LA VARIABILIDAD POR MODELOS MATEMÁTICOS.

Cuando la variabilidad de una variable aleatoria determinada puede expresarse también de forma sintética recurriendo a un tipo de modelo matemático, a través de la *distribución de probabilidad*.

Si la probabilidad de que una variable aleatoria es regida por un modelo matemático para adoptar cualquier valor dentro de su rango, se le llama *distribución de probabilidad* de la variable aleatoria.

Por la forma como una variable aleatoria es regida a asumir un valor dentro de la población, la caracterización de la pauta de variabilidad de la distribución de probabilidad se clasifican en:

- a. *Función de probabilidad o cuantía*: cuando el conjunto de valores posibles es discreto (variables aleatorias discretas)
- b. *Función de densidad*: cuando el conjunto de valores posibles de la variables aleatorias es un infinito continuo(variables aleatorias continuas).

2.8.1 FUNCIÓN DE DISTRIBUCIÓN.

Cuando una variable aleatoria discreta asociada a una población como una característica expresable numéricamente cuyo valor fluctúa según el elemento en la población. La probabilidad de que dicha variable aleatoria tome valores comprendidos en un determinado intervalo puede interpretarse intuitivamente como la proporción de los elementos de la población en los que el valor que tome dicha característica pertenezca al intervalo considerado.

Cuando al conjunto de valores posibles que puede tomar una variable aleatoria es discreto, es decir finito o infinito numerable, la forma más sencilla de caracterizar la distribución de probabilidad correspondiente es a partir de la *función de probabilidad*, también llamada algunas veces *función de cuantía o función de masa*, función que da la probabilidad de cada uno de los valores posibles.

En la figura 2.8 se ilustra una distribución de probabilidad discreta hipotética. Nótese que la distribución de probabilidad discreta la probabilidad esta representada por la altura de la función.

Con el fin de caracterizar dicha distribución de probabilidad se define la **Función de Distribución** $P(x)$ de una variable X (Hines. Et al ,1997), de la forma siguiente:

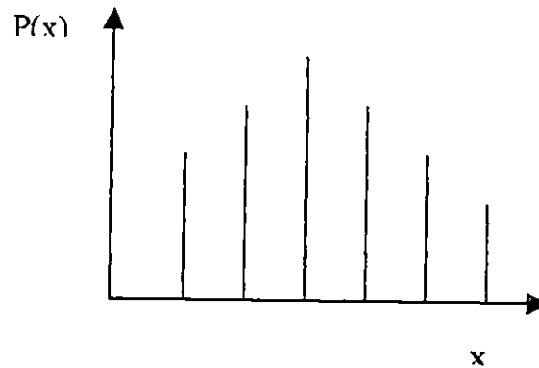


Figura 2.8. Distribución de Probabilidad discreta.

$$\begin{aligned}
 0 \leq p(x) \leq 1 & \quad \text{para todos los valores de } x_i \\
 P(x = x_i) = p(x_i) & \quad \text{para todos los valores de } x_i \\
 \sum_{x_i} p(x_i) = 1 & \quad \quad \quad 2.3
 \end{aligned}$$

2.8.1.1 Valores Esperados, Media, Varianza y Desviación Típica

La media de una distribución de probabilidad de variables discretas es simplemente la esperanza matemática de una variable aleatoria correspondiente. Si una variable aleatoria discreta X adopta los valores x_1, x_2, \dots, x_n , con las probabilidades $p(x_1), p(x_2), \dots, p(x_n)$, (Hines. Et al, 1997) la esperanza matemática o valor esperado es:

$$x_1 \cdot p(x_1) + x_2 \cdot p(x_2) + \dots + x_n \cdot p(x_n) = \sum (\text{valor}) * (\text{probabilidad})$$

Usando la notación \sum . Así pues, se puede definir como:

$$\mu = \sum_{i=1} x_i \cdot p(x_i) \quad 2.4$$

La media de una función de distribución mide su centro en el sentido de un promedio. También puede expresarse la media en términos del valor esperado o promedio a lo largo de una variable aleatoria como $E(x)=\mu$. Donde E es el operador valor esperado.

La extensión o dispersión de una distribución de probabilidad para variables discretas se mide por medio de la varianza (Hines. Et al, 1997), como:

$$\sigma^2 = \sum_{i=1} (x_i - \mu)^2 \cdot p(x_i) \quad 2.5$$

Esta medida no se encuentra en las mismas unidades (o dimensión) que los valores de la variable aleatoria, pero se puede ajustar sacando la raíz cuadrada, denomina desviación estándar, definiéndose como

$$\sigma = \sqrt{\sum_{i=1} (x_i - \mu)^2 \cdot p(x_i)} \quad 2.6$$

Entre las distribuciones de probabilidad para variables aleatorias discretas de mayor importancia por sus aplicaciones en ingeniería, se encuentran:

1. Binomial
2. Hipergeométrica
3. Poisson

En la tabla 2.4 se hace una descripción general de las funciones de distribución antes mencionadas para variables aleatorias discretas.

2.8.1.2 Resumen

Las distribuciones presentadas en esta sección tienen gran uso en las aplicaciones en ingeniería, por consiguiente en el áreas de Ingeniería Química e Ingeniería de Alimentos. La selección de una distribución discreta específica dependerá del grado en el fenómeno de interés que se va a modelar cumpla con las suposiciones relativas a la distribución. Las distribuciones que se presentaron se seleccionaron por su amplia aplicabilidad. En tabla 2.4 se presenta un resumen de estas distribuciones.

2.8.2 FUNCIÓN DE DENSIDAD

Graficando la Función de Densidad se observa como se muestra en la Figura 2.9 (Hines. Et al, 1997).

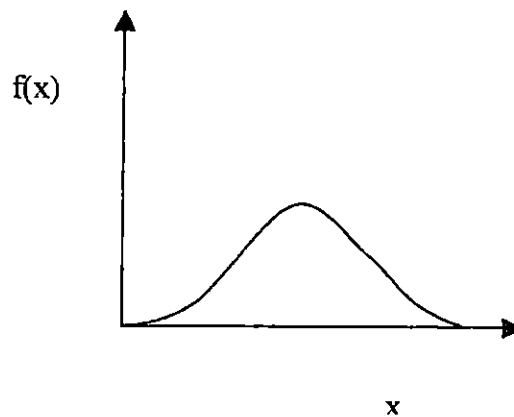


Figura 2.9. Distribución de probabilidad continua.

Con el fin de caracterizar dicha distribución de probabilidad se define la **Función de Densidad** $f(x)$ de una variable X . En general, se formula $P(a \leq x \leq b)$ para la probabilidad asociada con los puntos del espacio muestral, para los cuales el valor de una variable aleatoria cae en el intervalo de a a b .

TABLA 2.4. Resumen de distribuciones discretas (Hines. Et al, 1997).

Distribución	Parámetros	Función de Probabilidad: p(x)	Media	Varianza	Observaciones
Binomial	$n = 1, 2, \dots$ $0 \leq p \leq 1$	$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{para } x=0, 1, 2, \dots, n$ = 0 en otro caso	$\mu = np$	$\sigma^2 = np(1-p)$	<ul style="list-style-type: none"> ▪ Se interesa obtener la probabilidad de tener x éxitos en n ensayos o, otras palabras, x éxitos y $n-x$ fracasos en n intentos. Donde: p y $1-p$ son las probabilidades de éxito y fracaso en cualquier ensayo. ▪ Cumple con los ensayos Bernoulli, la probabilidad de éxito es la misma que la del rechazo, además n es independiente. ▪ Se aplica para determinar productos buenos y malos, es decir que no cumplan con las normas de calidad.
Hipergeométrica	$N = 1, 2, \dots$ $n = 1, 2, \dots, N$ $D = 1, 2, \dots, N$	$p(x) = \frac{\binom{a}{x} \binom{N-a}{n-x}}{\binom{N}{n}} \quad \text{para } x=0, 1, \dots, n$ = 0 en otro caso	$\mu = n \cdot \frac{a}{N}$	$\sigma^2 = \frac{na(N-a)(N-n)}{N^2(N-1)}$	<ul style="list-style-type: none"> ▪ La muestra se extrae de tal forma que en cada extracción sucesiva los elementos restantes del lote tienen la misma oportunidad de ser seleccionados, la probabilidad de que el primer elemento resulte un fracaso es $\frac{a}{N}$, pero para la segunda extracción es $\frac{a-1}{N-1}$, a $\frac{a}{N-1}$, los ensayos no son independientes ▪ Los ensayos de la muestra no se reemplazan, donde x no puede exceder de a y $n-x$ no exceder de $N-a$. ▪ Su aplicación sería la misma que para la distribución binomial, con la observación que son eventos no independientes, por el hecho que los ensayos no se reemplazan

Continuación de la Tabla 2.4

Poisson	c)0	$p(x_i) = \frac{\lambda^x \cdot e^{-\lambda}}{x!} \quad \text{para } x=0,1,2,\dots$ $= 0 \text{ en otro caso}$	$\mu = \lambda$	$\mu = \lambda$	<ul style="list-style-type: none"> ▪ Es cuando n es grande y p es pequeña, donde λ (lambda) igual al producto $n \cdot p$ ▪ Las distribuciones de Poisson tiene muchas aplicaciones aparte de aproximar las probabilidades binomiales, se han tabulado extensamente, tablas del apéndice B, da valores de λ en diversos incrementos de 0.02 a 25. ▪ Tiene su aplicación en los fenómenos de colas, para el caso de Ingeniería Química e Ingeniería de Alimentos, su aplicación son en tiempo de espera para materia prima y producto terminado.
---------	-----	--	-----------------	-----------------	---

Se define la *función de densidad*, $f(x)$ como

$$f(x) = \frac{d}{dx} F(x) \quad 2.7$$

y resulta que

$$F(x) = \int_{-\infty}^x f(t) dt \quad 2.8$$

2.8.2.1 Valores Esperados, Media, Varianza y Desviación Estándar

El valor esperado se de la siguiente manera (Hines. Et al, 1997):

$$\mu = \int_{-\infty}^{\infty} xf(x) dx \quad 2.9$$

Para la *varianza*, que se denota como σ^2 ; como anteriormente, mide la dispersión de una densidad de probabilidad en el sentido de que da el valor esperado de la desviación cuadrada respecto a la media (Hines. Et al, 1997), como:

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2 \quad 2.10$$

Asimismo, σ es llamada de nueva cuenta *desviación estándar*.

En la tabla 2.5 (Hines. Et al, 1997) se resume las Funciones de Densidad más importantes en las áreas de Ingeniería.

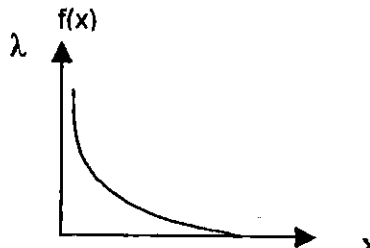
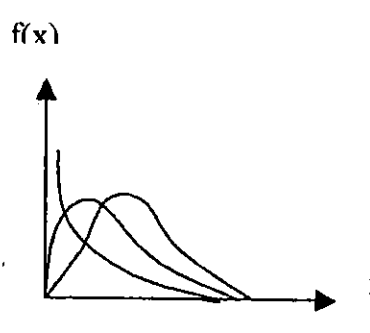
2.8.2.2 Distribución Normal

Se afirma que una variable aleatoria X tiene una distribución normal con un medio μ ($-\infty < \mu < \infty$) y varianza $\sigma^2 > 0$.

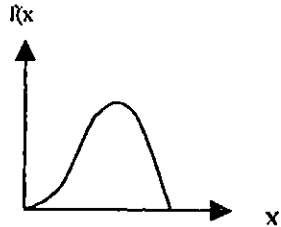
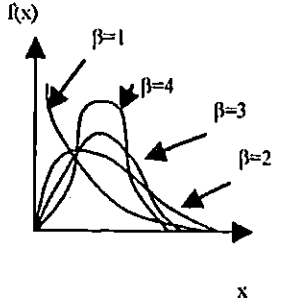
TABLA 2.5. Cuadro resumen de las distribuciones de densidad de mayor en el área de Ingeniería (Hines. Et al, 1997)

Distribución	Función de Densidad	Media	Varianza	Forma	Aplicaciones
Uniforme	$f(x) = \frac{1}{\beta - \alpha} \quad \alpha \leq x \leq \beta$ $= 0 \quad \text{en otro caso}$	$\frac{\alpha + \beta}{2}$	$\frac{(\beta - \alpha)^2}{12}$		<ul style="list-style-type: none"> ▪ Donde α y β son constantes reales con $\alpha < \beta$ ▪ Adviértase que todos los valores de x de α a β son "igualmente probables" en el sentido de que la probabilidad de que x se encuentre en un intervalo de amplitud ▪ Se aplica cuando se elige un punto al azar y que el valor elegido está distribuido de manera uniforme en α y β

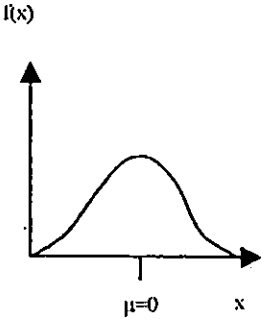
Continuación de la Tabla 2.5

<p>Exponencial</p>	$f(x) = \lambda e^{-\lambda x} \quad \text{para } x \geq 0$ $= 0 \quad \text{en otro caso}$	$1/\lambda$	$1/\lambda^2$		<ul style="list-style-type: none"> ▪ donde el parámetro λ es una constante positiva real ▪ La distribución exponencial está estrechamente relacionada con la distribución de Poisson. ▪ Se utiliza para medir el tiempo de falla de un producto manufacturado, es decir el tiempo de vida útil.
<p>Gamma</p>	$f(x) = \frac{\lambda}{\Gamma(\alpha)} (\lambda x)^{\alpha-1} e^{-\lambda x}$ $= 0 \quad \text{en otro caso}$	α/λ	α/λ^2		<ul style="list-style-type: none"> ▪ Los parámetros son $\alpha > 0$ y $\lambda > 0$. El parámetro α suele llamarse el <i>parámetro de forma</i>, y λ recibe el nombre de <i>parámetro de escala</i>. ▪ Si $\alpha=1$ la distribución gamma se reduce a la distribución exponencial. ▪ Su aplicación es para cuando una población se desvíe de la distribución exponencial y básicamente cuando se requiera una función de confiabilidad y tiene una aproximación a la distribución ji cuadrada

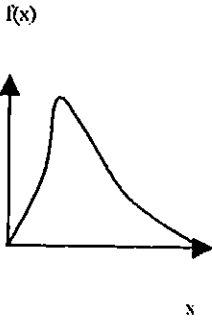
Continuación de la Tabla 2.5.

Beta	$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$ <p>para $0 < x < 1, \alpha > 0, \beta > 0$ $= 0$ en otro caso</p>	$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$	 <p>Para $\alpha=3$ y $\beta=2$</p>	<ul style="list-style-type: none"> ▪ Si $\alpha=1$ y $\beta=1$ se tiene un caso especial de distribución uniforme ▪ Su curva característica son negativas ▪ Su mayor aplicación es cuando se necesita una variable aleatoria adopte valores en el intervalo de 0 a 1
De Weibull	$f(x) = \frac{\beta}{\delta} \left(\frac{x-\gamma}{\delta}\right)^{\beta-1} \exp\left[-\left(\frac{x-\gamma}{\delta}\right)^\beta\right]$ <p>$= 0$ en otro caso</p>	$\gamma + \delta \left(1 + \frac{1}{\beta}\right)$	$\sigma^2 = \delta^2 \left\{ \Gamma\left(1 + \frac{2}{\beta}\right) - \left[\Gamma\left(1 + \frac{1}{\beta}\right) \right]^2 \right\}$	 <p>Para $\gamma=0$ y $\delta=1$</p>	<ul style="list-style-type: none"> ▪ Esta estrechamente relacionada con la distribución exponencial ▪ Sus parámetros son γ, ($-\infty < \gamma < \infty$) el parámetro de localización, $\delta > 0$ el parámetro de escala, y $\beta > 0$, el parámetro de forma. ▪ La más importante área de aplicación ha sido como un modelo para el tiempo de falla en componentes y sistemas eléctricos y mecánicos.

Continuación de la Tabla 2.5.

<p>Normal</p>	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\left(\frac{1}{2}\right)\left[\frac{(x-\mu)}{\sigma}\right]^2\right\}$ <p>= 0 en otro caso</p>	μ	σ^2	 <p>The graph shows a bell-shaped curve on a coordinate system. The vertical axis is labeled f(x) and the horizontal axis is labeled x. The peak of the curve is at the origin, which is marked with a vertical line and labeled mu=0.</p>	<ul style="list-style-type: none"> ▪ La gráfica para poblaciones normales es simétrica, con media cero y con áreas iguales en ambos lados de la gráfica ▪ Es la distribución de densidad de mayor uso, dado que para muestras grandes la mayorías de las poblaciones tienden a comportarse de manera normal. Además de existir muchas aproximaciones, e incluso para variables aleatorias discretas.
---------------	--	-------	------------	---	--

Continuación de la Tabla 2.5.

<p>lognormal</p>	$f(x) = \frac{1}{x\sigma_Y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu_Y}{\sigma_Y}\right)^2}$ <p>para $x > 0$ $= 0$ en otro caso</p>	$e^{\mu_Y + \frac{1}{2}\sigma_Y^2}$	$\sigma_X^2 = e^{2\mu_Y + \sigma_Y^2} (e^{\sigma_Y^2} - 1)$		<ul style="list-style-type: none"> ▪ La distribución lognormal ocurre en la práctica cada vez que encontramos una variable aleatoria tal que su logaritmo tiene una distribución normal ▪ Su importancia radica por el hecho de que a menudo se encuentran poblaciones que no se comporten normales y, al tratarlas con propiedades de logaritmo se normaliza y se trata como tal. Su aplicación es amplia en el área. La distribución lognormal ocurre en la práctica cada vez que encontramos una variable aleatoria tal que su logaritmo tiene una distribución normal de Ingeniería.
------------------	--	-------------------------------------	---	---	--

La distribución se emplea de manera tan amplia que a menudo se recurre en la notación abreviada $X \sim N(\mu, \sigma^2)$ para indicar que la variable aleatoria X se distribuye normalmente con la media μ y varianza σ^2 . Entre las propiedades de la distribución normal se tiene (Romero. Et al, 1993):

1. $\int_{-\infty}^{\infty} f(x)dx = 1$ para las funciones de densidad
2. $f(x) \geq 0$ para todo x
3. $\lim_{x \rightarrow \infty} f(x) = 0$ y $\lim_{x \rightarrow -\infty} f(x) = 0$
4. $f[(x + \mu)] = f[-(x - \mu)]$. La densidad es simétrica alrededor de μ
5. El valor máximo de f ocurre en $x = \mu$.
6. Los puntos de inflexión de f están en $x = \mu \pm \sigma$
7. El coeficiente de asimetría será igual a cero
9. El coeficiente de curtosis igual a 3.

Además la distribución normal debe cumplir con lo siguiente:

1. $P((\mu - \sigma) < X < (\mu + \sigma)) \approx 0.68$
2. $P((\mu - 2\sigma) < X < (\mu + 2\sigma)) \approx 0.95$
3. $P((\mu - 3\sigma) < X < (\mu + 3\sigma)) \approx 0.997$

En otras palabras es bastante raro (probabilidad menor que el 5%) encontrar un valor de una variable normal que difiera de su media en más de dos desviaciones típicas, y es extremadamente improbable (probabilidad menor que el 3 por mil) encontrar un valor que difiera de la media en más de tres desviaciones típicas.

2.8.2.3 Distribución Normal Estándar

Puesto que la densidad de probabilidad normal no puede integrarse en forma cerrada entre cada par de límites a y b , probabilidades relacionadas con las distribuciones normales se obtienen usualmente de tablas especiales, apéndice C. Esta tabla corresponde a la *distribución normal estándar*, es decir, la distribución normal con $\mu=0$ y $\sigma=1$, y sus valores de entrada se muestra en la figura 2.10 (Johnson, 1997).

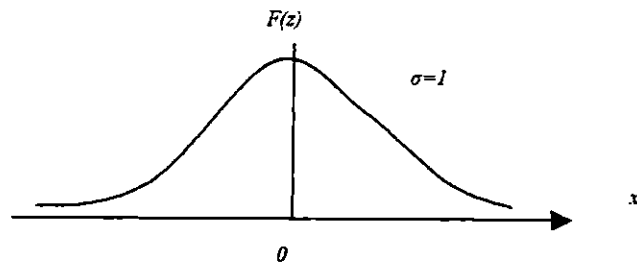


Figura 2.10. Distribución normal estándar

$$F(z) = \frac{1}{2\pi} \int_{-\infty}^z e^{-t^2/2} dt \quad 2.11$$

Para hacer uso del apéndice C en relación con una variable aleatoria X con una distribución normal con la media μ y la varianza σ^2 , se remite a la correspondiente *variable aleatoria estandarizada* (Johnson, 1997), de la siguiente manera:

$$Z = \frac{X - \mu}{\sigma} \quad 2.12$$

Asimismo, si se desea determinar la probabilidad de que una variante aleatoria con distribución normal con la media μ y la varianza σ^2 adopte un valor entre a y b ,

solo se tiene que calcular la probabilidad de que una variable aleatoria con la distribución normal estándar adopte un valor entre $\frac{a-\mu}{\sigma}$ y $\frac{b-\mu}{\sigma}$; simbólicamente,

$$P(a < X < b) = F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right) \quad 2.13$$

2.8.2.4 Comprobación si los Datos son Normales

Además, se comprobar haciendo uso de métodos gráficos, tales como:

1. Análisis de histogramas (visto en la descripción de la variabilidad en forma grafica). y
2. Análisis de gráficos en papel probabilístico.

2. Gráficas en papel probabilístico normal¹

La finalidad de utilizar papel probabilístico es que: Si los puntos se sitúan aproximadamente a lo largo de una línea recta, se estaría hablando que los datos o resultados obedecen a una distribución normal. Así como se muestra en la Figura 2.11 (Romero. Et al, 1993).

2.8.2.5 Aproximaciones Normales

Un teorema de gran importancia en la Estadística Matemática postula que, en condiciones muy generales respecto a las distribuciones de los sumandos, la suma de variables aleatorias independientes tiende a una distribución normal a medida aumenta el numero de sumandos. A este resultado se le conoce con el nombre de Teorema Central del Limite.

¹ Se utiliza para distintos tipos de distribuciones

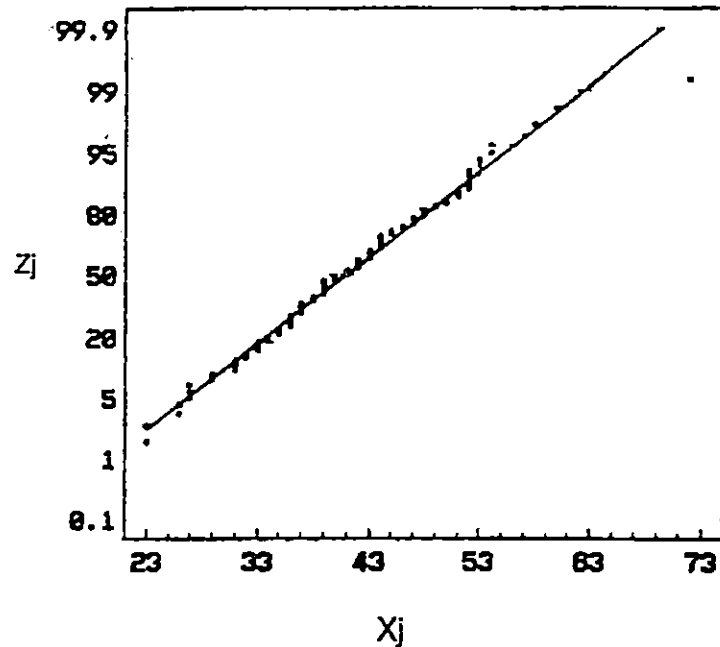


Figura 2.11. (Ref. 10) Representación gráfica en papel probabilístico para datos normales.

Teorema 1: Si X_1, X_2, \dots, X_n es una secuencia de n variables independientes con $E(X_i) = \mu_i$ y $V(X_i) = \sigma_i^2$ (ambas finitas) y $Y = X_1 + X_2 + \dots + X_n$, entonces, bajo ciertas condiciones generales (Hines. Et al, 1997), son:

$$Z_n = \frac{Y - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \quad 2.14$$

tiene una distribución $N(0, 1)$ aproximada conforme n se aproxima a infinito.

Si X_1, X_2, \dots, X_n es una secuencia de n variables aleatorias independientes y distribuidas idénticamente con $E(X_i) = \mu_i$ y $V(X_i) = \sigma_i^2$ (ambas finitas) y $Y = X_1 + X_2 + \dots + X_n$, entonces

$$Z_n = \frac{Y - n\mu}{\sigma\sqrt{n}} \quad 2.15$$

tiene una distribución aproximada $N(0,1)$ en el sentido que si F_n es la función de distribución de Z_n .

2.8.2.6 Transformaciones de Observaciones Hasta Casi la Normalidad

En la Tabla 2.6 (Johnson, 1997) se muestran algunas de las transformaciones de datos de mayor utilidad.

Tabla 2.6. Algunas transformaciones útiles.

Reducción de valores grandes	Incrementos de valores grandes
$-\frac{1}{x}, \ln x, x^{1/4}, \sqrt{x}$	x^2, x^3

Si las observaciones transformadas ofrecen un diagrama de puntaje normal casi en línea recta, suele ser conveniente emplear la normalidad de esta nueva escala para realizar cualquier análisis estadístico.

2.9 EJEMPLOS DEL CAPÍTULO

Ejemplo 2.1

En las lecturas de presión de rompimiento de las botellas para refrescos no retornables de un 1 litro de capacidad, presentadas en la Tabla 2.7. la presión de rompimiento en psi, el cliente desea averiguar si el fabricante de las botellas cumple con la norma de calidad, siendo la norma de calidad de 230 psi, para ello se sometieron a prueba 100 botellas. Los resultados de la prueba se presentan en la Tabla 2.7.

- Hacer una distribución de frecuencia de las presiones de rompimiento de las

botellas

- b. Elaborar los diagramas de histograma y de frecuencia acumulada para los resultados de la pruebas obtenidas.
- c. La media, mediana y los cuartiles (1, 2, 3).
- d. Elaborar un diagrama de caja para los resultados de la prueba.
- e. La desviación estándar, varianza, coeficiente de variación, coeficiente de curtosis y el coeficiente de asimetría.

Tabla 2.7. Resistencia al rompimiento en libras por pulgadas cuadradas de 100 botellas de vidrio no retornables de refresco de 1 litro.

285	197	346	280	265	200	221	265	261	278
205	286	317	242	254	235	176	262	248	250
293	274	242	260	281	246	248	271	260	265
307	243	258	321	294	328	263	245	274	270
220	231	276	228	223	296	231	301	337	298
268	267	300	250	260	276	334	280	250	257
260	281	208	299	308	264	280	274	278	210
234	265	187	258	235	269	265	253	254	280
299	214	264	267	283	235	272	287	274	269
215	318	271	293	277	290	283	258	275	251

Solución:

- a Se determina el limite inferior (Li) y superior (Ls) de la serie estadística. para el caso $Li=176$ psi y 346 psi respectivamente. El rango o amplitud (R) se determina de la siguiente manera:

$$R=Ls-Li=(346-176) \text{ psi}-170 \text{ psi}$$

Como para satisfacer y cubrir todos los resultados de la serie estadística se tomará como limite inferior real (Lir) ligeramente menor al limite inferior de la serie estadística y análogamente con el limite superior real (Lsr) ligeramente mayor al

limite superior. Para el caso más y menos 6 (en forma arbitraria).

$$Rr - Lsr \cdot Lir = (Ls + 4) - (Li - 4) = (346 + 6) - (176 - 6) = 180 \text{ psi}$$

Se tomará el número de clases (N_c) arbitrariamente que cumpla con el rango o intervalo real de la serie estadística. Para el caso $N_c = 9$ (siendo el recomendable un número de clases de 5 a 20). Luego se determina el intervalo de clase (I_c), de la manera siguiente:

$$I_c = Rr / N_c = 180 / 9 = 20$$

Con esto se puede hacer una distribución de frecuencias como se muestra en la Tabla 2.8.

Tabla 2.8. . Distribución de frecuencia para los datos de resistencia al rompimiento perteneciente al ejemplo 2.1.

Intervalo de clases	Frecuencia	Frecuencia relativa	Frecuencia acumulada
170-190	2	0.02	2
190-210	5	0.05	7
210-230	6	0.06	13
230-250	16	0.16	29
250-270	30	0.30	59
270-290	24	0.24	83
290-310	10	0.10	93
310-330	4	0.04	97
330-350	3	0.03	100
Total	100	1.00	

- b. Para elaborar un diagrama de histograma, se gráfica la frecuencia relativa vrs. el intervalo de clases respectivo. Como se muestra en la figura 2.12.

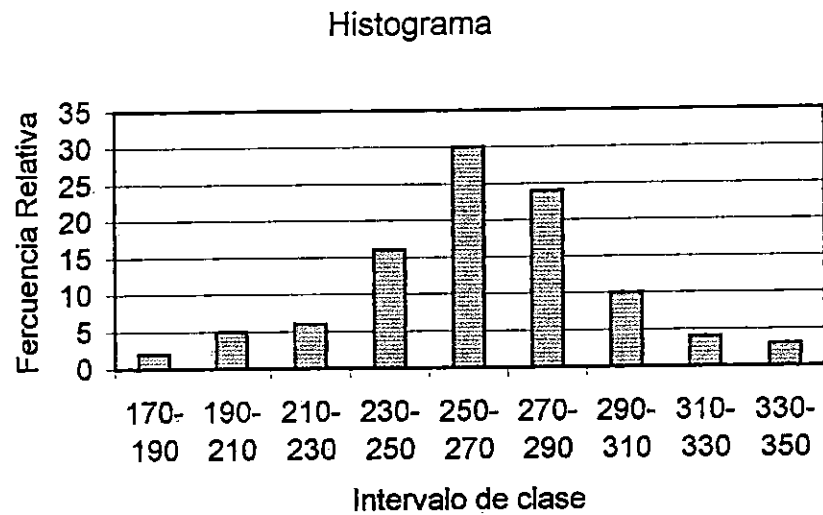


Figura 2.12. Histograma para los datos del ejemplo 3.1.

- c. Para elaborar un diagrama de frecuencia acumulada o tipo ojiva se gráfica la media de cada intervalo de clase vs. la frecuencia acumulada, como se muestra en la Figura 2.13. Por ejemplo, para calcular la media del primer intervalo de clases, es de la siguiente manera: $lcm_i = ((L_{i+1} - L_i) / 2 + L_i) = ((190 - 170) / 2 + 170) = 180$. De forma análoga para los siguientes intervalos medios.

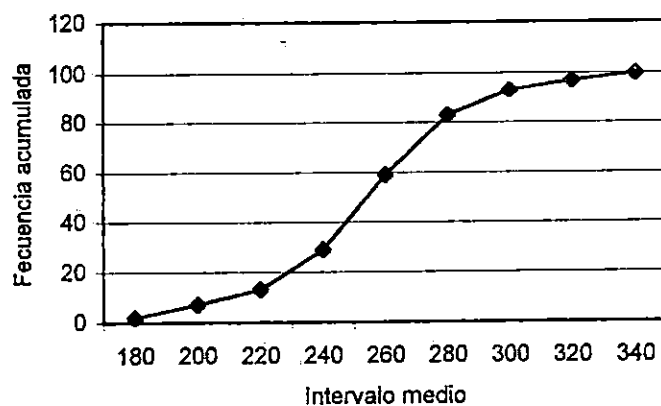


Figura 2.13. Gráfica de frecuencia acumulada vs. Intervalo de clases medio

d. Para calcular la media aritmética para una muestra, se calcula por la formula:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{26406}{100} = 264.06 \text{ psi}$$

para la mediana, como el número de lecturas es par, se determina por la media de las lecturas proporcionadas por la ecuación:

$$\tilde{x} = \frac{n}{2} \text{ y } \frac{n+2}{2}$$

resultando las lecturas 50 y 51 en forma ascendente, siendo estas 265 y 265 respectivamente, entonces la mediana es: 265 psi

Para los cuartiles 1, 2 y 3, se determina por la ecuación np , siendo p : 0.25, 0.50 y 0.75 respectivamente para cada cuartil.

- i. $Q_1=100*0.25$, se determina sacando el valor medio de las lecturas, 25 y 26. Para el caso sería 248 psi.
 - ii. $Q_2=$ a la mediana =265 psi.
 - iii. $Q_3= 100*0.75=75$, se determina sacando el valor medio de las lecturas, 75 y 76. Para el caso es: 280 psi.
- e. Para elaborar el diagrama de caja o de bigotes mostrados en la Figura 2.14, se considera los parámetros siguientes:

$$Li=176 \text{ psi}; Ls=346 \text{ psi}; Q_1=248 \text{ psi}; Q_2=265 \text{ psi y } Q_3=280.5 \text{ psi.}$$

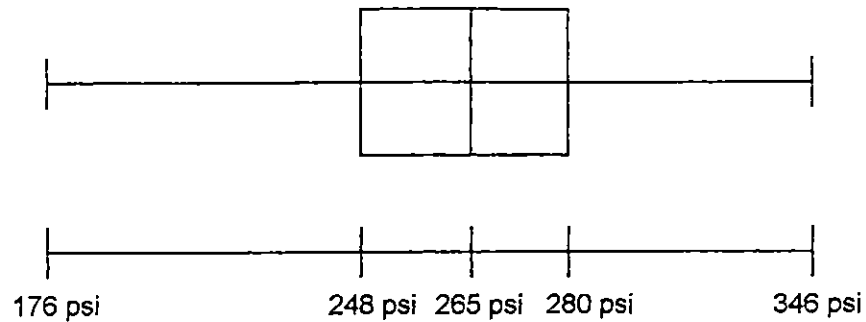


Figura 2.14. Diagrama de bigotes para el ejemplo 2.1

- f. Para determinar la varianza se utiliza la ecuación presentada en el Tabla 2.2, siendo esta:

$$s^2 = \frac{\sum_i^n (x_i - \bar{x})^2}{n-1}$$

así como la sumatoria al cuadrada de la diferencia de las lecturas con la media aritmética resultó: 1025.148 psi².

Para la desviación estándar σ se procede a sacar raíz cuadrada de la varianza de muestra, resultando: 32.018 psi.

Para el coeficiente de variación se relaciona la desviación estándar con la media aritmética de la muestra, de la siguiente manera:

$$CV = \frac{s}{\bar{x}} * 100 = \frac{32.018}{26206} * 100 = 12.21\%$$

- g. Para los parámetros de forma (coeficiente de asimetría y de curtosis), se determina por las ecuaciones 2. y 2. .

Para el coeficiente de asimetría., se tiene:

Para el coeficiente de asimetría., se tiene:

$$CA = \frac{\sum_i^n (x_i - \mu)^3 / (n-1)}{s^3} = -0.129$$

para el coeficiente de curtosis, se tiene:

$$CC - 3 = \frac{\sum_i^n (x_i - \mu)^4 / (n-1)}{s^4} = 0.129$$

entonces la $CC=3.129$ entonces el confortamiento de los datos da como resultado una curva mas pronunciada que una distribución normal (leptocúrticos).

Ejemplo 2.2

En el proceso de empaque de bolsas de detergente se tomaron 80 mediciones, del contenido en peso en gramos en bolsitas de detergentes: tal como se muestra en la tabla 2.9.

Tabla 2.9. Contenido en gramos de las bolsas de detergentes.

4.1	5.0	2.0	2.6	4.5	8.1	5.7	2.5
3.5	6.3	5.5	1.6	6.1	5.9	9.3	4.2
4.9	5.6	3.8	4.4	7.1	4.6	7.4	3.5
4.9	5.1	4.6	6.3	8.3	6.3	8.8	1.0
5.3	5.4	4.4	2.9	7.5	5.7	5.3	3.0
4.2	5.2	7.0	3.7	6.7	5.8	6.9	2.8
6.0	8.2	6.1	7.3	8.2	6.2	4.3	2.2
5.2	5.5	3.5	7.1	7.9	5.6	5.4	3.9
6.8	8.2	4.2	4.2	5.5	6.2	3.5	3.4
6.8	4.7	4.6	4.1	4.7	5.0	3.4	7.1

Determinar:

- a. Resumir los datos en forma tabular

- b. Resuma los datos en forma gráfica
- c. Calcule los estadísticos
- d. Concluya sobre la normalidad de los datos
- e. Calcule los siguientes tres conjuntos de límites: $\bar{x} \pm 1s$, $\bar{x} \pm 2s$ y $\bar{x} \pm 3s$. Para cada conjunto, calcule el porcentaje de los valores de los datos que caen dentro de los límites. Compare estos porcentajes con los porcentajes teóricos con base a la distribución normal.

Solución:

a Para resumir los datos en forma tabular se necesita determinar: el límites superior e inferior, amplitud o rango de la serie estadística, además establecer los límites superior real e inferior real, el número de clases y ancho de clase. Para el caso:

$$Ls=9.3 \text{ gramos}$$

$$Li=1.0 \text{ gramos}$$

$$R=8.3 \text{ gramos}$$

$$Lsr=Ls+0.3=9.6 \text{ gramos, para abarcar los valores máximos}$$

$$Lir=Li-0.3=0.7 \text{ gramos para abarcar los valores mínimos}$$

$$Rr=8.9 \text{ gramos}$$

$$\text{Número de clases}=10$$

$$\text{Intervalo de clases (Ic)}=0.89$$

Considerando lo anterior se resume los datos agrupados en la Tabla 2.10.

- b. Para resumir los datos en forma gráfica, se tienen los histogramas y polígonos de frecuencia. Presentados en las figuras 2.15 y 2.16.

Tabla 2.10. Resumen de los datos del contenido en gramos de las bolsitas de detergentes

Intervalo de clases	Frecuencia	Frecuencia relativa	Frecuencia acumulada
0.70- 1.59	1	0.0125	1
1.59-2.48	3	0.0375	4
2.48 - 3.37	5	0.0625	9
3.37-4.26	15	0.1875	24
4.26-5.15	14	0.1750	38
5.15-6.04	16	0.2000	54
6.04-6.93	11	0.1375	65
6.93-7.82	7	0.0875	72
7.82-8.71	6	0.0750	78
8.71 - 9.60	2	0.0250	80
Total	80	1.0000	

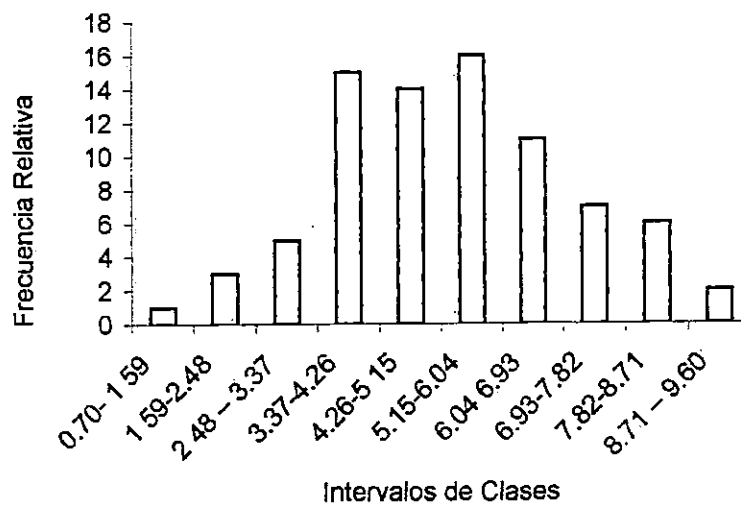


Figura 2.15. Histograma de frecuencia para los datos del ejemplo 2.2.

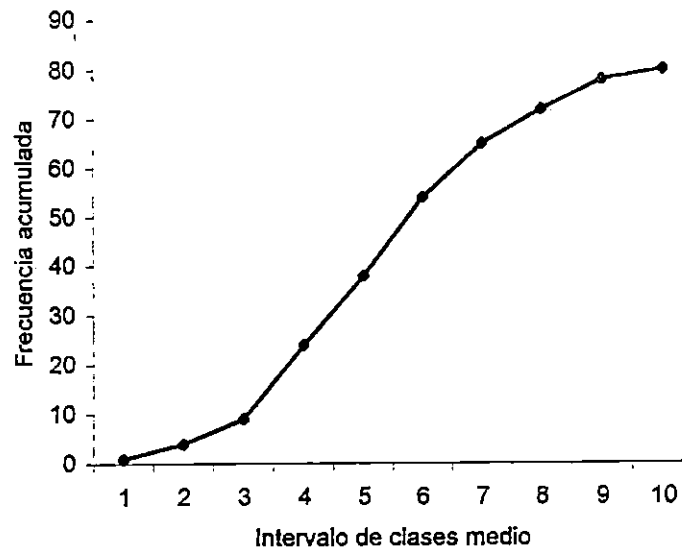


Figura 2.16. Diagrama de polígono para los datos del ejemplo 2.2

- b. Los estadísticos de mayor importancia son: la media muestral y varianza muestral, calculados a partir de las ecuaciones ocupadas en el ejemplo 2.1. Entonces se tiene:

$$\bar{x} = 5.25375 \text{ gramos}$$

$$s^2 = 3.104036 \text{ gramos}^2$$

- c. Para poder concluir sobre la normalidad de los datos presentados, se procede de la siguiente forma:
- i. Los histogramas. Observando si los datos se comportan siguiendo la campana de Gauss. Graficada en literal b.
 - ii. Graficando los datos en papel probabilístico. Presentado en la Figura 2.17.

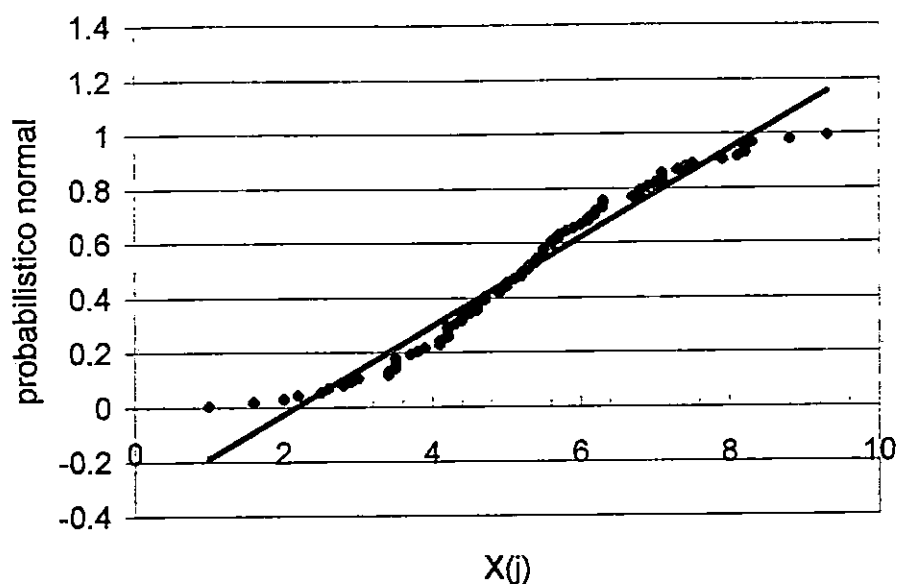


Figura 2.17. Gráfica de probabilístico normal para los datos del ejemplo 2.2.

- iii. Calculando los coeficientes de asimetría y de curtosis. Utilizando las mismas ecuaciones usadas en el ejemplo 2.1, resultando:

para el coeficiente de asimetría:

$$CA = 0.026306$$

con el resultado se tiene una distribución ligeramente desviada de una distribución normal ($CA=0$), siendo ligeramente positiva con cola alargada a la derecha.

Para el coeficiente de curtosis:

$$CC - 3 = -0.307212$$

siendo $CC=3$ para una distribución normal, los datos da un

CC=2.693, siendo ligeramente menor a 3 de una distribución normal, presentándose los datos como planicúrticos. (NOTA: Los datos fueron procesados en el STAGRAPHICS, la cual define el coeficiente de curtosis ya directamente como el valor CC-3, por lo que en datos normales el coeficiente debe resultar próximo a cero)

En conclusión: *Los datos están distribuidos aproximadamente en forma normal.*

- d. Para evaluar los porcentajes de los datos que caerían en los tres intervalos y compararlos con los de una distribución normal, se toman los valores que caen en los límites propuesto, tal como se muestra en la tabla 2.11.

Tabla 2.11. Comparación de los porcentajes de los datos de la serie estadística con los de una población normal, según los intervalos propuestos

Limites	Intervalos	% de datos que caen de la serie estadística	% de datos que caería en una distribución normal.
$\bar{x} \pm S$	$P(x)(3.49 < x < 7.02)$	71.25	68.00
$\bar{x} \pm 2S$	$P(x)(1.73 < x < 8.78)$	97.50	95.00
$\bar{x} \pm 3S$	$P(x)(-0.03 < x < 10.54)$	100.00	100.00

Ejemplo 2.3

Se esta probando una nueva pasta dentífrica, a cualquier persona con una probabilidad de 0.20, el fenómeno es regido por una distribución binomial, determinar (Johnson, 1997):

- La probabilidad de que les desagrade a 5 de 15 personas.
- La media
- La varianza

Solución:

- a. Por las tablas del Anexo H, se tiene:

$$B(5; 18, 0.20) = B(5; 18, 0.20) - B(4; 18, 0.20)$$

$$= 0.8671 - 0.7164$$

$$= 0.1507.$$

- b. La media se determina:

$$\mu = np = 18 \cdot 0.20 = 3.6$$

- c. La varianza es:

$$\sigma^2 = np(1-p) = 18 \cdot 0.20 \cdot 0.80 = 2.88$$

Ejemplo 2.4

Un embarque de 20 grabadoras contiene 5 defectuosas. Si 10 de ellas se eligen al azar para su inspección, asumiendo que tiene una distribución hipergeométrica, determinar (Johnson, 1997):

- La probabilidad de que 2 grabadoras sean defectuosas.
- La media
- La varianza

Solución:

- a. La probabilidad de que 2 unidades salgan defectuosas, se determina por la relación:

$$h(2x : 10, 5, 20) = \frac{\binom{a}{x} \binom{N-a}{n-x}}{\binom{N}{n}}$$

donde: a y N son grabadoras defectuosas y número de grabadoras originalmente, x y n son para el caso particular. Entonces:

$$h(2x : 10, 5, 20) = \frac{\binom{5}{2} \binom{15}{8}}{\binom{20}{10}} = \frac{10 * 6435}{184756} = 0.348$$

b. Para la media:

$$\mu = n \frac{a}{N} = 10 \frac{5}{20} = 2.5$$

seria la media de una muestra de 10 grabadoras.

c. Para la varianza:

$$\sigma^2 = \frac{na(N-a)(N-n)}{N^2(N-1)} = \frac{10 * 5 * (20-5)(20-10)}{20^2 * (20-1)} = 0.987$$

Ejemplo 2.5

En la inspección de hojalata producida por un proceso electrolito continuo, se identifica 0.2 imperfecciones en promedio por minuto. Determinar: a) Una imperfección en 3 minutos, b) La media y c) La varianza (Johnson, 1997)

Solución:

Como es una distribución de Poisson

- a. La probabilidad se puede determinar por la Tabla del Anexo G o por la relación:

$$f(x, \lambda) = \frac{\lambda^x e^{-\lambda}}{x} = \frac{0.6^1 * e^{-0.6}}{1} = 0.329$$

por tablas:

$$F(1; 0.6) - F(0; 0.6) = 0.878 - 0.549 = 0.329$$

- b. La media es $\lambda = 0.6$
- c. La varianza es $\lambda = 0.6$.

Ejemplo 2.6

En una planta de investigación, cuentan con aparatos de análisis químicos que utilizan isótopos radiactivos como patrones. Desean modelar el riesgo referente a la resistencia de los soportes para resistir la aceleración máxima causada por los sismos de la zona localizada. La opinión de expertos sugiere que la Ln(resistencia) esta distribuida con media de 4.0 y varianza de 0.99. Determine la probabilidad de que los soportes sobrevivan a una aceleración máxima de 33 (Hines. Et al, 1997).

Solución:

Por el hecho de tener un término con Ln, referente a la resistencia, el problema se comporta como una distribución log-normal. Entonces:

$$\alpha = \mu = 4 \text{ y } \beta = \sqrt{\sigma^2} = 0.30$$

como la probabilidad es un éxito (resistir) o un fracaso (no resistir) referente a un patrón desconocido para la cual la probabilidad es igual a 1, la distribución queda:

$$F\left(\frac{\ln b - \alpha}{\beta}\right) - F\left(\frac{\ln a - \alpha}{\beta}\right) = \int_{\ln a}^{\ln b} \frac{1}{2\pi\beta} e^{-(y-\alpha)^2 / 2\beta^2} dy$$

la función acumulativa y luego sustituyendo y calculando:

$$1 - F\left(\frac{\ln 33 - 4}{0.3}\right) = 0.9535$$

Ejemplo 2.7

El promedio de la llegada de camiones a cargar mercadería en una bodega industrial es de 3 por hora ¿Cuál es la probabilidad de que el tiempo de arribo de sucesivos camiones sea

- Menor que 5 minutos
- De al menos 45 minutos?.

Solución: suponiendo que los arribos siguen un proceso de Poisson con $\alpha=3$, entonces $\beta=1/3$, con lo que se obtiene:

$$\int_0^{1/12} 3e^{-3x} dx = 1 - e^{-1/4} = 0.221$$

Para el literal b), se tiene un tiempo de llegada de 3/4 de hora, la cual se procede:

$$\int_0^{3/4} 3e^{-3x} dx = 1 - e^{-9/4} = 0.205$$

Ejemplo 2.8

En cierta industria, la proporción de partes instaladas que requieren de

reparaciones en una es una variable aleatoria con distribución beta con $\alpha=3$ y $\beta=2$ (Hines. Et al, 1997). Determine:

- Que porcentaje en promedio requieren reparaciones en un año dado
- La probabilidad de cuando menos más la mitad requieran reparaciones en un año dado.

Solución:

a) Como: $\mu = \frac{\alpha}{\alpha + \beta} = 0.60$ lo que significa que en promedio es el 60% de las instalaciones requieran reparaciones.

b) Como: $\Gamma(\alpha + \beta) = \Gamma(3 + 2) = \Gamma(5) = 4! = 24$

$$\Gamma(\alpha) = \Gamma(3) = 2! = 2 \text{ y}$$

$$\Gamma(\beta) = \Gamma(2) = 1! = 1$$

ahora sustituyendo y calculando: $f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} = 12(1-x)$ para

$0 < x < 1$, como función, entonces:

$$F = \int_0^{1/2} 12(1-x) dx = 5/16$$

Ejemplo 2.9

Se tiene que el ciclo de vida de cierto tipo de baterías de respaldo de emergencia (en horas) es una variable aleatoria X con la distribución de Weibull con $\alpha=0.1$ y $\beta=0.5$ (Hines. Et al, 1997). Determinar:

- El ciclo de vida medio de estas baterías.

- b. La probabilidad de que una batería de este tipo dure más de 300 horas.

Solución:

- a. La media una distribución de Weibull es:

$$\mu = \alpha^{-1/\beta} \Gamma\left(1 + \frac{1}{\beta}\right), \text{ y sustituyéndola, se tiene:}$$

$$\mu = 0.1^{-2} \Gamma(3) = 200 \text{ horas.}$$

- b. Realizando la integración necesaria se tiene:

$$F = \int_{300}^{\alpha} (0.05)x^{-0.5} e^{-0.1x^{1.5}} dx = 0.177$$

Ejemplo 2.10

El monto real de café instantáneo que una maquina de relleno deposita en frascos de "4 onzas" puede considerarse una variable aleatoria con una distribución normal con $\sigma=0.04$ onzas. Si solo 2% de los frascos contienen menos de 4 onzas (Hines. Et al, 1997), ¿cuál deberá ser el relleno medio de esos frascos?.

Solución:

Relación es $Z = \frac{(X - \mu)}{\sigma}$, entonces $F\left(\frac{4 - \mu}{0.04}\right) = 0.02$ y por lo tanto, $F\left(-\frac{4 - \mu}{0.04}\right) =$

0.98, se busca en la tabla del Anexo B, con entrada cercana a 0.98 y se obtiene 0.9798, que corresponde a $z = 2.05$. Así,

$$-\frac{4 - \mu}{0.04} = 2.05, \text{ despejando } \mu, \text{ se tiene que la media es } 4.082 \text{ onzas}$$

Ejemplo 2.11

Después de una gran número de ensayos de contenido de oro en rocas recolectadas en una mina de pozo abierto, un ingeniero en minas postula que el registro (log) natural del contenido de oro (oz/st de oro) sigue una distribución normal con media de -4.6 y varianza de 1.21 (Hines. Et al, 1997), Con esta distribución, ¿sería inusual obtener 0.0015 oz/st de oro o menos en un ensayo?

Solución:

Dado que $\ln(X)$ es la que tiene una distribución normal, la pregunta se refiere al valor estandarizado

$$\frac{\ln(0.0015) - (-4.6)}{\sqrt{1.21}} = -1.729$$

Entonces la probabilidad normal de obtener este valor o menor es:

$$F\left(\frac{\ln(0.0015) - (-4.6)}{\sqrt{1.21}}\right) = 1 - F(-1.729), \text{ que por las tabla del Anexo B se tiene:}$$

$$= 1 - 0.9581 = 0.0419$$

La probabilidad es pequeña de modo que se sospecha que la distribución con media -4.6 no se aplica. Un ensayo con esta reducida cantidad de contenido de oro podría sugerir que el espécimen fue sustraído fuera de la veta.

2.10 EJERCICIOS PROPUESTOS

2.1 En la siguiente tabla se presentan los resultados obtenidos en el laboratorio de Operaciones Unitarias I, para determinar la viscosidad de 3 mezclas problemas. Para ello se toman 8 lecturas de viscosidades obtenidas, estas son determinadas en centipoise (Hines. Et al 1997).

Muestras de viscosidades leídas del viscosímetro, dadas en Centipoise.

Mezcla 1	Mezcla 2	Mezcla 3
22.02	21.49	20.33
23.83	22.67	21.67
26.67	24.62	24.67
25.38	24.18	22.45
25.49	22.78	22.28
23.50	22.56	21.95
25.90	24.46	20.49
24.98	23.79	21.81

Determinar:

- a. La media, mediana y los cuartiles 1, 2 y 3
- b. La varianza desviación típica y el coeficiente de variación
- c. Elaborar el diagrama de cajas o bigotes y parámetros de formas

2.2 En Applied Life Data Analysis (Análisis de Datos Aplicada a la Vida). Presenta el tiempo de falla de un fluido aislante entre electrodos a KV. Los tiempos en minutos, son como sigue (Hines. Et al, 1997): 0.19, 0.78, 0.96, 1.31, 2.78, 3.16, 4.15, 4.67, 4.85, 6.50, 7.35, 8.01, 8.27, 12.06, 31.75, 32.52, 33.91, 36.71 y 72.89.

Determinar:

- a. La elaboración de la distribución de frecuencia y histograma para los datos
- b. La media, mediana, varianza, desviación típica, coeficiente de variación y los cuartiles 1, 2 y 3.
- c. Los parámetros de formas y elaborar el diagrama de bigotes
- d. Concluir sobre la normalidad de los datos.

2.3 Un artículo en Techno metric, presenta los siguientes datos de las trazas de

octanajes de los combustibles de motor de varias mezclas de gasolinas (Hines. Et al 1997), ver la Tabla siguiente:

Datos en porcentaje de octanos para mezclas de gasolinas del mercado.

88.5	87.7	83.4	86.7	87.5	91.5	88.6
94.2	87.8	89.9	88.3	87.6	84.3	86.7
93.2	91.0	90.3	93.4	88.5	90.1	89.2
96.1	93.3	91.8	92.3	90.4	90.1	93.0
88.9	91.2	89.3	94.4	92.7	91.8	91.6
90.4	89.3	89.7	90.3	91.6	90.5	93.7
90.0	90.7					

Determinar:

- a. Resumir los datos en forma tabular
 - b. Los diagramas de frecuencia e histograma.
 - c. La media, mediana, varianza, cuartiles 1, 2, y 3, coeficientes de asimetría y curtosis
 - d. El diagrama de cajas o de bigotes.
 - e. concluya sobre la normalidad de los datos de la muestra.
- 2.4 En una procesadora de alimentos se afirma que cuando más el 10% de sus frascos de café instantáneo contienen menos café que lo declarado en la etiqueta. Probar esta afirmación, 16 frascos de su café instantáneo se selecciona aleatoriamente para pesar su contenido; la afirmación se aceptará si menos de 3 frascos contienen menos café que el declarado en la etiqueta (Hines. Et al, 1997). Determine las probabilidades de que el argumento de la procesadora de alimentos sea aceptado cuando el porcentaje real de sus frascos con menos café que el declarado en la etiqueta sea: a) 5%, b) 10%, c), 15% y d) 20%.
- 2.5 Un ingeniero de control de calidad inspecciona una muestra aleatoria de 3 baterías de cada lote de 24 baterías de automóvil listo para ser embarcado

(Hines. Et al 1997). Si uno de esos lotes contiene seis baterías con ligeros defectos, ¿qué probabilidad hay de que la muestra del inspector contenga

- a. Ninguna batería con defectos
- b. Solo una batería con defectos
- c. Al menos dos baterías con defectos

2.6 Si la probabilidad de que una lámpara fluorescente tenga una vida útil de al menos 500 horas es de 0.85, determine las probabilidades de que entre 20 lámparas de este tipo (Johnson, 1997)

- a. 18 tengan una vida útil de al menos 500 horas
- b. Como mínimo 15 tengan una vida útil de al menos 500 horas
- c. como mínimo dos no tensan una vida útil de al menos 500 horas.

2.7 El número de lecturas erróneas de un instrumento para medir la fuerza de tensión en los empaques, a la semana tiene una distribución de Poisson con $\lambda=0.3$ (Johnson, 1997). ¿Cuál es la probabilidad de que el aparato opere sin lecturas erróneas durante dos semanas consecutivas?

2.8 En ciertos experimentos, el error cometido en la determinación de la solubilidad de una sustancia es una variable aleatoria con la densidad uniforme, con $\alpha=-0.025$ y $\beta= 0.025$. ¿cuáles son las probabilidades de que ese error sea (Johnson, 1997):

- a. De entre 0.010 y 0.015;
- b. De entre -0.012 y 0.012?.



- 2.9 Si una variable aleatoria tiene una distribución log-normal con $\alpha=-1$ y $\beta=2$, determine su media, su desviación estándar y las probabilidades de que la variable aleatoria adopte un valor (Hines. Et al, 1997):
- De entre 3.2 y 8.4;
 - Mayor que 5.0.
- 2.10 Si una variable aleatoria tiene una distribución gamma con $\alpha=2$ y $\beta=2$, determinar la media, desviación estándar y la probabilidad de que la variable aleatoria adopte un valor menor que 4 (Hines. Et al, 1997).
- 2.11 Supóngase que la proporción de unidades defectuosas embarcadas por un distribuidor, que varía un tanto de embarque a embarque (Hines. Et al, 1997), puede considerarse una variable aleatoria con la distribución beta con $\alpha=1$ y $\beta=4$.
- Determine la medida de esta distribución Beta, esto es, la proporción promedio de unidades defectuosas en un embarque de este distribuidor.
 - Determine la probabilidad de que un embarque de este distribuidor contenga 25% o más unidades defectuosas.
- 2.12 Supóngase que el tiempo de falla (en minutos) de ciertas baterías de automóviles sujetas a vibraciones continuas puede considerarse una variable aleatoria con una distribución de Weibull con $\alpha=1/5$ y $\beta=1/5$ (Hines. Et al, 1997).
- ¿Cuánto puede esperarse que dure esta batería?
 - ¿qué probabilidad existe que una batería de este tipo falle en menos de 5 horas?
- 2.13 Con referencia al ejemplo 3.10, compruebe que si la variabilidad de la máquina de relleno se reduce a $\sigma=0.025$ onzas, eso reducirá el monto de café

promedio requerido a 4.05 onzas, aunque el 98% de los frascos seguirá conteniendo más de 4 onzas (Hines. Et al, 1997).

2.14 La resistencia al rompimiento (Newtons) de una tela sintética, denotada por X , se distribuye en $N(800,144)$. El comprador de la tela requiere que esta tenga una resistencia de por lo menos 772 N. Se selecciona al azar y se prueba una muestra de la tela (Hines. Et al, 1997).

2.15 La vida de servicio de un tipo particular de baterías de celda seca se distribuye normalmente con media de 600 días y desviación estándar de 60 días (Hines. Et al, 1997). ¿qué fracción de estas baterías se esperaría que dure mas de 680 días? y ¿qué fracción se esperaría que fallara antes de 550 días?.

2.16 La experiencia indica que el tiempo de revelado para un papel de impresión fotográfica se distribuye como $X \sim N(30 \text{ seg}, 1.21 \text{ seg}^2)$. Determine :a) la probabilidad de que X sea al menos 28.5 segundos (Hines. Et al, 1997), c) la probabilidad de que X sea alo mas 31 segundos c) la probabilidad de que X difiera de su valor esperado en mas de 2 segundos.

2.17 Un gerente de planta ordena interrumpir un proceso y efectuar un ajuste de lecturas siempre que el pH del producto final sea mayor que 7.20 o menor de 6.8. el pH de muestra se distribuye normalmente con μ desconocida y desviación estándar $\sigma = 0.10$ (Hines. Et al, 1997). Determine las siguientes probabilidades:

- a. De que se realice el reajuste cuando el proceso opere como se propuso con $\mu = 7.0$
- b. De que se realice el reajuste cuando el proceso se desvié ligeramente de lo planeado con el pH medio 7.05.

- c. De que falle el reajuste cuando el proceso sea demasiado alcalino y el pH medio sea $\mu=7.25$.
- d. De que falle el reajuste sea demasiado ácido y el pH medio sea $\mu=6.75$.

CAPÍTULO 3

MUESTRAS ALEATORIAS Y DISTRIBUCION DE MUESTREO

Si una población es infinita es imposible observar todos los valores, e incluso si es finita, puede resultar impráctico o antieconómico observarla en su totalidad. Por lo tanto, con frecuencia es necesario emplear una muestra, es decir una parte de una población, e inferir de ella resultados que correspondan a la población entera. Evidentemente, estos resultados sólo serán útiles si la muestra es en cierto modo "representativa" de la población (Hines, Et al, 1997). Por ejemplo, sería poco razonable esperar mediciones útiles acerca de la contaminación industrial en el área San Salvador, cuando se basan en medir la contaminación al final de las plantas de tratamiento de desechos; cuando esta trata una fracción del flujo de contaminantes generada en los procesos industriales. Por lo tanto las mediciones de contaminantes realizadas después de la planta de tratamiento no reflejará la magnitud de contaminantes generados por las industrias. Para garantizar que una muestra sea representativa de la población de la cual pertenece, así como para establecer una objetividad de los resultados para la aplicación de la probabilidad, las muestras deben ser aleatorias.

Hay muchas maneras de tomar una muestra de una población. Además, las conclusiones que pueden extraerse acerca de la población dependen a menudo de cómo se seleccionó la muestra. Por lo general, deseamos que la muestra sea representativa de la población. Un método importante para la selección de muestras es el muestreo aleatorio. La mayor parte de las técnicas estadísticas que presentará en el presente trabajo se suponen muestras aleatorias. En este capítulo se definirá una muestra aleatoria y se presentará varias distribuciones de probabilidad útiles en el análisis de la información de datos muestreados (Hines, Et al, 1997).

3.1 MUESTRAS ALEATORIAS

Para que los resultados obtenidos de las muestras en las investigaciones tecnológicas y científicas dentro del campo de Ingeniería Química e Ingeniería de Alimentos, sean objetivos y representativos de la población tomada; las muestras deben ser aleatorias

Una muestra aleatoria se define: que X sea una variable aleatoria con distribución de probabilidad $f(x)$. Entonces el conjunto de n observaciones X_1, X_2, \dots, X_n , tomadas en la variable aleatoria X , y con resultados numéricos x_1, x_2, \dots, x_n , se llama una muestra aleatoria si las observaciones se obtienen observando X de manera independiente bajo condiciones invariables por n veces (Hines. Et al, 1997).

El muestreo aleatorio se clasifica según el tamaño de la población de donde provenga, y éstos son: poblaciones finitas y poblaciones infinitas.

Para muestreo de poblaciones finitas se define: " *un conjunto de observaciones x_1, x_2, \dots, x_n , constituye una muestra aleatoria de tamaño n de una población finita de tamaño N , si se elige de tal forma que cada subconjunto de n de los N elementos de la población tenga la misma probabilidad de ser seleccionados*". Y, para poblaciones infinitas se define de la siguiente manera: *un conjunto de observaciones x_1, x_2, \dots, x_n , constituye una muestra aleatoria de tamaño n de una población infinita $f(x)$, si:*

1. *Cada x_i es un valor de una variable aleatoria tiene la misma distribución de probabilidad $f(x)$, y*
2. *Estas n variables aleatorias son independientes.*

No siempre es fácil obtener una muestra aleatoria. Algunas veces podemos utilizar tablas de números aleatorios uniformes. Para emplear este enfoque para una población finita se utiliza el Anexo I. Este método es en particular útil cuando se tiene un tamaño de población pequeña. En otras ocasiones, el investigador técnico o científico no puede utilizar fácilmente procedimientos formales para ayudar a asegurar la aleatoriedad y deben confiar en otros métodos de selección. Para tal caso se tiene (Hines. Et al, 1997):

*Se desea tomar una muestra aleatoria de 5 lotes de materia prima de 25 sin procesar. Se numera los lotes con los enteros del 1 al 25. Después de esto se elige arbitrariamente del Anexo I, un renglón y una columna como punto de partida. Se lee hacia abajo la columna elegida, la cual debe tener 2 dígitos hasta que se encuentren 5 números aceptables (un número aceptable se encuentra entre 1 y 25). Considerando el proceso anterior brinda una secuencia de números que se lee, 37, 48, 55, **02**, **17**, 61, 70, 43, **21**, 82, 73, **13**, 60, **25**. Los números en negritas especifican qué lotes de materia prima se van a elegir como muestra aleatorias.*

3.2 ESTADÍSTICOS

El propósito de la mayoría de las investigaciones en el campo de la Ingeniería Química y de Alimentos es hacer uso de las técnicas estadísticas para generalizar a partir de la información contenida en muestras aleatorias acerca de la población de la cual se extraen. En particular, a menudo se interesa en hacer inferencias sobre los parámetros de poblaciones, como la media μ o la desviación estándar σ . Al hacer tales inferencias, se usa la estadísticos como \bar{x} y s , esto es, cantidades calculadas con base a observaciones muestrales.

Una estadística ó estadístico es cualquier función de las observaciones en una muestra aleatoria que no depende de parámetros desconocidos (Hines. Et al,

1997). Por ejemplo, si X_1, X_2, \dots, X_n es una muestra aleatoria de tamaño n , para el caso, la media de la muestra \bar{x} , la varianza de muestra s^2 , y la desviación estándar s son estadísticas. Nótese que puesto que una estadística es una función de los datos a partir de una muestra aleatoria, ella misma es también una variable aleatoria.

3.2.1 MEDIA MUESTRAL, VARIANZA MUESTRAL Y DESVIACIÓN ESTÁNDAR MUESTRAL

Como se mencionó anteriormente la media muestral, varianza muestral y desviación estándar muestral, son ejemplos de medidas estadísticas calculados de muestras aleatorias (Hines. Et al, 1997), entonces la media muestral es:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad 3.1$$

la varianza muestral es:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad 3.2$$

y la desviación estándar muestral es:

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad 3.3$$

3.3 DISTRIBUCIÓN DE MUESTREO

Para todo proceso que se desee extraer conclusiones en torno a poblaciones basadas en los datos de las muestras se utiliza en forma considerable las estadísticas. Los procedimientos requieren que se entiendan el comportamiento probabilístico de ciertas estadísticas. Para ello se hace uso a lo que se llamará

distribución de muestreo. Y se entenderá como: la distribución de probabilidad de una estadística (Hines. Et al, 1997).

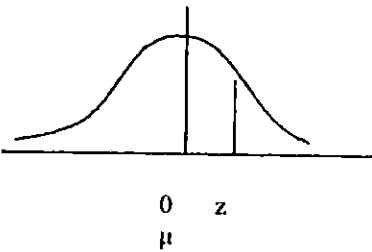
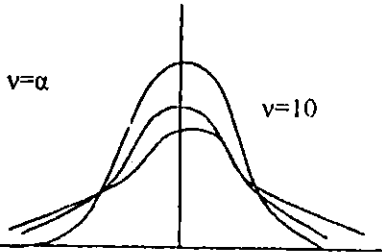
Se encuentran varias distribuciones de muestreo de uso importante en el campo de la ingeniería que se utilizará de manera extensiva posteriormente. Según la naturaleza de las variables aleatorias de donde provengan las muestras, estas pueden ser continuas o discretas. Para la distribución de muestreo de variables continuas en la tabla 3.1 se define y resume brevemente. Siendo las más importantes las siguientes:

- a. Distribución Normal $N(0,1)$
- b. Distribución t – student (t)
- c. Distribución Gi – dos (X^2)
- d. Distribución F de Snedecor (F)

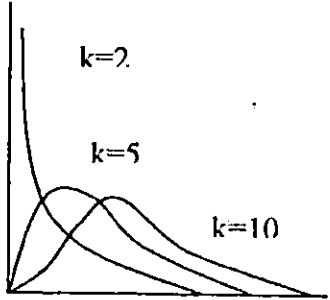
En la tabla 3.2 se define e ilustra brevemente la distribución de muestreo para variables aleatorias discretas. Siendo las más importantes las siguientes:

- a. Distribución de muestreo de Bernoulli (binomial)
- b. Distribución de muestreo de Poisson.

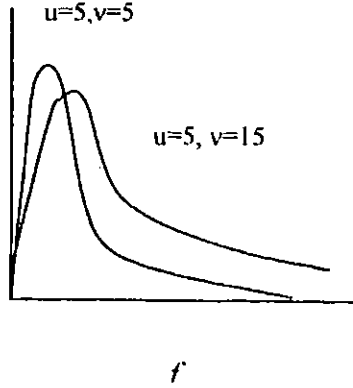
Tabla 3.1. Descripción breve de las distribuciones de muestreo a partir de poblaciones normales (variables aleatorias continuas).

DISTRIBUCIÓN	VARIABLE	ESTADÍSTICA	CARACTERÍSTICAS	FORMA
Normal N(0,1)	$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	\bar{X} Media muestral	<ul style="list-style-type: none"> Si la población es normal, entonces la distribución de muestreo de \bar{X} es normal. Con una $\mu=0$ y $\sigma=1$. Si la población no la rige una distribución normal de muestreo de \bar{X}, entonces será aproximadamente normal para muestras grandes (por el teorema central de limite), lo anterior se cumple para $n \geq 25$. 	
t - student	$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$	\bar{X} Media muestral para n muestras pequeñas	<ul style="list-style-type: none"> Tiene forma acampanada y es perfectamente simétrica y unimodales, y el valor máximo se alcanza cuando $\mu=0$ y $\sigma^2 = \nu/(\nu-2)$, para $\nu \geq 2$, a diferencia de Z que es muy sensible, alargándose con rapidez hacia la izquierda y la derecha. La distribución t tiene colas más pesadas que la normal; esto debido que tiene mayor probabilidad hacia fuera. El estadístico t de una muestra de tamaño n extraída de una población normal de media μ sigue una distribución t-student con $n-1$ grados de libertad. En realidad se aproxima a la distribución normal estándar $\nu=\alpha$. Observando que los valores de t para 29 o mas grado de libertad se acercan a los correspondientes valores de la distribución normal (Z). En conclusión la distribución normal estándar ofrece una adecuada aproximación a la distribución t para muestras de tamaño de 30 o mas. 	

Continuación de la Tabla 3.1.

<p>Gi - dos (χ^2)</p>	$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$	<p>S^2 Varianza muestral</p>	<ul style="list-style-type: none"> ▪ La forma de la distribución χ^2 variará con los grados de libertad asociados con la varianza. ▪ Es la forma de distribución de muestreo de la varianza de una muestra de distribución normal. Se puede construir estimación de intervalo de confianza e hipótesis estadística de prueba alrededor de la varianza de una población normal. ▪ Nótese que la variable aleatoria ji cuadrada es no negativa, y que la distribución de probabilidad es asimétrica hacia la derecha. Sin embargo, a medida que v aumenta la distribución simétrica. Cuando $v \rightarrow \infty$, la distribución ji cuadrada es la distribución normal. Se sigue una distribución ji cuadrada, si X_1, \dots, X_n son n variables aleatorias $N(0,1)$ independientes $Y = X_1^2 + \dots + X_n^2$. 	
---------------------------------------	--------------------------------------	---	---	---

Continuación de la Tabla 3.1.

<p>F de Snedecor</p>	$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2}$	<p>Razón de varianza de muestras aleatorias normales independientes</p>	<ul style="list-style-type: none"> ▪ Siendo X_1 y X_2 dos variables aleatorias ji cuadrada independiente con u y v grados de libertad respectivamente. la razón entre ellas divididas cada una por su grado de libertad se distribuye como F Snedecor. ▪ La función de densidad es asimétrica hacia la derecha cuando aumenta los u grados de libertad numerador y v de denominador ▪ La distribución F se usa con el objeto de comparar la variabilidad de dos fuentes diferentes independientes. $S_1^2 / \sigma_1^2 \rightarrow n_1 - 1$, para la fuente 1; $S_2^2 / \sigma_2^2 \rightarrow n_2 - 1$, para la fuente 2. 	
----------------------	---	---	---	---

Nota: se define el numero de grado libertad (v) como una suma de cuadrados ($SS = \sum(X_i - \bar{X})^2$) como el numero de términos independientes. Estos elementos no son todos independientes ya que $\sum(X_i - \bar{X}) = 0$ siendo independiente solo $(n-1)$ de ellos. Donde n es el tamaño de muestra. Los grados de libertad (v)= $n-1$; cuando se tienen dos poblaciones diferentes los grados de libertad para la fuente 1, está dada por $v=n$

Tabla 3.2. Descripción breve de las distribuciones de muestreo a partir de una función de distribución (variables aleatorias discretas).

DISTRIBUCIÓN	VARIABLE	ESTADÍSTICA	CARACTERÍSTICAS
Bernoulli (Binomial)	$\mu_x = p$ y $\sigma_x^2 = \frac{p(1-p)}{n}$	Media y varianza de \bar{X}	<ul style="list-style-type: none"> • Cuando $x=1$, se le llama éxito y cuando $x=0$ se le llama fracaso. Por lo tanto x_i es 0 o 1. • p es la probabilidad de éxito y n observaciones • es una variable aleatoria discreta que puede tomar los valores $(0, 1/n, 2/n, \dots, (n-1)/n, 1)$. • P se le llama a menudo, fracción muestral de disconformidades
Poisson	$\mu_x = \lambda$ y $\sigma_x^2 = \frac{\lambda}{n}$	Media y varianza de \bar{X} (procedimientos de demérito)	<ul style="list-style-type: none"> • El parámetro λ es la tasa de ocurrencia media. Y n el tamaño de la muestra aleatoria tomada. • Generalmente la suma de n variables aleatorias independientes de Poisson tienen una distribución de Poisson, con parámetro igual a la suma de los parámetros individuales de Poisson. Tomando valores $(0, 1/n, 2/n, \dots)$. • Una variable aleatoria con esta distribución es útil en casos en que una unidad puede tener n diferentes defectos o disconformidades, llamándola procedimientos de demérito

3.4 EJERCICIOS DEL CAPÍTULO

Ejemplo 3.1

Un fabricante de torres de enfriamiento asegura que con una sobre carga de agua del 20%, la torre se inunda en un promedio en 12.40 minutos en promedio. Para probar esta afirmación, una muestra de 20 torres de enfriamiento fueron sometidas a una sobre carga del 20%, y los tiempos que tardaron en inundarse tubo un promedio de 10.63 minutos y una desviación estándar de 2.48 minutos. Si puede suponerse que los datos constituyen una muestra aleatoria de una población normal, ¿Tenderían estos apoyar o refutar la afirmación del fabricante?.

Solución: primero se calcula t del problema, de la manera

Datos:

$$\bar{X} = 10.63 \text{ minutos}$$

$$\mu = 12.40 \text{ minutos}$$

$$S = 2.48 \text{ minutos}$$

$$n = 20$$

Utilizando la distribución t student (Tabla 3.1), sustituyéndola y calculando se tiene:

$$t = \frac{\bar{X} - \mu}{S / \sqrt{n}} = \frac{10.63 - 12.40}{2.48 / \sqrt{20}} = -3.19$$

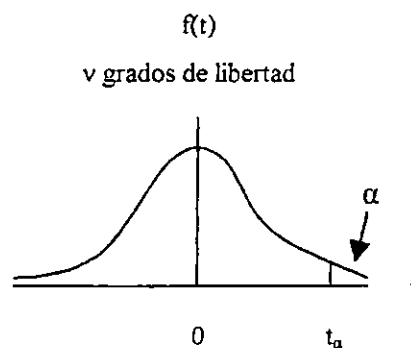


Figura 3.1. Valores tabulados de t

Ahora con $v=20-1=19$ grados de libertad y con un nivel de significancia (α) de 0.005 se lee en la tabla del Anexo E el valor t_{α} , siendo este de 2.861, por consecuencia la probabilidad de que t sea menor que -2.861 también es de 0.005. Dado que $t=-3.19$ es menor que -2.861 y 0.005 es una probabilidad muy reducida, se concluye que los datos tienden a refutar la afirmación del fabricante. Con toda probabilidad, el tiempo de inundación medio de estas torres de enfriamiento con una sobrecarga del 20% es inferior a los 12.40 minutos.

Ejemplo 3.2

Una empresa de análisis de aguas, compra agua desmineralizada, y sabe por experiencia que la varianza micromhos es de $1.26 \cdot 10^{-6}$ mmhos². Como es importante que el agua tenga la misma concentración de minerales, la empresa rechaza un pedido si la varianza de las 20 muestras del lote pedido seleccionado al azar excede de $2.00 \cdot 10^{-4}$ mmhos² (Johnson, 1997). Suponiendo que los valores de la muestra pueden ser considerados como una muestra aleatoria de una población normal, ¿cuál es la probabilidad de que el lote sea rechazado aunque $\sigma^2=1.26 \cdot 10^{-6}$ mmhos²?

Solución:

Los grados de libertad es: $v=20-1=19$

Al sustituir en la fórmula de la estadística ji cuadrada, se tiene

$$\chi^2 = \frac{19(2.00 \times 10^{-4})}{1.26 \times 10^{-6}} = 30.2$$

y luego se determina en la tabla del Anexo D χ^2_{α} , con 19 grados de libertad y 0.05 como nivel de significancia, resultando $\chi^2_{0.05} = 30.1$. Así, la probabilidad de que el lote sea rechazado es menos del 0.05 o menor del 5%.

Ejemplo 3.3

Si dos muestras aleatorias independientes de tamaño $n_1=7$ y $n_2=13$ se toman de una población normal (Johnson, 1997), ¿cuál es la probabilidad de que la varianza de la primera muestra sea al menos tres veces mayor que la segunda muestra, asumiendo un nivel de significancia del 95%?

Solución:

Cálculo de los grados de libertad para ambas muestras:

Muestra 1:

$$v_1=n_1-1=7-1=6$$

Muestra 2:

$$v_2=n_2-1=13-1=12$$

En la tabla del Anexo F se encuentra $F_{0,05}=3.00$ para los grados de libertad de las muestras calculadas; así la probabilidad calculada es de 0.05 para un nivel de significancia del 95%. Entonces:

$$F_{0,95}(6,12) = \frac{1}{F_{0,05}(12,6)} = \frac{1}{3.00} = 0.3333 \text{ para la cola izquierda}$$

3.5 EJERCICIOS PROPUESTOS

- 3.1 Un proceso para la fabricación de jabones se halla bajo el control si los pesos de los jabones tienen una media de 0.5000 lbs. Que se puede decir de este proceso si una muestra de 10 de esos jabones tiene un peso medio de 0.5060 lbs. Y una desviación estándar de 0.0040 lbs. (Hines. Et al, 1997).
- 3.2 Un proceso de tratamientos de aguas embotellas de un litro el contenido de calcio (Ca) como reglamento de consumo humano debe tener como media de 50 ppm. Se tomo una muestra de 15 botellas de agua para su análisis y se

obtuvo una media de 51 ppm y una desviación estándar de 6 ppm (Johnson, 1997). ¿qué probabilidad tiene el lote para ser apta a consumo humano?

3.3 Un comprador de café instantáneo de 4 onzas por bolsa de contenido acepta un lote con una varianza de 0.4 onzas², si una muestra de tamaño de 15 bolsa tiene una varianza de 0.45 onzas² (Hines. Et al, 1997). Asumiendo que es una población normal. ¿Cuál será la probabilidad de aceptar el lote?

3.4 Para el proceso descrito del ejercicio 3.2 se exige una varianza de 8 ppm, si una muestra de 20 jabones da como resultados una varianza de 6 ppm (Johnson, 1997). ¿cuál será la probabilidad de aceptar el lote, sumiendo una población aleatoria con distribución normal?

3.5 Obtenga los siguientes valores empleando la tabla del Apéndice F

- a. $F_{0.25,4,9}$
- b. $F_{0.05,15,10}$
- c. $F_{0.95,6,8}$
- d. $F_{0.90,24,24}$

3.6 Si muestras aleatorias independientes de tamaño $n_1=n_2=8$ proceden de poblaciones normales con la misma varianza (Johnson, 1997), ¿cuál es la probabilidad de que cualquiera de las varianzas muestrales sea al menos siete veces mayor que la otra?

CAPÍTULO 4

INFERENCIA ESTADÍSTICA

Aunque la inferencia estadística abarca el muestreo y distribución de muestreo, estas ya fueron objeto estudio en el capítulo 3.

Sobre los diseños muestrales se estudiaron los conceptos de estadístico. Siendo un estadístico (estadística), como la medida estadística calculada a partir de una muestra aleatoria, y como parámetro, la medida estadística calculada a partir de la población.

La inferencia estadística es el proceso mediante el cual se utiliza la información de los datos de una muestra para extraer o inferir conclusiones acerca de la población de la que se seleccionó la muestra. Las técnicas de la inferencia estadística pueden dividirse en dos áreas principales: *estimación de parámetros* y *pruebas de hipótesis*.

La diferencia entre ambos es que los problemas de *estimación de parámetros* se debe elegir un valor que describa la variabilidad de las muestras con respecto a la población proveniente, mientras que en las pruebas de hipótesis se debe decidir entre aceptar o rechazar un valor o un conjunto de valores especificados de un parámetro.

5.1 ESTIMACION DE PARÁMETROS.

La estimación es un proceso para usar un estadístico (estadística) muestral en orden a estimar el correspondiente parámetro poblacional desconocido. Como un ejemplo del problema de la estimación por parámetros, supóngase que ingenieros químicos están analizando el contenido de impurezas en materias primas.

Existiendo en la realidad una variabilidad entre cada zona individual de la población. En consecuencia, el ingeniero está interesado en estimar la impureza en la materia prima promedio para la población. Es posible también que le interese estimar la variabilidad de las impurezas en la materia prima. A partir de cálculos de los estadísticos de muestras representativas de la población proveniente. Los procedimientos de estimación en general, se dividen, en:

- a. Estimación puntual de parámetros
- b. Estimación de intervalo de parámetros o intervalos de confianza

4.1.1 ESTIMACIÓN PUNTUAL

La estimación por puntos de un parámetro de población, es un solo valor numérico de una estadística que corresponde a ese parámetro. Esto es, la estimación puntual es una selección única para el valor de un parámetro desconocido. En forma más precisa, si x es una variable aleatoria con distribución de probabilidad $f(x)$, caracterizada por el parámetro desconocido θ , y si X_1, X_2, \dots, X_n es una muestra aleatoria de tamaño n de x , entonces la estadística $\hat{\theta} = h(X_1, X_2, \dots, X_n)$ correspondiente a θ ; que se le llama estimador de θ . Nótese que la estimación $\hat{\theta}$ es una variable aleatoria, porque es una función de los datos de muestreo. Después de que la muestra se ha seleccionado, $\hat{\theta}$ toma un valor numérico particular llamado estimación por puntos o puntuales de θ (Hines. Et al, 1997).

En la Tabla 4.1 (Hines. Et al, 1997), se muestrann las estimaciones que con frecuencia se necesita en ingeniería y sus respectivas estimaciones puntuales razonables de estos parámetros.

Un parámetro puede tener varios estimadores puntuales diferentes. Como por ejemplo: si se desea estimar la media de una variable aleatoria, se podría considerar la media de la muestra, la mediana de la muestra, o quizás el promedio de las observaciones extremas más pequeña y mas grande en la muestra como estimadores puntuales. Para decidir cuál es el mejor estimador puntual que se puede usarse de un parámetro particular, se necesita examinar sus propiedades estadísticas y desarrollar algunos estimadores comparativos.

Tabla 4.1. Los parámetros que con frecuencias se necesita estimar, con sus respectivos estimadores.

PARÁMETRO DESCONOCIDO (θ)	ESTIMACIÓN ($\hat{\theta}$)	OBSERVACION
Media de una sola población (μ)	$\hat{\mu} = \bar{X}$	La estimación es la media de la muestra.
Varianza (σ^2) o desviación estándar (σ) de una sola muestra	$\hat{\sigma}^2 = S^2$	La estimación es la varianza y desviación estándar de la muestra.
Proporción p de artículos en una población que pertenece a la clase de interés.	$\hat{p} = X/n$	La estimación es la proporción de la muestra, donde X es el número de objetos en una muestra aleatoria de tamaño n que pertenece a la clase de interés.
Diferencia de media entre dos poblaciones ($\mu_1 - \mu_2$)	$\hat{\mu}_1 - \hat{\mu}_2 = \bar{X}_1 - \bar{X}_2$	La estimación es la diferencia entre las medias de las muestras de dos muestras aleatorias independientes.
Diferencia en dos proporciones de población, $p_1 - p_2$.	$\hat{p}_1 - \hat{p}_2$	La estimación es la diferencia entre dos proporciones de muestra calculadas a partir de dos muestras aleatorias independientes.

4.1.1.1 Propiedades de los Estimadores

Una propiedad deseable de un estimador es: que debe estar "cerca" en cierto sentido al valor verdadero del parámetro desconocido. Formalmente se dice que $\hat{\theta}$ es un estimador *neutral de parámetro* θ cuando (Hines. Et al, 1997):

$$E(\hat{\theta}) = \theta \quad 4.1$$

Esto es, $\hat{\theta}$ es un estimador neutral de θ si "en promedio" sus valores son iguales a θ . Nótese que esto equivale a requerir que la medida de la distribución de la muestra de $\hat{\theta}$ sea igual a θ .

En consecuencia, la media de la muestra \bar{X} es un estimador neutral de la media de la población μ . Por tanto, la varianza de la muestra S^2 es un estimador neutral de la varianza σ^2 . Sin embargo, la desviación estándar S de la muestra es un estimador sesgado de la desviación estándar σ de la población. En el caso de muestras grandes este sesgo es despreciable.

El error cuadrático medio estimador $\hat{\theta}$ se define como (Hines. Et al, 1997):

$$ECM(\hat{\theta}) = E\left(\hat{\theta} - \theta\right)^2 \quad 4.2$$

El error en el cuadrático medio se puede describirse de la siguiente manera:

$$ECM(\hat{\theta}) = E\left[\hat{\theta} - E(\hat{\theta})\right]^2 + \left[\theta - E(\hat{\theta})\right]^2 = V(\hat{\theta}) + (\text{sesgo})^2 \quad 4.3$$

Esto es, el error cuadrático medio de $\hat{\theta}$ es igual a la varianza del estimador más el sesgo al cuadrado. Si $\hat{\theta}$ es un estimador neutral de θ , el error cuadrático medio de $\hat{\theta}$ es igual θ .

El error cuadrático medio es un criterio importante para comparar dos estimadores. Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores del parámetro θ , y $ECM(\hat{\theta}_1)$ y $ECM(\hat{\theta}_2)$ los errores cuadráticos medio de $\hat{\theta}_1$ y $\hat{\theta}_2$. Entonces la eficiencia relativa de $\hat{\theta}_2$ a $\hat{\theta}_1$ se define de la siguiente manera:

$$\frac{ECM(\hat{\theta}_1)}{ECM(\hat{\theta}_2)} \quad 4.4$$

Si esta eficiencia relativa es menor que uno, se concluye que $\hat{\theta}_1$ es más eficiente de θ que $\hat{\theta}_2$, en el sentido que tiene un error cuadrático medio más pequeño.

Obsérvese que el estimador $\hat{\theta}$ es neutral para que θ , el error cuadrático medio se reduce la varianza del estimador $\hat{\theta}$. Dentro la clase de estimadores neutrales se pretende buscar el estimador que tiene la varianza más pequeña. Este se llama estimador neutral de la varianza mínima. La figura 4.1 (Hines. Et al, 1997) muestra la distribución de probabilidad de dos estimadores neutrales $\hat{\theta}_1$ y $\hat{\theta}_2$, teniendo $\hat{\theta}_1$ una varianza más pequeña que $\hat{\theta}_2$. El estimador $\hat{\theta}_1$ producirá con mayor probabilidad que $\hat{\theta}_2$ una estimación más cercana al valor verdadero del parámetro desconocido θ .

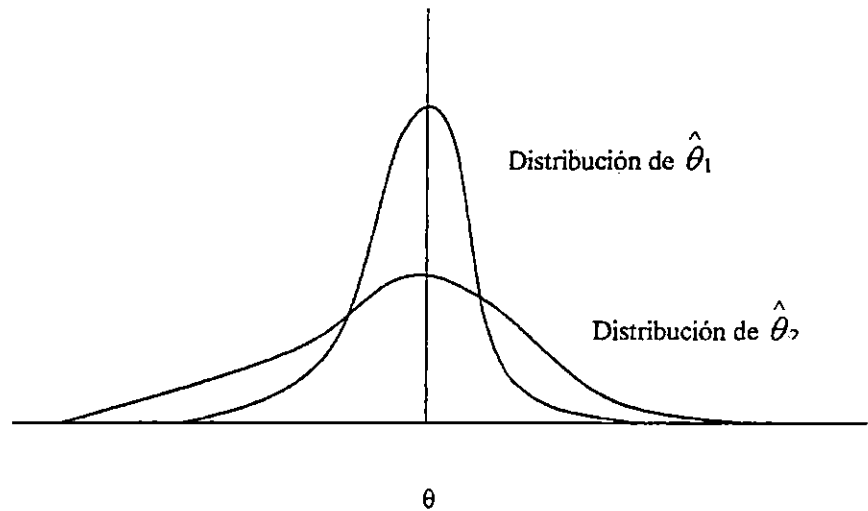


Figura 4.1. distribución de probabilidad de dos estimadores neutrales $\hat{\theta}_1$ y $\hat{\theta}_2$.

Es posible obtener una cota inferior en la varianza de todos los estimadores neutrales θ . Sea $\hat{\theta}$ un estimador neutral del parámetro θ , basado en una muestra aleatoria de n observaciones y considere que en $f(x, \theta)$ denota la distribución de probabilidad de la variable aleatoria X . Entonces una cota inferior en la varianza¹.

$$v(\hat{\theta}) \geq \frac{1}{nE\left[\frac{d}{d\theta} \ln f(X, \theta)\right]^2} \quad 4.5$$

Esta desigualdad se denomina cota inferior de Cramér-Rao. Si un estimador neutral $\hat{\theta}$ satisface la ecuación 4.5 con desigualdad, se tratará del estimador de *varianza mínima* de θ (Hines. Et al, 1997).

¹ Se requieren ciertas condiciones en la función $f(\bar{X}, \theta)$ al obtener la desigualdad Cramér-Rao. Las mayores partes de las distribuciones de probabilidad estándar satisfacen estas condiciones.

Se encuentra algunas veces que algunos estimadores sesgados son preferibles a los neutrales por ellos tienen un error cuadrático medio más pequeño. Esto es, se puede reducir la varianza del estimador de manera considerable introduciendo una cantidad relativamente pequeña de sesgo. En tanto que la reducción de la varianza sea mayor que el sesgo al cuadrado, se obtendrá un estimador mejorado en el sentido del error cuadrático medio. Por ejemplo la figura 4.2 muestra la distribución de probabilidad de un estimador sesgado $\hat{\theta}_1$ con varianza menor que el estimador neutral $\hat{\theta}_2$. Sería más probable que una estimación basada en $\hat{\theta}_1$ estuviera más cerca del valor verdadero θ que una basada en $\hat{\theta}_2$, tal como se muestra en la figura 4.2.

Un estimador $\hat{\theta}^*$ que tiene un error cuadrático medio que es menor o igual al error cuadrático medio de cualquier otro estimador $\hat{\theta}$, para todos los valores del parámetro θ , se llama estimador óptimo de θ .

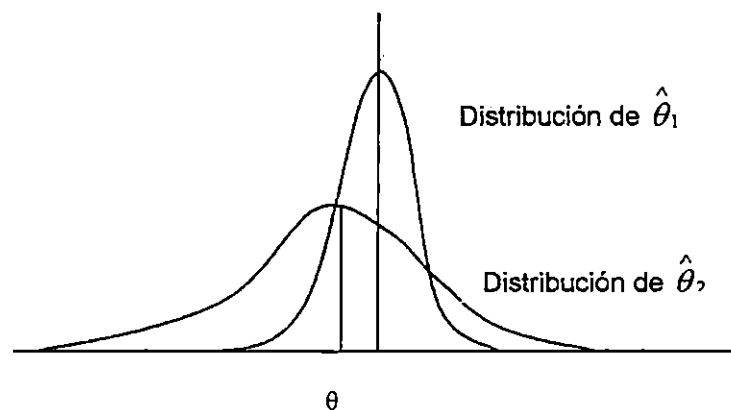


Figura 4.2. Estimador sesgado $\hat{\theta}_1$ que tiene varianza menor que un estimador neutral $\hat{\theta}_2$



Otra manera de definir la cercanía de un estimador $\hat{\theta}$ a un parámetro θ se da en término de la *consistencia*. Si $\hat{\theta}_n$ es un estimador de θ basado en una muestra aleatoria de tamaño n , se dice que $\hat{\theta}_n$ es consistente para θ si:

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1 \quad 4.6$$

La consistencia es una propiedad de muestras grandes, puesto que describe el comportamiento en el límite del estimador $\hat{\theta}$ conforme el tamaño de la muestra tiende a infinito.

4.1.1.2 Método de Máxima Similitud.

Uno de los mejores métodos para obtener un estimador puntual es el de máxima similitud. Sea X una variable aleatoria con distribución de probabilidad $f(x, \theta)$, donde θ es un parámetro desconocido único. Sean X_1, X_2, \dots, X_n los valores observados en una muestra aleatoria de tamaño n (Hines. Et al, 1997). Entonces la función de probabilidad de la muestra es:

$$L(\theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \dots \cdot f(x_n, \theta)$$

La función de probabilidad es ahora función únicamente del parámetro desconocido θ . El estimador de máxima similitud de θ es el valor que maximiza la función de probabilidad $L(\theta)$. En esencia, el estimador de máxima similitud es el valor de θ que maximiza la probabilidad de ocurrencia de los resultados de la muestra. En la tabla 4.2 muestran los estimadores que maximiza algunas funciones de probabilidades importantes; para distribuciones normales y para distribuciones de las series de Bernoulli.

TABLA 4.2. Estimadores que maximiza las funciones de probabilidades normales y de las series de Bernoulli.

DISTRIBUCIÓN	FUNCIÓN DE PROBABILIDAD	ESTIMADOR DE MÁXIMA SIMILITUD	OBSERVACIONES
Bernoulli	$p(x) = p^x(1-p)^{1-x}$ $x=1,0$	$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$	\hat{p} es el estimador de máxima similitud de p es \hat{p} , para un tamaño de muestra n .
Normal con media y varianza conocida	$L(\mu) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/2\sigma^2}$	$\hat{\mu} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$	La media de la muestra es el estimador de máxima similitud.
Uniforme con intervalo de 0 a a	$L(a) = \prod_{i=1}^n \frac{1}{a} = \frac{1}{a^n}$	\hat{a}	La pendiente de esta función no es cero en todas partes y la función de probabilidad aumenta cuando a disminuye. Entonces se maximiza cuando se fija \hat{a} con el valor más pequeño que sea razonable (a no puede ser más pequeño que el valor más grande de la muestra).
Normal con media y varianza desconocida.	$L(\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)\sum_{i=1}^n (x_i-\mu)^2}$	$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$ $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$	Los estimadores de máxima similitud son la media y varianza de la muestra

Los estimadores de máxima similitud no son necesariamente insesgado, pero es usual que puedan modificarse con facilidad para hacerlos insesgados. El sesgo se aproxima a cero para muestras grandes. En general, los estimadores de máxima similitud tienen buenas propiedades de muestras grandes llamadas *asintóticas*. Específicamente, se distribuyen en forma asintótica, son insesgados y tienen una varianza que se aproxima a la cota inferior de Cramér-Rao para n grande. Los estimadores de máxima similitud también son consistentes. Además, tienen la propiedad de invarianza.

Cuando se presenta un valor de una estimación puntual, suele ser necesario dar una idea de su precisión. El *error estándar* es la medida usual de precisión que se emplea. Si $\hat{\theta}$ es un estimador de θ , entonces el *error estándar* de $\hat{\theta}$ es justamente la desviación estándar de $\hat{\theta}$, es:

$$\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})} \quad 4.7$$

Si $\sigma_{\hat{\theta}}$ involucra cualesquiera parámetro desconocido, entonces se sustituye estimaciones de estos parámetros en la ecuación 4.7, se obtiene el *error estándar estimado* de $\hat{\theta}$, se dice $\hat{\sigma}_{\hat{\theta}}$. Un error estándar pequeño implica que se ha presentado una estimación relativamente precisa.

4.1.2 ESTIMACIÓN DEL INTERVALO DE CONFIANZA

En muchas situaciones, una estimación puntual no proporciona suficiente información acerca del parámetro de interés. Por ejemplo, si nos interesa estimar la resistencia térmica media al calor del hierro, un solo número puede tener mucho significado. Una estimación de intervalo de la forma $L \leq \mu \leq U$ podría resultar más útil.

Los puntos extremos de este intervalo serán variables aleatorias, puesto que son funciones de datos de muestras (Hines. Et al, 1997).

En general, para construir un estimador de intervalo del parámetro desconocido, se debe encontrar dos estadísticas (estadísticos) L y U tal como sigue:

$$P\{L \leq \theta \leq U\} = 1 - \alpha \quad 4.8$$

El intervalo resultante es:

$$L \leq \theta \leq U \quad 4.9$$

se llama *intervalo de confianza* del $100(1-\alpha)$ por ciento del parámetro desconocido θ . A L y U se les denomina *límites de confianza* inferior y superior, respectivamente, y $(1-\alpha)$ recibe el nombre de *coeficiente de confianza*. La interpretación del intervalo de confianza es que si se coleccionan muchas muestras aleatorias y se calcula un intervalo de confianza de $100(1-\alpha)$ por ciento en θ de cada muestra, entonces $100(1-\alpha)$ por ciento de estos intervalos contendrán el valor verdadero de θ .

El intervalo de confianza en la ecuación 4.9 podría llamarse con mayor propiedad un *intervalo de confianza de dos lados*, en cuanto a que especifican tanto un límite inferior como uno superior en θ . En ocasiones, el intervalo de confianza de un solo lado podría ser más apropiado. Un intervalo de confianza inferior de $100(1-\alpha)$ por ciento de un lado en θ está dado por el intervalo siguiente:

$$L \leq \theta \quad 4.10$$

donde el límite de confianza inferior L se elige de la siguiente manera:

$$P\{L \leq \theta\} = 1 - \alpha \quad 4.11$$

De manera similar, para un intervalo de confianza superior de $100(1-\alpha)$ por ciento de un lado superior de θ .

La longitud de un intervalo de confianza observado es una medida importante de la calidad de la información obtenida de la muestra. La longitud de medio intervalo $(\theta-L)$ o $(U-\theta)$ se denomina la *precisión* del estimador. Cuanto mayor sea el intervalo de confianza, tanto mayor confianza se tendrá de que el intervalo contiene en realidad el verdadero valor de θ . Por otra parte, cuanto mayor sea el intervalo, tanto menor es la información que se tiene en torno al valor verdadero de θ . En una situación real, si se obtiene un intervalo relativamente corto con una confianza elevada mayor será la información en torno al valor verdadero.

Los problemas más comunes para estimar intervalos de confianzas en muestras aleatorias que enfrentará el investigador en las áreas de Ingeniería Química e Ingeniería de Alimentos, se muestra en forma resumida en la Tabla 4.3.

4.1.2.1 Nivel de Confianza y Precisión de la Estimación

Por ejemplo, el intervalo de confianza del 99 por ciento es más largo que el intervalo de confianza del 95 por ciento. Esto es por que se tiene un nivel de confianza más alto en el intervalo de confianza del 99 por ciento. En general, para una muestra fija de tamaño n y una desviación estándar σ , cuanto más alto es el nivel de confianza, tanto más largo el intervalo de confianza resultante.

Puesto que la *longitud* del intervalo de confianza mide la *precisión* de la estimación, se ve que la precisión se relaciona inversamente con el nivel de confianza. Para ello, es muy deseable obtener un intervalo de confianza lo

TABLA 4.3. Tipos problemas para cálculos de intervalos de confianzas para $\hat{\theta}$, como un estimador del parámetro θ

TIPO DE PROBLEMAS	ESTIMADOR POR PUNTOS	INTERVALOS DE CONFIANZA DE DOS LADOS 100(1-A) POR CIENTO	OBSERVACIONES
Media μ , varianza σ^2 conocida	\bar{X}	$\bar{X} - Z_{\alpha/2} \sigma / \sqrt{n} \leq \mu \leq \bar{X} + Z_{\alpha/2} \sigma / \sqrt{n}$	Sea X una variable aleatoria con media desconocida μ y varianza conocida, con tamaño de muestra aleatoria n, X_1, X_2, \dots, X_n . Con un intervalo de confianza de $(100-\alpha)$ por ciento en μ considerando la distribución de muestreo de X de la media de muestra \bar{X} . Se asume una distribución normal o casi normal.
Diferencia en dos medias μ_1 y μ_2 , varianza σ_1^2 y σ_2^2	$X_1 - X_2$	$X_1 - X_2 - Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq X_1 - X_2 + Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	Es un intervalo de confianza sobre dos medias, conocida la varianza Para dos variables independientes X_1 con media μ_1 desconocida y varianza conocida y X_2 con media μ y varianza σ_2^2 desconocida. Donde n_1 y n_2 son el numero de observaciones de las variables X_1 y X_2 respectivamente.

Continuación de la Tabla 4.3.

<p>Media μ de una distribución normal, varianza σ^2 desconocida</p>	<p>\bar{X}</p>	$\bar{X}_1 - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X}_1 + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$	<p>Es el intervalo de confianza sobre una media de una distribución normal con varianza desconocida.</p> <p>Donde n es el tamaño de muestra y \bar{X} y S son las medias y varianza de la muestra.</p> <p>Bajo la suposición de que la población la rige una distribución normal. Ello conduce a intervalos de confianza basadas en la distribución t.</p>
<p>Diferencia en las medias de dos distribuciones normales $\mu_1 - \mu_2$, varianza $\sigma^2_1 = \sigma^2_2$ desconocida</p>	<p>$\bar{X}_1 - \bar{X}_2$</p>	$\bar{X}_1 - \bar{X}_2 - t_{\alpha/2, n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X}_1 - \bar{X}_2 + t_{\alpha/2, n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ <p>donde $S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$</p>	<p>Si los tamaños de muestra n_1 y n_2 exceden entonces se comportara como una distribución normal, en caso contrario se supone que es una población normal y sustentar los intervalos de confianza en la distribución t</p> <p>Se asume dos variables aleatorias independientes desconociendo ambas medias y varianzas, entonces es razonable suponer que ambas varianzas son iguales.</p>

Continuación de la Tabla 4.3.

<p>Diferencias en las medias de dos distribuciones normales para muestras en pares $\mu_0 = \mu_1 - \mu_2$</p>	<p>D</p>	$D - t_{\alpha/2, n-1} S_D / \sqrt{n} \leq \mu_D \leq D + t_{\alpha/2, n-1} S_D / \sqrt{n}$	<p>Son dos muestras aleatorias independientes de las dos poblaciones de interés. Esto es, n_1 observaciones se seleccionaron al azar de la primera población y una muestra completamente independiente de n_2 observaciones se seleccionó al azar de la segunda población.</p> <p>Donde \bar{D} y S_D son la media y desviación estándar de muestra de las diferencias D_i.</p>
<p>Varianza σ^2 de una distribución normal</p>	<p>S^2</p>	$\frac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}}$	<p>Es para una muestra de tamaño n que se distribuye en forma normal con media y varianza desconocida</p> <p>El intervalo de confianza se basa en una distribución ji cuadrada</p>
<p>Cociente de dos varianzas σ_1^2 / σ_2^2 de dos distribuciones normales</p>	<p>$\frac{S_1^2}{S_2^2}$</p>	$\frac{S_1^2}{S_2^2} F_{1-\alpha/2, n_2-1, n_1-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2} F_{\alpha/2, n_2-1, n_1-1}$	<p>Con dos variables aleatorias normales con medias y varianzas desconocidas, se desea encontrar el intervalo de confianza respecto al cociente de las varianzas. Rige una distribución F</p>

Continuación de la Tabla 4.3.

<p>Proporción o parámetro de una distribución binomial p</p>	\hat{p}	$\hat{p} - Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	<p>El intervalo de confianza se rige por una distribución binomial. Por los parámetros n y p. Y es normal con media p y varianza $p(1-p)/n$. Y se aplica para poblaciones grandes</p>
<p>Diferencia en dos proporciones o dos parámetros binomiales $p_1 - p_2$</p>	$\hat{p}_1 - \hat{p}_2$	$\hat{p}_1 - Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \leq p_1 - p_2 \leq \hat{p}_1 - \hat{p}_2 + Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	<p>Dos proporciones de interés p_1 y p_2 es posible obtener un intervalo de confianza respecto a su diferencia. Si las dos muestras de tamaño n_1 y n_2 se toman de poblaciones infinitas de manera que X_1 y X_2 son variables aleatorias binomiales independientes</p>

suficientemente corto para propósitos de toma de decisiones y que a la vez brinde una confianza adecuada. Una manera de asegurar esto es eligiendo el tamaño de muestra n lo bastante grande para brindar un intervalo de confianza de longitud especificada con una confianza preestablecida.

4.1.2.2 Elección del Tamaño de Muestra

La precisión del intervalo de confianza para intervalos de confianza sobre la media, conocida la varianza (Tabla 4.3). Esto significa que usar \bar{x} para estimar μ , el error $E = |\bar{x} - \mu|$ es menor que $Z_{\alpha/2} \sigma / \sqrt{n}$ con confianza de $100(1-\alpha)$. En situaciones donde el tamaño de muestra puede controlarse, se puede elegir n como $100(1-\alpha)$ por ciento confiable de que el error al estimar μ sea menor que un error especificado E (Hines. Et al, 1997). El tamaño de muestra apropiado será:

$$n = \left(\frac{Z_{\alpha/2} \sigma}{E} \right)^2 \quad 4.12$$

Si el lado derecho de la ecuación 4.12 no es entero, debe redondearse hacia arriba. Adviértase que $2E$ es la longitud del intervalo de confianza resultante.

4.1.2.3 Intervalos de Confianza de un Lado

Es posible obtener intervalos de confianza de un lado para μ ajustando $L = -\alpha$ o $U = \alpha$ y reemplazando $Z_{\alpha/2}$ por Z_{α} . El intervalo de confianza superior de $100(1-\alpha)$ por ciento, para μ es:

$$\mu \leq \bar{X} + Z_{\alpha} \sigma / \sqrt{n} \quad 4.13$$

y el intervalo de confianza inferior del $100(1-\alpha)$ por ciento para μ es:

$$\bar{X} - Z_{\alpha} \sigma / \sqrt{n} \leq \mu \quad 4.14$$

4.1.2.4 Intervalos de Confianza Aproximados en la Estimación de Máxima Similitud.

Si se utiliza el método de máxima similitud para la estimación de parámetros, pueden emplearse las propiedades asintóticas de estos estimadores para obtener intervalos de confianza aproximados. Sea $\hat{\theta}$ el estimador de máxima similitud de θ . En muestra grandes $\hat{\theta}$ se distribuye aproximadamente de manera normal con media θ y varianza $V(\hat{\theta})$ dada por la cota inferior de Cramér-Rao (ecuación 4.5). por consiguiente, un intervalo de confianza aproximado del $100(1-\alpha)$ por ciento para θ es:

$$\hat{\theta} - Z_{\alpha/2} \left[V(\hat{\theta}) \right]^{1/2} \leq \theta \leq \hat{\theta} + Z_{\alpha/2} \left[V(\hat{\theta}) \right]^{1/2} \quad 4.15$$

Usualmente, la $V(\hat{\theta})$ es una función del parámetro conocido θ . En estos casos, sustitúyase θ por $\hat{\theta}$.

4.1.2.5 Intervalos de Confianza Simultáneos.

En ocasiones es necesario construir varios intervalos de confianza respecto a más de un parámetro, y se desea que haya una probabilidad de $(1-\alpha)$ que la totalidad de tales intervalos de confianza produzcan en forma simultánea enunciados correctos. Por ejemplo, supóngase que se toma una muestra de una población

normal con media y varianza desconocida, y se desea construir intervalos de confianza para μ y σ^2 tales que la probabilidad de que ambos intervalos produzcan simultáneamente conclusiones correctas sea $(1-\alpha)$. Puesto que \bar{X} y S^2 son independientes, se podría asegurar este resultado construyendo intervalos de confianza del $100(1-\alpha)^{1/2}$ por ciento para cada parámetro por separado, y ambos intervalos producirán en forma simultánea conclusiones correctas con probabilidad $(1-\alpha)^{1/2}(1-\alpha)^{1/2}=(1-\alpha)$ (Hines. Et al, 1997).

Si las estadísticas de la muestra en las cuales se basan los intervalos de confianza no son variables aleatorias independientes, entonces los intervalos de confianza no son independientes, y deben emplearse otros métodos. En general, supóngase que se requieren m intervalos de confianza. La desigualdad de Bonferroni establece lo siguiente:

$$P\{m \text{ enunciados}\} \equiv 1 - \alpha \geq 1 - \left(\sum_{i=1}^m \alpha_i \right) \quad 4.16$$

donde m son los enunciados simultáneamente correctos y $(1-\alpha_i)$ es el nivel de confianza i ésimo. En la práctica, se selecciona un valor para el nivel de confianza simultáneo $(1-\alpha)$, después se elige la α_i individual tal que el sumatorio de todos los α_i sean igual a α . Usualmente se hace $\alpha_i = \alpha/m$. Como ejemplo, supóngase que se desea construir dos intervalos de confianza respecto a las medias de dos distribuciones normales tal que tenga el 90 por ciento de confianza de que ambos enunciados sean simultáneamente correctos. Por tanto que $1-\alpha=0.90$, se tiene $\alpha=0.10$, y puesto que se requieren dos intervalos de confianza, cada uno de éstos debe construirse con $\alpha_i = \alpha/2=0.10/2=0.05$, $i=1,2$. Esto es, dos intervalos de confianza individuales del 95 por ciento en μ_1 y μ_2 conducirán simultáneamente a enunciados correctos con probabilidad al menos de 0.90.

4.2 PRUEBAS DE HIPÓTESIS

Muchos problemas en investigaciones en las áreas de Ingeniería Química e Ingeniería de Alimentos requieren decidir si se acepta o rechaza un enunciado acerca de algún parámetro. El enunciado suele llamarse hipótesis, y el procedimiento de toma de decisiones en torno a la hipótesis recibe el nombre de prueba de hipótesis. Esto es uno de los aspectos más útiles de la inferencia estadística, puesto que muchos tipos de problemas de investigación con decisión pueden formularse como prueba de hipótesis. En esta sección se desarrollarán procedimientos de prueba de hipótesis para varias situaciones importantes en el campo de investigación en las áreas de interés.

4.2.1 HIPÓTESIS ESTADÍSTICAS

Una hipótesis estadística es un enunciado acerca de la distribución de probabilidad de una variable aleatoria. Las hipótesis estadísticas a menudo involucran uno o más parámetros de esta distribución.

Por lo general, los elementos de cualquier prueba estadística son (Hines. Et al, 1997).:

- a *hipótesis nula*: Es aquella hipótesis que desea contrastar a la hipótesis de investigación o de trabajo, como medio que se sirve el investigador para probar la hipótesis alternativa, se simboliza por H_0 . El planteamiento de la hipótesis nula permite elaborar un modelo probabilístico a partir del cual se puede llegar a una decisión final.
- b *Hipótesis alternativa*: También se conoce como experimental o de trabajo y se representa por H_1 , esta es la hipótesis de investigación. De modo que se

espera que hay un argumento para la hipótesis de investigación (alternativa) H_1 , demostrando que no lo hay para su contraria, la hipótesis nula.

- c *Estadístico de prueba*: Los valores muestrales se emplean un solo número que corresponde a un punto en una recta. Tal número sirve para tomar decisiones y se denomina *estadístico de prueba* y se simboliza en general como T . Tal estadístico, será, una variable aleatoria, con una distribución muestral determinada, que proporcionará las probabilidades asociadas a un valor o un determinado intervalo de valores del Estadístico del Contraste
- d *Regla de decisión*: Una regla de decisión es un criterio utilizado para decidir si se acepta o rechaza la hipótesis nula, a partir del espacio muestral de valores del estadístico de contraste y de las probabilidades asociadas. Este criterio consiste en dividir tal espacio en dos zonas mutuamente excluyentes y exhaustivas: *la zona de rechazo* (región crítica) y *la zona de aceptación*.

Es importante recordar que las hipótesis son siempre anunciados relativos a la población o distribución bajo estudio, no enunciados en torno a la muestra. El valor del parámetro de la población especificado en la hipótesis nula suele determinarse en una de tres maneras. Primero, puede resultar de la experiencia o conocimiento pasado del proceso, o incluso de experimentación previa. El objetivo de la prueba de hipótesis suele ser entonces determinar si la situación experimental a cambiado. Segundo, este valor puede determinarse a partir de alguna teoría o modelo con respecto al objeto que se estudia. Aquí el objeto de la prueba de hipótesis es verificar la teoría o modelo. Una tercera situación surge cuando el valor del parámetro de la población es resultado de consideraciones experimental, tales como especificaciones de diseño o ingeniería, o de obligaciones contractuales. En esta situación, el objetivo usual de la prueba de hipótesis es la prueba de conformidad.

Si se esta interesados en tomar un decisión en torno a la veracidad o falsedad de una hipótesis. Un procedimiento que conduce a tal decisión se llama *prueba de una hipótesis* los procedimientos de la prueba de hipótesis dependen del uso de la información en una muestra aleatoria de la población de interés. Si esta información es consistente con la hipótesis, entonces se concluye que la hipótesis es verdadera; sin embargo, si esta información es inconsistente con la hipótesis, se concluye que la hipótesis es falsa.

4.2.2 ERRORES DE TIPO I Y TIPO II

La decisión para aceptar o rechazar la hipótesis nula se basa en una estadística de prueba calculada a partir de los datos en una muestra aleatoria. Cuando se toma una decisión utilizando la información en una muestra aleatoria, esta decisión está sujeta a error. Pueden producirse dos tipos de errores cuando se prueban hipótesis. Si la hipótesis nula se rechaza cuando es verdadera, entonces se ha cometido un error de tipo I. Si la hipótesis nula se acepta cuando es falsa, entonces el error cometido es de tipo II. Esta situación se describe en la Tabla 4.4.

Tabla 4.4. Decisiones en la prueba de hipótesis

DECISIÓN	H_0 es verdadera	H_0 es falsa
Aceptación de H_0	Ningún error	Error del tipo II
Rechazo de H_0	Error del tipo I	Ningún error

Las probabilidades de ocurrencia de los errores de tipo I y de tipo II tienen símbolos especiales como se muestra a continuación (Hines. Et al, 1997):

$$\alpha = P \{ \text{error tipo I} \} = P \{ \text{rechazar } H_0 / H_0 \text{ verdadera} \} \quad 4.17$$

$$\beta = P \{ \text{error tipo II} \} = P \{ \text{aceptar } H_0 / H_0 \text{ es falsa} \}$$

Algunas veces es conveniente trabajar con la *potencia* de la prueba, donde

Potencia = $1 - \beta$ {rechazar H_0/H_0 es falsa

4.18

Adviértase que la potencia de la prueba es la probabilidad de que una hipótesis nula falsa se rechace correctamente. Debido a que los resultados de una prueba de hipótesis están sujetos a error, no se puede "probar" o "desaprobar" una hipótesis estadística. Sin embargo, es posible designar procedimientos de prueba que controlen las probabilidades de error α y β a valores adecuadamente pequeños.

La probabilidad α del error de tipo I a menudo se llama *nivel o tamaño de significación* de la prueba. En general, la probabilidad del error de tipo I está controlada la localización de la región crítica. Por consiguiente, en la práctica el analista fija la probabilidad de error de tipo I en (cerca de) cualquier valor deseado. Es decir que la probabilidad de rechazar en forma errónea H_0 está controlada directamente por el que toma las decisiones. Además, disminuye cuando el tamaño de muestra n aumenta. Por lo tanto, la probabilidad β del error de tipo II está controlada por el tamaño de muestra como del grado al que es falsa la hipótesis nula H_0 , es común considerar a la decisión de aceptar H_0 como una *conclusión débil*, a menos que se sepa que β es aceptablemente pequeña. Por tanto, en vez de decir que "se acepta H_0 ", se prefiere la terminología "no se rechaza H_0 ". El no rechazar H_0 implica que no se tiene la evidencia suficiente para rechazarla, esto es, para realizar un enunciado fuerte. De tal manera, que para un tamaño de muestra dado, la disminución α provoca que β aumente, y el incremento α provoca que β disminuya.

4.2.3 HIPÓTESIS UNILATERALES O BILATERALES

Se debe a que rechazar H_0 es una conclusión fuerte en un tanto que no H_0 puede ser una conclusión débil al menos que se sepa que β es pequeña, se prefiere en la mayor parte de los casos construir hipótesis tales que el enunciado entorno al cual

se desea una conclusión fuerte, esté en la hipótesis alternativa H_1 . los problemas para los cuales es apropiada una hipótesis alternativa de dos lados en realidad se desea probar la hipótesis de que la media de una distribución μ es igual a un valor arbitrario, digamos μ_0 , y es importante detectar valores de la media verdadera μ que podrían ser más grandes o mas pequeños μ_0 , entonces debe utilizarse la alternativa de dos lados (bilateral) como se muestra a continuación (Hines. Et al, 1997):

$$\begin{aligned} H_0 : \mu &= \mu_0 \\ H_1 : \mu &\neq \mu_0 \end{aligned} \quad 4.19$$

Muchos problemas de pruebas de hipótesis involucran de manera natural las hipótesis alternativas de un solo lado (unilateral). Por ejemplo, supóngase que se desea rechazar H_0 solo cuando el valor verdadero de la media supera a μ_0 . las hipótesis serian:

$$\begin{aligned} H_0 : \mu &= \mu_0 \\ H_1 : \mu &> \mu_0 \end{aligned} \quad 4.20$$

Esto implica que la región crítica se localiza en la cola superior que la distribución que la estadística de prueba. Esto es, si la decisión se basará en el valor de la media de muestra \bar{x} , entonces se rechazaría H_0 en la ecuación 4.20 si \bar{x} es demasiado grande. La curva característica de operación para la prueba correspondiente a esta hipótesis se muestra en el Anexo N, junto con la curva característica de operación para prueba bilateral. Se observa que cuando es cierto que la media verdadera μ supera μ_0 , la prueba unilateral es superior a la prueba bilateral en el sentido de que tiene una curva característica de operación con pendiente más pronunciada. Cuando la media verdadera $\mu = \mu_0$, las prueba unilaterales y bilaterales son equivalentes. Sin embargo, cuando la media

verdadera μ es menor que μ_0 , las dos curvas características de operación difieren. Si $\mu < \mu_0$, la prueba bilateral tiene una mayor probabilidad de detectar esta desviación respecto a μ_0 , que la prueba unilateral. Esto es intuitivamente atrayente ya que la prueba unilateral está diseñada suponiendo que μ no puede ser menor que μ_0 , o bien que, si es menor que μ_0 , sería deseable aceptar la hipótesis nula.

4.2.4 PRUEBA HIPÓTESIS EN MEDIAS Y VARIANZAS.

Para las pruebas de hipótesis, tanto para variables continuas (para medias y varianzas) como discontinuas (proporción) se utilizará la comparación de la hipótesis nula H_0 y la hipótesis alternativa H_1 , con el fin de sacar conclusiones del problema de estudio y tomar la decisión de rechazar o aceptar la hipótesis alternativa. En forma general Proporcionando con ello el criterio de rechazo tal como se muestra en la tabla 4.5. En la cual se da una breve descripción de las problemas de pruebas de hipótesis, tales como:

- 1 Hipótesis referente a una medias
- 2 Hipótesis referente a dos medias
- 3 Hipótesis referente a una varianza
- 4 Hipótesis referente a dos varianza
- 5 Hipótesis referente a una proporción
- 6 Hipótesis referente a varias proporciones

Tabla 4.5. Ilustración del procedimientos de los problemas de pruebas de hipótesis en medias, varianzas

HIPÓTESIS NULA	ESTADISTICA DE PRUEBA	HIPÓTESIS ALTERNATIVA	CRITERIO DE RECHAZO	ELECCIÓN DEL TAMAÑO DE MUESTRA	OBSERVACIONES
$H_0: \mu = \mu_0$ σ^2 conocida	$Z_0 = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$	$H_1: \mu \neq \mu_0$ $H_1: \mu > \mu_0$ $H_1: \mu < \mu_0$	$Z_0 > Z_{\alpha/2}$ $Z_0 > Z_\alpha$ $Z_0 < -Z_\alpha$	$d = \mu - \mu_0 / \sigma$ $d = (\mu - \mu_0) / \sigma$ $d = (\mu_0 - \mu) / \sigma$ $n = \frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{\delta^2}$ $d = \frac{ \delta }{\sigma}$	Es una prueba de hipótesis sobre la media, con varianza conocida, para una población X representado por una variable aleatoria, con distribución normal. En la cual lo zona de aceptación define la región crítica o región de rechazo viene dada por el planteamiento del problema de la prueba de hipótesis. Para determinar la elección de muestra, se puede determinar directamente por la ecuación, en función de la probabilidades de la zona crítica (en función de los errores de tipo I y II). Dado que β depende del tamaño de muestra. O utilizando las curvas características de operación (CO), de las graficas del Anexo Na y Nb. Utilizando el parámetro d.

Continuación de la Tabla 4.5.

$H_0: \mu = \mu_0$ σ^2 desconocida	$t_{0\alpha} \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$	$H_1: \mu \neq \mu_0$ $H_1: \mu > \mu_0$ $H_1: \mu < \mu_0$	$ t_0 > t_{\alpha/2, n-1}$ $t_0 > t_{\alpha, n-1}$ $t_0 < -t_{\alpha, n-1}$	$d = \mu - \mu_0 / \sigma$ $d = (\mu - \mu_0) / \sigma$ $d = (\mu_0 - \mu) / \sigma$	<p>Es la relación existente en la prueba de hipótesis con una media de una población cuando la varianza se desconoce, el procedimiento se aplica para poblaciones grandes ($n > 30$) y, se asume una distribución normal. Sin embargo para muestras pequeñas se supone en muchos casos una distribución normal.</p> <p>Para determinar el tamaño de muestras se hace uso el parámetro d, ilustrado en las graficas del Anexo Ne, Nf, Ng y Nh.</p>
---	--	---	--	--	--

Continuación de la Tabla 4.5.

$H_0: \mu_1 = \mu_2$ σ_1^2 y σ_2^2 conocidas	$Z_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$H_1: \mu_1 \neq \mu_2$ $H_1: \mu_1 > \mu_2$ $H_1: \mu_1 < \mu_2$	$Z_0 > Z_{\alpha/2}$ $Z_0 > Z_\alpha$ $Z_0 < -Z_\alpha$	$d = \frac{ \mu_1 - \mu_2 }{\sqrt{\sigma_1^2 + \sigma_2^2}}$ $d = \frac{(\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}}$ $d = \frac{(\mu_2 - \mu_1)}{\sqrt{\sigma_1^2 + \sigma_2^2}}$ $n = \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ $= \frac{(Z_\alpha + Z_\beta)^2 (\sigma_1^2 + \sigma_2^2)}{\delta^2}$ $d = \frac{\delta}{\sqrt{\sigma_1^2 + \sigma_2^2}}$	<p>Sean X_1 y X_2 dos poblaciones que tienen medias desconocidas (μ_1 y μ_2) y varianzas conocidas (σ_1^2 y σ_2^2). Estando interesados en la prueba de hipótesis sean iguales.</p> <p>Se puede utilizar la fórmula directa para el cálculo del tamaño de muestra n, para $n_1 \neq n_2$. O haciendo uso de las curvas CO representadas en las gráficas del Anexo Nc y Nd. Utilizando el parámetro d, para $n = n_1 = n_2$.</p>
--	--	---	---	--	---

Continuación de la Tabla 4.5.

$H_0: \mu_1 = \mu_2$ $\sigma^2_1 = \sigma^2_2 = \sigma^2$ desconocida s	$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $v = n_1 + n_2 - 2$ $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	$H_1: \mu_1 \neq \mu_2$ $H_1: \mu_1 > \mu_2$ $H_1: \mu_1 < \mu_2$	$ t_0 > t_{\alpha/2, n_1 + n_2 - 2}$ $t_0 > t_{\alpha, n_1 + n_2 - 2}$ $t_0 < -t_{\alpha, n_1 + n_2 - 2}$	$d = \mu_1 - \mu_2 / 2\sigma$ $d = (\mu_1 - \mu_2) / 2\sigma$ $d = (\mu_2 - \mu_1) / 2\sigma$	Son pruebas de hipótesis respecto a la igualdad de medias de dos distribuciones normales donde se desconocen las varianzas, siendo estas iguales, empleando una estadística t para probar las hipótesis. No afectando las desviaciones moderadas la normalidad de los datos. Dado que se desconoce las varianzas, la determinación del tamaño de muestras, se hace uso del parámetro d, utilizando las graficas del Anexo Ne, Nf, Ng y Nh, asumiendo $n_1 = n_2$
--	--	---	--	---	---

Continuación de la Tabla 4.5.

$H_0: \sigma^2 = \sigma_0^2$	$X_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$	$H_1: \sigma^2 \neq \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$	$X_0^2 > X_{\alpha/2, n-1}^2$ o $X_0^2 < X_{1-\alpha/2, n-1}^2$ $X_0^2 > X_{\alpha, n-1}^2$ $X_0^2 < X_{1-\alpha}^2$	$\lambda = \sigma / \sigma_0$ $\lambda = \sigma / \sigma_0$ $\lambda = \sigma / \sigma_0$	<p>Son pruebas de hipótesis sobre la varianza o desviación típica de una población normal. En la cual se desea probar la hipótesis de una distribución normal σ^2 es igual a un valor especificado σ_0^2.</p> <p>Para determinar el tamaño de muestra se basa en las curvas CO donde se ilustran en los diagrama del Anexo Ni, al Ng. Para la hipótesis alternativa de dos lados, los diagramas Ni y Nj, graficando β contra el parámetro de abscisas (λ) para diversos tamaños de muestras, donde σ denota el valor verdadero de la desviación estándar. Los diagramas Ik y Il corresponden a la alternativa $H_1: \sigma^2 > \sigma_0^2$, en tanto Im y In son para la otra alternativa de un lado $H_1: \sigma^2 < \sigma_0^2$. al emplear los valores, considera σ como el valor de la desviación estándar que se desea detectar.</p>
------------------------------	---------------------------------------	---	---	---	---

Continuación de la Tabla 4.5.

$H_0: \mu_1 = \mu_2$ $\sigma^2_1 \neq \sigma^2_2$ desconocidas	$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ $v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1+1} + \frac{\left(S_2^2/n_2\right)^2}{n_2+1}}$	$H_1: \mu_1 \neq \mu_2$ $H_1: \mu_1 > \mu_2$ $H_1: \mu_1 < \mu_2$	$ t_0 > t_{\alpha/2, v}$ $t_0 > t_{\alpha, v}$ $t_0 < -t_{\alpha, v}$		Esta es la situación que las varianzas son desconocidas y desiguales (caso II de la prueba de hipótesis anterior).
--	---	---	--	--	--

Continuación de la Tabla 4.5.

$H_0: \sigma_1^2 = \sigma_2^2$	$F_0 = \frac{S_1^2}{S_2^2}$ $F_0 = \frac{S_1^2}{S_2^2}$	$H_1: \sigma_1^2 \neq \sigma_2^2$ $H_1: \sigma_1^2 > \sigma_2^2$ $H_1: \sigma_1^2 < \sigma_2^2$	$F_0 > F_{\alpha/2, n_1-1, n_2-1}$ o $F_0 < F_{1-\alpha/2, n_1-1, n_2-1}$ $F_0 > F_{\alpha, n_1-1, n_2-1}$ $F_0 > F_{\alpha/2, n_2-1, n_1-1}$	$\lambda = \sigma_1 / \sigma_2$ $\lambda = \sigma_1 / \sigma_2$ $\lambda = \sigma_2 / \sigma_1$	<p>Es una prueba para comparar dos varianzas (prueba F) para dos poblaciones normales y pruebas de muestras grandes que pueden aplicarse a poblaciones no normales. En la cual son dos poblaciones X_1 y X_2, donde μ_1, σ_1^2, μ_2 y σ_2^2 se desconocen, para dos muestras aleatorias de tamaño n_1 de la población 1 y de tamaño n_2 para la población 2, con n_1-1 y n_2-1 como sus grados de libertad. Para la elección del tamaño de muestras se utilizan los diagramas o al r del Anexo N, que proporciona las curvas CO para la prueba F suponiendo que $n=n_1=n_2$. los diagramas No y Np se emplea para hipótesis alternativas de dos lados y, los diagramas Nq y Nr para alternativa de un solo lado. Ellos grafican β contra el parámetro de las abscisas λ.</p>
--------------------------------	--	---	--	---	---

NOTA: las pruebas de hipotes ilustradas en el cuadro resumen se consideran pruebas de hipótesis de ambos lados (bilaterales) y de un lado (mayor y menor que la hipótesis nula) o hipótesis unilateral. El método para elección de muestras para la pruebas de hipótesis consiste en analítico y por lo general utilizando las curvas características de operación (CO). Asumiendo con esto que el error de tipo II (β) es función del tamaño de muestras.

4.3 PRUEBAS NO PARAMÉTRICAS.

La mayor parte de la influencia estadística (estimación de parámetro, procedimientos de intervalo de confianza y las pruebas de hipótesis) vistos anteriormente se han basado la suposición de que estas son muestras aleatorias de poblaciones con distribuciones establecidas. Más sin embargo, la mayor parte de estos procedimientos son relativamente insensibles a ligeras desviaciones respecto a la normalidad. En general, las pruebas de t y F y los intervalos de confianza t tendrán niveles de significación o niveles de confianza reales que difieren de los niveles nominales o problemas elegidos por el experimentador, aunque la diferencia entre los niveles reales y enunciado suele ser sin duda mas pequeña cuando la población de base no es demasiado diferente a la normal. Por lo común, se le llaman a estos procedimientos, métodos *paramétricos* debido a que se basa en una familia paramétrica particular de distribuciones normal. Alternativamente existen procedimientos denominados *no parámetros o libres de distribución* no se suele hacer suposiciones a la distribución normal (o otra distribución de población, aparte de que es continua. Dentro de las ventajas ofrecidas de las por los métodos paramétricos (Johnson, 1997), se encuentran:

- 1 Los procedimientos paramétricos utilizan toda la información proporcionada por la muestra (no se realizan aproximaciones) por lo tanto ofrece mayor eficiencia que el procedimiento correspondiente cuando la población de base es normal (o otra distribución). Será menos eficiente para requerimientos de tamaño de muestra mas grandes.
- 2 Realizan inferencias exactas cuando los métodos estándar son imposibles de cumplir.
- 3 Las pruebas solo dependen de relaciones de orden entre observaciones (no dependen del tamaño de muestra ni se ponen mucho menos aproximaciones sobre la forma de las poblaciones). La pérdida de eficiencia por lo común no

es grande, y a menudo la diferencia de en el tamaño de muestra es muy pequeña.

- 4 No se ven afectadas por la simetría u otras desviaciones respecto a la normalidad (su fuerza es usualmente satisfactoria aun cuando las poblaciones se desvíen de la normalidad).
- 5 La elección de escala de medición numérica es arbitraria. No obstante, su mayor virtud es el echo que el nivel de significación o intervalo de confianza es exacto incluso si las poblaciones son excesivamente normales.

Dentro de las pruebas no paramétricas de mayor uso en el campo de Ingeniería química e Ingeniería de alimentos, se tiene:

- a Prueba de signo.
- b Prueba de Wilcoxon del rango con signo
- c Prueba de Wilcoxon de la suma de rango.
- d Prueba de Kruskal – Wallis.

Estas pruebas se describirá en forma general en la Tabla 4.6, los estadísticos de prueba de hipótesis, criterio de rechazo, parámetros y definición de variables para cada una de las pruebas.

Tabla 4.6. Resumen de procedimientos de pruebas de hipótesis en medias y varianzas para las pruebas no paramétricas.

Prueba	Hipótesis nula	Hipótesis alternativa	Parámetro estadístico	Región de rechazo	Observaciones
Prueba de signo	$H_0 : \mu = \mu_0$	$H_1 : \mu \neq \mu_0$ $H_1 : \mu > \mu_0$ $H_1 : \mu < \mu_0$	R	$R < R_\alpha^*$ $R^- < R_\alpha^*$ $R^+ < R_\alpha^*$	<p>Se emplea para probar hipótesis en torno a la media de una distribución continua y una variable aleatoria, cuyo probabilidad del valor observado X sea mayor que la mediana es 0.5 y que sea menor que la mediana es 0.5.</p> <p>Cualquier diferencia $X_i - \mu_0$ es igualmente posible de ser negativa o positiva. De modo que R^+ denota el número de observaciones de las diferencias positivas y R^- las diferencias negativas. Y R el número de observaciones.</p> <p>Cuando las desigualdades se cumplen, entonces las hipótesis se rechazan.</p> <p>El valor crítico R_α^* se elige para un nivel de significación tan cercano al α del enunciado. Y se muestra en el Anexo K. El nivel de significación para una prueba de un lado es la mitad mostrada en la tabla del Anexo K.</p>

Continuación de la Tabla 4.6.

<p>Prueba de Wilcoxon del rango con signo</p>	$H_0 : \mu = \mu_0$	$H_1 : \mu \neq \mu_0$ $H_1 : \mu > \mu_0$ $H_1 : \mu < \mu_0$	<p>R</p>	$R < R_\alpha^*$ $R > R_\alpha^*$ $R < R_\alpha^*$	<p>Considera que las poblaciones de interés son continuas y simétricas, es decir la mediana $\tilde{\mu}$ es igual a la media μ. Una desventaja de la prueba de signo es que solo considera el signo de la desviación $X_i - \tilde{\mu}_0$ y no sus magnitudes y la Wilcoxon se diseñó para superar esta desventaja.</p> <p>Los niveles de significación se encuentran en el Anexo L. Y si las desigualdades se cumplen con el valor crítico la hipótesis se rechaza.</p> <p>Para hipótesis de un lado será la mitad del nivel de significación del Anexo L.</p> <p>Para poblaciones mayores de 20, se comporta aproximadamente como distribución normal con media $(\mu_R) = (n(n+1))/4$ y una varianza $(\sigma^2_R) = (R-n(n+1)(2n+1))/24$ y como estadística de prueba $Z_0 = (R - \mu_R) / \sigma_R$.</p>
---	---------------------	--	----------	--	---

Continuación de la Tabla 4.6.

<p>Prueba de Wilcoxon de la suma de rango</p>	<p>$H_0 : \mu_1 = \mu_2$</p>	<p>$H_1 : \mu_1 \neq \mu_2$ $H_1 : \mu_1 > \mu_2$ $H_1 : \mu_1 < \mu_2$</p>	<p>R_1 y R_2</p>	<p>$R_1 < R_\alpha^*$ o $R_2 < R_\alpha^*$ $R_2 \leq R_\alpha^*$ $R_1 \leq R_\alpha^*$</p>	<p>Se tiene dos poblaciones independientes continuas X_1 y X_2 con media μ_1 y μ_2. las distribuciones de X_1 y X_2 tienen la misma forma de dispersión, y se difieren solo (posiblemente) en sus medidas. La prueba de Wilcoxon de la suma de rango se puede utilizar para doblar la hipótesis $H_0: \mu_1 = \mu_2$. algunas veces este procedimiento recibe el nombre de prueba de Mann-Whitney, aunque la estadística de prueba de Mann-Whitney suele expresarse de forma diferente. Sean X_{11}, \dots, X_{n1} y $X_{21}, X_{22}, \dots, X_{2n1}$ dos muestras aleatorias independientes de las poblaciones continuas X_1 y X_2 descritas antes. Se supone que $n_1 \leq n_2$. Arreglase las observaciones $n_1 + n_2$ en orden de magnitud ascendente asignándoles rangos. Sea R_1 la suma de los rangos en la muestra mas pequeña (1), y defínase.</p> $R_2 = n_1(n_1 + n_2 + 1) - R_1$ <p>si después de esto no se difieren las medidas de muestra, se espera que la suma de los rangos fuera casi igual para ambas muestras. En consecuencia, si la suma de los rangos difiere en modo considerable, se concluye que las medidas no son iguales. La tabla del Anexo L contiene el valor critico de las sumas del rango para $\alpha = .05$ y $\alpha = .01$ con los tamaños de muestra n_1 y n_2 apropiados, puede obtenerse el valor critico R_α^*. La hipótesis nula $H_0: \mu_1 = \mu_2$ se rechaza a favor de $H_1 : \mu_1 \neq \mu_2$ ya sea R_1 y R_2 sean menores o iguales que el valor critico tabulador R_α^*.</p> <p>Cuando las desigualdades del criterio se cumple las hipótesis rechaza para el problema de hipótesis.</p> <p>Para muestras grandes (mayores de 8) se puede aproximar a la distribución normal con media μ_R y varianza σ_R^2</p>
---	---	--	--	--	--

Continuación de la Tabla 4.6.

<p>Prueba Kruskal- Wallis</p>	$H_0 : \mu_1 = \mu_2 = \dots$		$K = \frac{12}{N(N+1)} \sum_{i=1}^a \frac{R_i^2}{n_i} - 3(N+1)$	<p>K</p> <p>Consiste en un modelo de análisis de varianza de un solo factor y es una prueba alternativa de la prueba F (distribución normal) y se necesita una misma distribución continua para todos los tratamientos $i=1, 2, \dots, a$, donde N es número total de observaciones y todas las posibles asignaciones de los N rangos 1, 2, ..., N se mezclen a lo largo de las muestras a.</p> <p>Esta prueba mide el grado en el cual los rangos promedios reales observados R_i difieren del valor esperado $(N+1)/2$. si esta diferencia es grande, entonces la se rechaza la hipótesis nula H_0.</p> <p>La hipótesis nula se rechaza para datos de muestra que generen un valor grande de K.</p> <p>Para muestras grandes , K se aproxima a una distribución ji cuadrada con a-1 como grados de libertad, por medio de la relación</p> $K \geq X_{\alpha, a-1}^2$ <p>con apropiadas regiones criticas de dos colas o cola superior o inferior.</p>
---------------------------------------	-------------------------------	--	---	---

4.4 EJEMPLOS DEL CAPÍTULO¹

Ejemplo 4.1

Un artículo en el Journal of Heat Transfer (Trans. ASME, Ses.) Describe un método para medir la conductividad térmica de hierro Armco. Al emplear una temperatura de 100 °F y una entrada de potencia de 550 W, se obtuvieron 10 mediciones de conductividad térmica (en Btu/hr -pie-°F):

41.60, 41.48, 42.34, 41.95, 41.86

42.18, 41.72, 42.26, 41.81, 42.04

- La media de la muestra y desviación estándar estimada.
- Calcular el tamaño de muestra suponiendo que se desea un error en la estimación de la conductividad térmica media del hierro Arcmo sea menor a 0.05 Btu/hr -pie-°F y $\sigma=0.10$.

Solución:

Para el literal (a), se tiene que:

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = 41.924 \quad \text{Btu/hr-pie-°F}$$

y la varianza de la muestra, es

$$s^2 = \frac{\sum_{i=1}^{10} (x_i - \bar{x})^2}{10-1} = 0.081 \quad (\text{Btu/hr-pie-°F})^2$$

Como σ se desconoce, se sustituye por la varianza de muestra $s=0.284$ para obtener el error y estándar estimado de la media muestral

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = 0.284 / \sqrt{10} = 0.0898$$

¹ Todos los ejemplos desarrollados en esta sección fueron extraídos de Hines. Et al, 1997.

entonces el error estándar es cerca del 0.2 por ciento de la media de la muestra, lo que implica que se ha obtenido una buena precisión de la estimación puntual de la conductividad térmica.

Para el literal (b), con una confianza del 95 por ciento y $\sigma=0.10$ y $Z_{0.025}=1.96$, se puede encontrar el tamaño de la muestra requerido a partir de la ecuación 4.12 como

$$n = \left[\frac{(1.96)0.10}{0.05} \right]^2 = 15.37 = 16$$

Nótese cómo, en general, el tamaño de muestra se comporta como una función de la longitud del intervalo de confianza $2E$, el nivel de confianza del $100(1-\alpha)$ por ciento, y la desviación estándar σ , de las siguientes maneras:

Ejemplo 4.2

Considérese los datos de conductividad térmica del ejemplo 4.1. Encontrar un intervalo de confianza de 95 por ciento en la conductividad térmica media de hierro Armco. Suponiendo que se conoce que la desviación estándar de la conductividad térmica a 100 °F y 550 W es $\sigma=0.10$ Btu/hr-pie-°F.

Solución:

Un intervalo de confianza del 95% implica que $1-\alpha=0.95$, así que $\alpha=0.05$, como se supone que se trata de una variable aleatoria que se comporta normalmente por tabla del Anexo C se lee $Z_{\alpha/2}=Z_{0.025}=1.96$

Límite de confianza inferior, se tiene:

$$L = \bar{x} - Z_{0.025} \cdot \frac{\sigma}{\sqrt{n}}$$

sustituyendo y calculando

$$L = 41.924 - 1.96 * 0.10 / \sqrt{10} = 41.862$$

Para el limite de confianza superior, se tiene:

$$U = x + Z_{0.025} * 0.10 / \sqrt{10}$$

sustituyendo y calculando

$$U = 41.924 + 1.96 * 0.10 / \sqrt{10} = 41.986$$

Por consiguiente el limite de confianza de dos lados del 95% es:

$$41.862 \leq \mu \leq 41.986$$

Ejemplo 4.3

Se efectuaron pruebas de resistencias a la presión de dos clases de tuberías empleados en las instalaciones de vapor en la industria. De la experiencia pasada en las instalaciones de las tuberías, se suponen conocidas las desviaciones estándares de la resistencia a la presión. Los datos se muestran en la Tabla 4.7. Asumiendo que se comporta como una variable aleatoria con distribución normal, calcular los intervalos de confianza del 90 por ciento.

Tabla 4.7. Resultados de prueba de resistencia a la presión de las tuberías

Clase de tubo	Tamaño de muestra	Resistencia a la presión media de muestra (kg/mm ²)	Desviación estándar (kg/mm ²)
1	n ₁ =10	87.6	1.0
2	n ₂ =12	74.5	1.5

Solución:

Como $1-\alpha=0.90$ para un intervalo de confianza del 90 por ciento, entonces $\alpha=0.10$, por tabla del Anexo C se lee $Z_{0.05}=1.645$. Para el limite de confianza inferior se tiene:

$$L = X_1 - X_2 - Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\begin{aligned}
 &= 87.6 - 74.5 - 1.645 \sqrt{\frac{1.0^2}{10} - \frac{1.5^2}{12}} \\
 &= 12.22 \text{ kg/mm}^2
 \end{aligned}$$

Para el límite de confianza superior se tiene:

$$\begin{aligned}
 U &= X_1 - X_2 + Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \\
 &= 87.6 - 74.5 + 1.645 \sqrt{\frac{1.0^2}{10} - \frac{1.5^2}{12}} \\
 &= 13.98 \text{ kg/mm}^2
 \end{aligned}$$

En consecuencia el intervalo de confianza de dos lados del 90 por ciento en la diferencia en la resistencia media es:

$$12.22 \text{ kg/mm}^2 \leq \mu_1 - \mu_2 \leq 13.98 \text{ kg/mm}^2$$

Se tiene una confianza del 90 por ciento de que la resistencia a la presión media de la tubería 1 excede a la tubería 2 entre 12.22 y 13.98 kg/mm².

Ejemplo 4.4

Un artículo en *Journal of Testing and Evaluation* (Vol. 10, Núm. 4, 1982, p. 133) presenta las siguientes 20 mediciones del tiempo residual de flama (en segundos) en muestras tratadas de ropa de dormir para niños:

9.85, 9.93, 9.75, 9.77, 9.67,
 9.87, 9.67, 9.94, 9.85, 9.75,
 9.83, 9.92, 9.74, 9.99, 9.88,
 9.95, 9.95, 9.93, 9.92, 9.89,

Determinar el límite de confianza del 95 por ciento respecto al tiempo residual de flama media.

Solución:

Calculando la media, varianza y desviación estándar para la muestra, se tiene:

$$\bar{x} = 9.8475 \text{ seg.}$$

$$s^2 = 0.009101 \text{ seg}^2$$

$$s = 0.0954 \text{ seg}$$

Como es el intervalo de confianza del 95 por ciento, $\alpha=0.05$ y $v=20-1=19$. Entonces de la tabla del Anexo E se encuentra que $t_{0.025,19}=2.093$. por lo tanto los límites de confianza inferior y superior del 95 por ciento son:

$$\begin{aligned} L &= \bar{x} - t_{\alpha/2, n-1} s / \sqrt{n} \\ &= 9.8475 - 2.093(0.0954) / \sqrt{20} \\ &= 9.8029 \text{ seg} \end{aligned}$$

y

$$\begin{aligned} L &= \bar{x} + t_{\alpha/2, n-1} s / \sqrt{n} \\ &= 9.8475 + 2.093(0.0954) / \sqrt{20} \\ &= 9.8921 \text{ seg.} \end{aligned}$$

Por tanto, el intervalo de confianza de dos lados del 95 por ciento es:

$$9.8029 \leq \mu \leq 9.8921 \text{ seg.}$$

Se tiene una confianza del 95 por ciento de que el tiempo de flama media está entre 9.8025 y 9.8921 segundos.

Ejemplo: 4.5

En un proceso de baño químico utilizado para grabar tarjetas de circuito impreso, se están comprobando dos diferentes catalizadores para determinar si quieren

diferentes tipos de inmersión para remover cantidades idénticas de mineral foto resistente. Se efectuaron 12 baños con el catalizador 1, resultando un tiempo inmerso medio de muestra de $\bar{X}_1 = 24.6$ minutos y una desviación estándar de $s_1 = .85$ minutos. Con el catalizador 2 se efectuaron 15 baños, siendo el tiempo de inmersión medio de $\bar{X}_2 = 22.1$ y una desviación estándar de $s_2 = .98$ minutos. Determinar el intervalo de confianza del 95 por ciento en la diferencia de dos medias $\mu_1 - \mu_2$ asumiendo que las dos desviaciones de ambas poblaciones son iguales.

Solución:

Determinar la estimación estimada por medio de la relación siguiente:

$$\begin{aligned} S_p &= \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \\ &= \frac{11(0.85)^2 + 14(0.98)^2}{12 + 15 - 2} \\ &= 0.925 \end{aligned}$$

Puesto que $\alpha = 0.05$, se lee de tabla del Apéndice E $t_{\alpha/2, n_1 + n_2 - 2} = t_{0.025, 25} = 2.060$, se puede calcular los límites de confianza inferior y superior del 95 por ciento como:

$$\begin{aligned} L &= \bar{X}_1 - \bar{X}_2 - t_{\alpha/2, n_1 + n_2 - 2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\ &= 24.6 - 22.1 - 2.060(0.925) \sqrt{\frac{1}{12} + \frac{1}{15}} \\ &= 1.76 \text{ min.} \end{aligned}$$

y

$$U = \bar{X}_1 - \bar{X}_2 + t_{\alpha/2, n_1 + n_2 - 2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 24.6 - 22.1 + 2.060(0.925) \sqrt{\frac{1}{12} + \frac{1}{15}}$$

$$= 3.24 \text{ min.}$$

y para el intervalo de confianza de dos colas, se tiene:

$$1.76 \text{ min.} \leq \mu_1 - \mu_2 \leq 3.24 \text{ min.}$$

Se está el 95 por ciento seguro de que el catalizador 1 requiere un tiempo de inmersión que está entre 1.76 minutos y 3.24 minutos más largo que el requerido por el catalizador 2.

Ejemplo 4.6

Según los datos de la tabla 4.8 concernientes en el tiempo (segundos) de reacción de saponificación en paralelo de dos líneas de jabón (1 cajas) con dos materias primas diferentes, se tomaron de 14 medidas de tiempo. Determinar los intervalos de confianza para 90 por ciento para μ_D .

Tabla 4.8. Tiempo en segundos para la producción de jabón en forma paralela.

Medidas	Proceso		Diferencia
	1	2	
1	37.0	17.8	19.2
2	25.8	20.2	5.6
3	16.2	16.8	-0.6
4	24.2	41.4	-17.2
5	22.0	21.4	0.6
6	33.4	38.4	-5.0
7	23.8	16.8	7.0
8	58.2	32.2	26.0
9	33.6	27.8	5.8
10	24.4	23.2	1.2
11	23.4	29.6	-6.2
12	21.2	20.6	0.6
13	36.2	32.2	4.0
14	29.8	53.8	-24.0
			17.0

Solución:

Como \bar{D} y S_D son la media y desviación estándar de muestra de las diferencias D_i , siendo 1.21 y 12.68 respectivamente. El intervalo de confianza de 90 por ciento se tiene $\alpha=0.1$, entonces por tabla del Anexo E se tiene que $t_{0.05,13}=1.771$ y para $\mu_D = \mu_1 - \mu_2$ se tiene el intervalo por medio de la siguiente ecuación:

$$\bar{D} - t_{\alpha/2, n-1} \frac{S_D}{\sqrt{n}} \leq \mu_D \leq \bar{D} + t_{\alpha/2, n-1} \frac{S_D}{\sqrt{n}}$$

$$1.21 - 1.771(12.68)/\sqrt{14} \leq \mu_D \leq 1.21 + 1.771(12.68)/\sqrt{14}$$

$$-4.79 \leq \mu_D \leq 7.21$$

Nótese que el intervalo de confianza incluye el cero. Esto implica que, en el nivel de confianza del 90 por ciento, los datos no soportan dos cajas de jabón tengan diferentes tiempos de saponificación medios. Esto es, el valor cero no es inconsistencia con los datos observados.

Ejemplo 4.7

Un productor de refrescos está interesado en el funcionamiento uniforme de la maquina que se utiliza para llenar latas. En particular, está en que la desviación estándar σ del proceso sea menor de 0.2 onzas líquidas; en otro caso, habrá un porcentaje más alto que el tolerable de latas que no estarán completamente llenas. Suponiendo que el volumen de llenado se distribuye aproximadamente en forma normal. Determinar el intervalo de confianza superior del 95 por ciento, cuando una muestra aleatoria de 20 latas resulta con una varianza de muestra de $s^2=0.0225$ (onzas líquidas)².

Solución:

Esto es un requerimiento del intervalo de confianza superior del 95 por ciento. Como $\alpha=0.05$ entonces se lee en la tabla del Anexo D $\chi^2_{95,19} = 10.117$ y utilizando la relación siguiente:

Como \bar{D} y S_D son la media y desviación estándar de muestra de las diferencias D_i , siendo 1.21 y 12.68 respectivamente. El intervalo de confianza de 90 por ciento se tiene $\alpha=0.1$, entonces por tabla del Anexo E se tiene que $t_{0.05,13}=1.771$ y para $\mu_D = \mu_1 - \mu_2$ se tiene el intervalo por medio de la siguiente ecuación:

$$\bar{D} - t_{\alpha/2, n-1} \frac{S_D}{\sqrt{n}} \leq \mu_D \leq \bar{D} + t_{\alpha/2, n-1} \frac{S_D}{\sqrt{n}}$$

$$1.21 - 1.771(12.68)/\sqrt{14} \leq \mu_D \leq 1.21 + 1.771(12.68)/\sqrt{14}$$

$$-4.79 \leq \mu_D \leq 7.21$$

Nótese que el intervalo de confianza incluye el cero. Esto implica que, en el nivel de confianza del 90 por ciento, los datos no soportan dos cajas de jabón tengan diferentes tiempos de saponificación medios. Esto es, el valor cero no es inconsistencia con los datos observados.

Ejemplo 4.7

Un productor de refrescos está interesado en el funcionamiento uniforme de la maquina que se utiliza para llenar latas. En particular, está en que la desviación estándar σ del proceso sea menor de 0.2 onzas liquidas; en otro caso, habrá un porcentaje más alto que el tolerable de latas que no estarán completamente llenas. Suponiendo que el volumen de llenado se distribuye aproximadamente en forma normal. Determinar el intervalo de confianza superior del 95 por ciento, cuando una muestra aleatoria de 20 latas resulta con una varianza de muestra de $s^2=0.0225$ (onzas liquidas)².

Solución:

Esto es un requerimiento del intervalo de confianza superior del 95 por ciento. Como $\alpha=0.05$ entonces se lee en la tabla del Anexo D $\chi^2_{95,19} = 10.117$ y utilizando la relación siguiente:

$$\sigma^2 \leq \frac{(n-1)s^2}{\chi_{95,19}^2} = \frac{19 * 0.0225}{10.117} = 0.0423 \text{ (onzas liquidas)}^2$$

Aplicando raíz cuadrada a ambos miembros se tiene

$$\sigma \leq 0.21 \text{ onzas liquidas}$$

El nivel de confianza del 95 por ciento, los datos no soportan el requerimiento de que la desviación estándar del proceso sea menor de 0.20 onzas liquidas.

Ejemplo 4.8

En un proceso de grabado por baño químico descrito en el ejemplo 4.5. Recuérdese que se están comparando dos catalizadores para medir la eficacia en los tiempos de inmersión para tarjetas de circuito impreso. $n_1=12$ baños se efectuaron con el catalizador 1 y $n_2=15$ baños se realizaron con el catalizador 2, resultando $s_1=8.5$ minutos y $s_2=0.98$ minutos. Determinar el intervalo de confianza de 90 por ciento respecto al cociente de varianzas $\frac{s_1^2}{s_2^2}$.

Solución:

Primero se lee en la tabla del Anexo F para un intervalo de confianza del 90 por ciento, dando: $F_{0.95,14,11} = \frac{1}{F_{0.05,11,14}} = 0.39$ y $F_{0.05,14,11} = 2.74$. Luego utilizando la relación

siguiente:

$$\frac{s_1^2}{s_2^2} F_{1-\alpha/2, n_2-1, n_1-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} F_{\alpha/2, n_2-1, n_1-1}$$

$$\frac{(0.85)^2}{(0.98)^2} 0.39 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{(0.85)^2}{(0.98)^2} 2.74$$

$$0.29 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq 2.06$$

Como el intervalo de confianza incluye la unidad, se podría no requerir desviaciones estándar de los tiempos de inmersión para los dos catalizadores sean diferentes en el nivel de confianza del 90 por ciento.

Ejemplo 4.9

En una muestra aleatoria 75 laminas, 12 tienen un acabado superficial que es más riguroso que lo permitido por las especificaciones. Determinar un intervalo de confianza del 95 por ciento de dos lados para p .

Solución:

Como $\hat{p} = \frac{x}{n} = \frac{12}{75} = 0.16$ es la población que excede de la rugosidad especificada, luego si el 95 por ciento es el nivel de confianza se tiene que $Z_{0.025} = 1.96$ (leído de la tabla del Anexo C), ahora p se determina por medio de la siguiente relación:

$$\hat{p} - Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.16 - 1.96 \sqrt{\frac{0.16 * 0.84}{75}} \leq p \leq 0.16 + 1.96 \sqrt{\frac{0.16 * 0.84}{75}}$$

$$0.08 \leq p \leq 0.24$$

Entonces el intervalo de confianza del 95 por ciento de que la rugosidad excedida caiga entre 0.08 y 0.24, siendo un 5 por ciento que se salga de este intervalo de confianza la rugosidad excedida.

Ejemplo 4.10

Considerando los datos del ejemplo 4.8. Supóngase que se efectúa una modificación en el proceso de acabado de la superficie y subsecuentemente se obtiene una segunda muestra aleatoria de 85 laminas. El número de unidades defectuosas en la segunda muestra es 10. Determinar el intervalo de confianza aproximado del 95 por ciento y concluir si en el proceso ocurrió mejora.

Solución:

La probabilidad de unidades defectuosas en el proceso modificado es: $p_2 = 10/85 = 0.12$ con esto calcular el intervalo de confianza del 95 por ciento mediante la relación siguiente:

$$\hat{p}_1 - Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \leq p_1 - p_2 \leq \hat{p}_1 - \hat{p}_2 + Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$0.16 - 0.12 - 1.96 \sqrt{\frac{0.16 * 0.84}{75} + \frac{0.12 * 0.88}{85}} \leq p_1 - p_2 \leq 0.16 - 0.12 + 1.96 \sqrt{\frac{0.16 * 0.84}{75} + \frac{0.12 * 0.88}{85}}$$

$$-0.07 \leq p_1 - p_2 \leq 0.15$$

Este intervalo incluye el cero, de modo que, con base en los datos de la muestra parece poco probable que las modificaciones realizadas en el proceso de acabado de la superficie hayan reducido la proporción de las laminas defectuosas que están produciendo.

Ejemplo 4.11

Supóngase que se está interesado en el coeficiente a la transferencia de calor media del acero AMCO. Específicamente, se está interesado en decidir si el coeficiente de transferencia de calor media es o no 40.0 Btu/hr pie. °F. Se puede expresar de la manera formal como

$$H_0: \mu = 40.0 \text{ Btu/hr.pie. } ^\circ\text{F}$$

$$H_1: \mu \neq 40.0 \text{ Btu/hr.pie. } ^\circ\text{F}$$

Al enunciado $H_0: \mu = 40.0 \text{ Btu/hr.pie. } ^\circ\text{F}$ de la ecuación 5. se le llama *hipótesis nula* (H_0), y al enunciado $H_1: \mu \neq 40.0 \text{ Btu/hr.pie. } ^\circ\text{F}$, *hipótesis alternativa* (H_1). Puesto que la hipótesis alternativa especifica valores de μ que podría ser más grandes o pequeños que $40.0 \text{ Btu/hr pie. } ^\circ\text{F}$, se le llama *hipótesis alternativa de dos lados*. En algunas situaciones, se puede estar interesados en formular una hipótesis alternativa de un lado, como en

$$H_0: \mu = 40.0 \text{ Btu/hr.pie. } ^\circ\text{F}$$

$$H_1: \mu > 40.0 \text{ Btu/hr.pie. } ^\circ\text{F}$$

Ejemplo 4.12

Se está estudiando el flujo quemado de un combustible. Las especificaciones requieren que el flujo medio quemado sea $40 \text{ cm}^3/\text{s}$. Además, supóngase que se sabe que la desviación estándar del flujo quemado es aproximadamente $2 \text{ cm}^2/\text{s}$. El experimentador decide especificar una probabilidad de error de tipo I $\alpha = 0.05$, y el basará la prueba en una muestra aleatoria de tamaño $n = 25$. ¿El experimentador rechazará la hipótesis?

Solución:

Planteamientos de hipótesis:

$$H_0: \mu = 40 \text{ cm}^3/\text{s}$$

$$H_1: \mu \neq 40 \text{ cm}^3/\text{s}$$

$$\text{Estadístico de prueba: } Z_0 = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Ejemplo 4.13

Un gerente de planta de una fábrica enlatadora de jugos de naranjas está interesado en comparar el rendimiento de dos diferentes líneas de producción. Como la línea número 1 es relativamente nueva, sospecha que el número de cajas que se producen al día es mayor que el correspondiente a la vieja línea 2. Se toman al día al azar durante diez días para cada línea, encontrándose que la media de la línea 1 fue 824.9 cajas por día y para la línea 2 fue 818.6 cajas por día. De la experiencia con la operación de este tipo de equipo se sabe que la varianza para la línea 1 es 40 y para la línea 2 es 50. Probar:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 > \mu_2$$

$$\text{Estadístico de prueba: } Z_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\text{Entonces: } Z_0 = \frac{824.9 - 818.6}{\sqrt{\frac{40}{10} + \frac{50}{10}}} = 2.10$$

Criterio para rechazo: $Z_0 > Z_{\alpha/2}$ (95 por ciento tabla del Anexo C)

$$\text{Entonces: } Z_{0.05} = 1.645$$

$$\text{Por lo tanto: } 2.10 > 1.645$$

Con esto se rechazaría la H_0 y se concluye que el número medio de cajas producidas diariamente por la línea nueva de producción es mayor que el número medio de cajas producidas por la vieja línea.

Ejemplo 4.14

La resistencia al rompimiento de una fibra textil es una variable aleatoria distribuida normalmente. Las especificaciones requieren que la resistencia media al rompimiento deba igualar el valor de 150 psi. Al fabricante le gustaría detectar cualquier desviación significativa respecto a este valor. Una muestra aleatoria de 15 especímenes de prueba se selecciona y se determina a partir de los datos de la misma con media de muestra de 152.18 psi y varianza de muestra de 16.63. Probar:

$$H_0: \mu = 150 \text{ psi}$$

$$H_1: \mu \neq 150 \text{ psi}$$

$$\text{Estadístico de prueba: } t_0 = \frac{\bar{X} - \mu_0}{S / \sqrt{n}}$$

$$\text{Entonces: } t_0 = \frac{152.18 - 150}{\sqrt{16.63/15}} = 2.07$$

Criterio para rechazo: $|t_0| > t_{\alpha/2, n-1}$ (95 por ciento leído de tabla del Anexo E)

$$\text{Entonces: } t_{0.025, 14} = 2.145$$

$$\text{Por lo tanto: } 2.07 > 2.145$$

No hay evidencia suficiente para rechazar la hipótesis de que $\mu = 150$ psi.

Ejemplo 4.15

Se están analizando dos catalizadores para determinar como afectan la producción media de un proceso químico. Específicamente, se está empleando el catalizador 1, pero el catalizador 2 es aceptable. Puesto que el catalizador 2 es más barato, si no cambia la producción del proceso, debe adoptarse. Los datos de

la planta piloto producen $n_1=8$, con media de 91.73 y varianza de 3.89, $n_2=8$, con media de 93.75 y varianza de 4.02. Probar:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

$$\text{Estadístico de prueba: } t_0 = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{y } s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$\text{Entonces: } s_p^2 = \frac{7 * 3.89 + 7 * 4.02}{14} = 3.96$$

$$t_0 = \frac{91.73 - 93.75}{1.99 \sqrt{\frac{1}{8} + \frac{1}{8}}} = -2.03$$

Criterio para rechazo: $|t_0| > t_{\alpha/2, n_1+n_2-2}$ (95 por ciento leído de tabla del Anexo D)

$$\text{Entonces: } t_{0.025, 14} = -2.145$$

Por lo tanto: $-2.03 > -2.145$ (ligeramente)

En consecuencia, $H_0: \mu_1 = \mu_2$ no puede rechazarse. Esto es, no se tiene la suficiente evidencia para concluir que el catalizador 2 resulta en una producción media que difiera de la producción media cuando se emplea el catalizador 1.

Ejemplo 4.16

Un fabricante de aceite comestible prueba dos productos de aceites para determinar si ellos contienen colesterol. La ingeniería de desarrollo ha obtenido los

siguientes datos. Cuando se asume que ambas poblaciones son normales con varianzas de población diferentes.

Producto	Tamaño de muestra	Media	Varianza
1	15	24.2	10
2	10	23.9	20

Probar:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Solución:

$$\text{Estadístico de prueba: } t_0^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\text{Entonces: } t_0^* = \frac{24.2 - 23.9}{\sqrt{\frac{10}{15} + \frac{20}{10}}} = 0.18$$

$$\text{Calculo de grados de libertad: } \nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 + 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 + 1}} - 2 = \frac{\left(\frac{10}{15} + \frac{20}{10}\right)^2}{\frac{(10/15)^2}{16} + \frac{(20/10)^2}{11}} - 2 = 16$$

Criterio de rechazo: $|t_0| > t_{\alpha/2, \nu}$ (95 por ciento leer en la tabla del Apéndice D)

$$\text{Entonces: } t_{0.025, 16} = 2.120$$

0.18 no es mayor que 2.120

Entonces la hipótesis nula no se puede rechazar.

Ejemplo 4.17

Considérese la maquina descrita en el ejemplo 5.7, la cual se utiliza para llenar latas de refresco. Si la varianza del volumen de llenado excede .02 (onzas liquidas)², entonces un gran porcentaje inaceptable de latas no se llenará lo suficiente. El embotellador esta interesado en probar la hipótesis.

$$H_0 : \sigma^2 = .02$$

$$H_1 : \sigma^2 > .02$$

cuando se toma una muestra aleatoria de n=20 latas produce una varianza de muestra de 0.0225.

Solución:

$$\text{Estadístico de prueba: } \chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

$$\text{Entonces: } \chi_0^2 = \frac{(20-1)0.0225}{0.02} = 21.38$$

Criterio para rechazo: $\chi_0^2 > \chi_{\alpha, n-1}^2$ (para un nivel de confianza del 95 por ciento)

Entonces: $\chi_{0.05, 19}^2 = 30.14$ (leído en tabla del Anexo D).

Se concluye que hay suficiente evidencia de que la varianza del llenado excede 0.02 (onzas liquidas)².

Ejemplo 4.18

Se emplea el grabado químico para remover cobre de tarjetas de circuito impreso.

X_1 y X_2 representan las producciones del proceso cuando se utilizan los

concentrados diferentes. Asumiendo que las varianzas de ambas poblaciones son iguales. Supóngase que se desea probar

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Para dos muestras de tamaños $n_1=n_2=8$, con varianzas de población 1 de 3.89 y para población 2 de 4.02.

Solución:

$$\text{Estadístico de prueba: } F_0 = \frac{s_1^2}{s_2^2}$$

$$\text{Entonces: } F_0 = \frac{3.89}{4.02} = 0.97$$

Criterio para rechazo: $F_0 > F_{\alpha/2, n_1-1, n_2-1}$ (para un nivel de confianza del 95 por ciento)

$$\text{Entonces: } F_{0.025, 7, 7} = 4.99 \quad (\text{leído de tabla del Apéndice F}).$$

Por lo tanto no se puede rechazar la hipótesis nula, y se concluye que no hay suficiente evidencia de que la concentración afecta la varianza de la producción

Ejemplo 4.19

Dados los siguientes datos presentados en la tabla 4.9 que los constituyen una muestra aleatoria de 15 medidas de la calificación de octanaje de cierto tipo de gasolina.

Probar:

$$\bar{\mu} = 98$$

$$\bar{\mu} > 98$$

Para el nivel de significación de 0.01.

Tabla 4.9. Datos del contenido de octanajes en diferentes mezclas de gasolina.

Medida	Contenido de octanaje	Diferencia ($X_i - 98$)	Signo
1	99	+1.0	+
2	102.3	+4.3	+
3	99.8	+1.8	+
4	100.5	+2.5	+
5	99.7	+1.7	+
6	96.2	-1.8	-
7	99.1	+1.1	+
8	102.5	+4.5	+
9	103.3	+5.3	+
10	97.4	-0.6	-
11	100.4	+2.4	+
12	98.9	+0.9	+
13	98.0	0.0	
14	101.6	+3.6	+
15	98.3	+0.3	+

Solución:

Dado que uno de los valores muestrales es igual a 98 y debe descartarse, el tamaño de la muestra para la prueba de los signos es de solo $n=14$.

Hipótesis nula: $\mu = 98.0$ ($p = 0.5$)

Hipótesis alternativa: $\mu > 98.0$ ($p > 0.5$)

Criterio: El criterio puede basarse en el número de los signos de más o el número de los signos menos. Usando el número de signos más, denotado por x , rechazar la hipótesis nula si la probabilidad de obtener x o más signos es menor que o igual a 0.01.

Así, $x=12$, y la tabla 4.9. muestra que para $n=14$ y $p=0.5$ la probabilidad de $X \geq 12$ es de $1 - 0.9935 = 0.0065$.

Decisión: Dado que 0.0065 es menor que 0.01, la hipótesis nula debe rechazarse; se concluye que la calificación mediana de octanaje del tipo dado de gasolina excede de 98.0.

Ejemplo 4.20

Según el proceso descrito en el ejemplo 4.18. probar si se rechaza la hipótesis nula, dando un tratamiento de los valores por la prueba de Wilcoxon del rango con signo.

Solución:

La suma de los rangos positivos es $R^+ = 1.0 + 4.3 + 1.8 + 2.5 + 1.7 + 1.1 + 4.5 + 5.3 + 2.4 + 3.6 + 0.3 + 0.9 = 29.4$ y la suma de rangos negativos es $R^- = 1.8 + 0.6 = 2.4$.

Por lo tanto $R = \min(29.4, 2.4) = 2.4$. De la tabla del Anexo K con $n=14$ y $\alpha=0.01$, se encuentra el valor crítico $R_\alpha = 12$

Decisión: *como R es mucho menor que R_α , entonces se rechaza la hipótesis nula. Se concluye que la calificación mediana de octanaje del tipo dado de gasolina excede de 98.0.*

Ejemplo 4.21

Supóngase que en un estudio de rocas sedimentarias se obtuvieron los siguientes diámetros (en milímetros) de dos tipos de arena. ¿Probar si las dos poblaciones son iguales por la prueba de Wilcoxon de la suma de rango si es probable que una produzca observaciones más grandes que la otra?

Arena 1: 0.63, 0.17, 0.35, 0.49, 0.18, 0.43, 0.12, 0.20, 0.47, 1.36, 0.51, 0.45, 0.84, 0.32, 0.40.

Arena 2: 1.13, 0.54, 0.96, 0.26, 0.39, 0.88, 0.92, 0.53, 1.01, 0.48, 0.89, 1.07, 1.11, 0.58.

Ordenando:

Arena	Tamaño (milímetros)	Rango
1	0.12	1
1	0.17	2
1	0.18	3
1	0.20	4
2	0.26	5
1	0.32	6
1	0.35	7
2	0.39	8
1	0.40	9
1	0.43	10
1	0.45	11
1	0.47	12
2	0.48	13
1	0.49	14
1	0.51	15
2	0.53	16
2	0.54	17
2	0.58	18
1	0.63	19
1	0.84	20
2	0.88	21
2	0.89	22
2	0.92	23
2	0.96	24
2	1.01	25
2	1.07	26
2	1.11	27
2	1.13	28
1	1.36	29

La suma de los rangos de la arena 1 es

$$R1=1+2+3+4+6+7+9+10+11+12+14+15+19+20+29=162$$

$$R2=5+8+13+16+17+18+21+22+23+24+25+26+27+28=273$$

Hipótesis nula: las poblaciones son idénticas

Hipótesis alternativa: las poblaciones no son iguales

Nivel de significación: $\alpha=0.01$

Criterio de rechazo: Rechazar la hipótesis nula si $Z < -2.575$ o $Z > 2.575$, donde Z esta dada por la formula anterior.

Cálculos: con $n1=15$, $n2=14$ y $\alpha=0.01$.y ya se encontró que $R1=162$ (esto para tamaños muestras mayores a 28)

$$U1 = R1 - \frac{n_1(n_1 - 1)}{2} = 42$$

$$\mu_{U1} = \frac{n_1 n_2}{2} = 105$$

$$\sigma_{U1} = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12} = 525$$

entonces

$$Z = \frac{U1 - \mu_{U1}}{\sigma_{U1}} = -2.75$$

Decisión: Dado que $Z = -2.75$ es menor que -2.575 , la hipótesis nula debe rechazarse; se concluye que existe diferencia en las poblaciones de tamaño de gránulos.

Ejemplo 4.22

En *Design and Analysis of Experiments*, 2ª. Edición (John Wiley & Sons, 1984), presenta datos de un experimento en la cual cinco diferentes niveles de contenido de algodón en una fibra sintética se probaron para determinar si dicho contenido tiene algún efecto en la resistencia a la tensión de la fibra. Los datos de la muestra y el rango de este experimento se muestran en la tabla 4.10. Probar:

$$K \geq \chi_{\alpha, a-1}^2$$

Tabla 4.10. Datos y rangos para el experimento de la prueba de tensión

Porcentaje de algodón									
15		20		25		30		35	
Y_{1j}	R_{1j}	Y_{2j}	R_{2j}	Y_{3j}	R_{3j}	Y_{4j}	R_{4j}	Y_{5j}	R_{5j}
7	2.0	12	9.5	14	11.0	19	2.05	7	2.0
7	2.0	17	14.0	18	16.5	25	25.0	10	5.0
15	12.5	12	9.5	18	16.5	22	23.0	11	7.0
11	7.0	18	16.5	19	20.5	19	20.5	15	12.5
9	4.0	18	16.5	19	20.5	23	24.0	11	7.0
R_i	27.5		66.0		85.0		113.0		33.5

Solución:

Estadística de prueba:

$$S^2 = \frac{1}{N-1} \left[\sum_{i=1}^a \sum_{j=1}^m R_{ij}^2 - \frac{N(N+1)^2}{4} \right] = \frac{1}{24} \left[5497.79 - \frac{25(26)^2}{4} \right] = 53.03$$

y

$$K = \frac{1}{S^2} \left[\sum_{i=1}^a \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4} \right] = \frac{1}{53.03} \left[5245.0 - \frac{25(26)^2}{4} \right] = 19.25$$

Criterio para rechazo: $K \geq \chi_{\alpha, a-1}^2$ (para un nivel de significancia del 99 por ciento)

Entonces: $\chi_{0.01,4}^2 = 13.24$ (por tabla del Anexo D).

Decisión: Dado que K es mayor que la ji cuadrada tabulada, la hipótesis se rechaza y se concluyen que los tratamientos difieren.

4.5 EJERCICIOS PROPUESTOS

- 4.1 Un ingeniero en proceso de una industria cementera analiza la resistencia a la compresión de un tipo de cemento. Estas se distribuye aproximadamente en forma normal con varianza $\sigma^2=1000$ (psi²). Una muestra aleatoria de 12 especímenes tiene una resistencia media a la compresión de $\bar{x} = 3250$ psi.
- Construya un intervalo de confianza de dos lados de 95 por ciento respecto a la resistencia media a la compresión.
 - Construya un intervalo de confianza de dos lados de 99 por ciento respecto a la resistencia media a la compresión. Compare el ancho de este intervalo de confianza con el ancho del encontrado en la parte a).

4.2 Se emplean dos maquinas para llenar botellas de plásticos con detergentes para lavar platos. Se tienen como datos de desviaciones estándar del volumen llenado son $\sigma_1=0.15$ onzas de liquido y $\sigma_2=0.18$ onzas liquidas para las dos maquinas, respectivamente. Se selecciona dos muestras aleatorias de n_1 botellas de la maquina 1 y $n_2=10$ botellas de la maquina 2, y las media de muestras de los volúmenes de llenado son $\bar{x}_1 = 30.87$ onzas liquidas y $\bar{x}_2 = 30.68$ onzas liquidas.

- Construya un intervalo de confianza de dos lados del 90 por ciento con respecto a la diferencia de volúmenes de llenado.
- Construya un intervalo de confianza de dos lados del 95 por ciento respecto a la diferencia de volúmenes de llenado. Compare el ancho de este intervalo con el intervalo de la parte a).
- Construya un intervalo de confianza superior del 95 por ciento respecto a las diferencias de medidas de volumen de llenado.

4.3 Dos formulaciones diferentes de gasolina sin plomo se están probando para estudiar sus números de octanaje. La varianza para la formulación 1 es de $\sigma_1^2=1.5$ y para la formulación 2 $\sigma_2^2=1.2$. Se prueban dos muestras aleatorias de tamaño $n_1=15$ y $n_2=20$, y los números de octanaje medios son $\bar{x}_1 = 99.6$ y $\bar{x}_2 = 92.5$. Construya un intervalo de confianza de dos lados del 95 por ciento respecto a la diferencia de los números de octanaje.

4.4 Un ingeniero calidad esta probando la resistencia compresiva de concreto. Realiza la prueba con 16 especimenes, y obtiene los siguientes datos:

2216	2237	2249	2204
2225	2301	2281	2263
2318	2255	2275	2295
2250	2238	2300	2217

- Construya un intervalo de confianza de dos lados del 95 por ciento respecto a la resistencia media.

- b. Construya un intervalo de confianza de inferior del 95 por ciento respecto a la resistencia media.
- c. Construya un intervalo de confianza de dos lados del 95 por ciento respecto a la resistencia media suponiendo que $\alpha=36$. Compare este intervalo con el de la parte a).
- 4.5 Una maquina produce barras metálicas que se usan en el sistema de suspensión de un automóvil. Se selecciona una muestra aleatoria de 15 barras y se mide el diámetro en mm. Los datos resultantes se muestran a continuación. Suponga que el diámetro de las barras se distribuye normalmente. Construya un intervalo de confianza de dos lados del 95 por ciento respecto al diámetro de barra medio.

8.24	8.23	8.20
8.21	8.20	8.28
8.23	8.26	8.24
8.25	8.19	8.25
8.26	8.23	8.24

- 4.6 Un ingeniero de control de calidad midió el espesor de la pared de 25 botellas de vidrio de dos litros. La media de la muestra fue $\bar{x} = 4.05$ mm y la desviación estándar de la muestra $s=0.08$ mm. Determinar un intervalo de confianza inferior del 90 por ciento respecto al espesor de la pared medio.
- 4.7 Dos muestras aleatorias independientes de tamaños $n_1=18$ y $n_2=20$ se toman de dos poblaciones normales. Las medias de las muestras son $\bar{x}_1 = 200$ y $\bar{x}_2 = 190$. Se sabe que las varianzas son $\sigma_1^2 = 15$ y $\sigma_2^2 = 12$. Encuentre lo siguiente:
- Un intervalo de confianza de dos lados del 95 por ciento respecto a $\mu_1 - \mu_2$
 - Un intervalo de confianza inferior del 95 por ciento en $\mu_1 - \mu_2$
 - Un intervalo de confianza superior del 95 por ciento en $\mu_1 - \mu_2$.

- 4.8 Se extraen muestras aleatorias de tamaños $n_1=15$ y $n_2=10$ de dos poblaciones normales independientes. Las medias y varianzas de las muestras son $\bar{x}_1 = 300$, $s_1^2 = 16$, $\bar{x}_2 = 190$ y $s_2^2 = 49$. Suponiendo que $\sigma_1^2 \neq \sigma_2^2$, construya un intervalo de confianza de dos lados del 95 por ciento en $\mu_1 - \mu_2$.
- 4.9 Considere los datos en el ejercicio 5.4. Construya lo siguiente:
- Un intervalo de confianza de dos lados del 95 por ciento en σ^2 .
 - Un intervalo de confianza inferior del 95 por ciento en σ^2 .
 - Un intervalo de confianza superior del 95 por ciento en σ^2 .
- 4.10 Construya un intervalo de confianza de dos lados del 95 por ciento respecto al cociente $\frac{\sigma_1^2}{\sigma_2^2}$ utilizando los dos lados en el ejercicio 5.8.
- 4.11 La fracción de productos defectuosos producidos por dos líneas de productos se está analizando. Una muestra aleatoria de 1000 unidades de la línea 1 tiene 10 defectuosas, en tanto que una muestra aleatoria de 1200 unidades de la línea 2 tiene 25 defectuosas. Encuentre un intervalo de confianza del 99 por ciento respecto a la diferencia de unidades defectuosas producidas por las dos líneas.
- 4.12 Se requiere que la resistencia al rompimiento de una fibra utilizada en la fabricación de ropa no sea menor que 160 psi. La experiencia indica que la desviación estándar de la resistencia al rompimiento es de 3 psi. Se prueba una muestra aleatoria de cuatro especímenes y se encuentra que la resistencia al rompimiento es de 158 psi.
- ¿Debe considerarse aceptable la fibra con $\alpha=0.05$?
 - ¿Cuál es la probabilidad de aceptar $H_0: \mu \leq 160$ si la fibra tiene una resistencia al rompimiento verdadera de 165 psi?

- 4.13 Se está estudiando el rendimiento de un proceso químico. De la experiencia previa se sabe que la varianza del rendimiento con este proceso es 5 (unidades de $\sigma^2 = \text{porcentaje}^2$). Los últimos cinco días de operación de la planta han dado como resultado los siguientes rendimientos (en porcentajes): 91.6, 88.75, 90.8, 89.95, 91.3.
- ¿Hay razón para creer que el rendimiento es menor al 90%?
 - ¿Qué tamaño de muestra se requeriría para detectar un rendimiento medio verdadero de 85% con probabilidad de 0.95?
- 4.14 Se emplean dos máquinas para llenar botellas de plástico con un volumen neto de 16.0 onzas. El proceso de llenado puede suponerse normal, con desviaciones estándar de $\sigma_1 = 0.015$ y $\sigma_2 = 0.018$. Ingeniería de calidad sospecha que ambas máquinas llenan hasta el mismo volumen neto, sin importar que este volumen sea o no de 16 onzas. Se toma una muestra aleatoria de la salida de cada muestra.

Máquina 1		Máquina 2	
16.03	16.01	16.02	16.03
16.04	15.96	15.97	16.04
16.05	15.98	15.96	16.02
16.05	16.02	16.01	16.01
16.02	15.99	15.99	16.00

- ¿Piense usted que ingeniería está en lo correcto? Utilice $\alpha = 0.05$
 - Suponiendo tamaño de muestras iguales, que tamaño de muestra se utilizaría para asegurar que $\beta = 0.05$ si la diferencia en las medias reales es 0.075. suponiendo que $\alpha = 0.05$.
- 4.15 Considere los datos del número de octanaje de gasolina en el ejercicio 5.3. Al fabricante le gustaría detectar que la fórmula 2 produce un número de

octanaje más alto que la formula 1. Formule y pruebe una hipótesis apropiada, empleando $\alpha=0.05$.

- 4.16 La vida de almacenamiento de una película fotografica es de interes para el fabricante. Este observa los siguientes datos de vida de almacenamiento para ocho unidades elegidas al azar de la producción actual. Suponga que la vida de almacenamiento se distribuye normalmente.

Días	Días
108	128
134	163
124	159
116	134

¿Hay alguna evidencia relativa a que la vida de almacenamiento media es mayor o igual que 125 dias?

- 4.17 El contenido de titanio de una aleación se está estudiando con la esperanza de incrementar finalmente la resistencia a la tensión. Un análisis de seis calentamientos recientes elegidos al azar produce los siguientes contenidos de titanio.

8.0%	7.7%
9.9%	11.6%
9.9%	14.6%

¿Hay alguna evidencia de que el contenido medio de titanio sea mayor que el 9.5%?

- 4.18 Un ingeniero desea probar la hipótesis de que el punto de fusión de una aleación es 1000 °C. Si el punto de fusión real difiere de éste en más de 20 °C, él debe cambiar la composición de la aleación. Si suponemos que el punto de fusión es una variable aleatoria que se distribuye normalmente, $\alpha=0.05$, y $\sigma=10$ °C, ¿Cuántas observaciones deben efectuarse?

- 4.19 Se están investigando dos métodos para producir gasolina a partir de petróleo crudo. Se supone que el rendimiento de ambos procesos se distribuye normalmente. Los siguientes datos de rendimiento se han obtenido de la planta piloto.

Proceso	Rendimiento %					
1	24.2	26.6	25.7	24.8	25.9	26.5
2	21.0	22.1	21.8	20.9	22.4	22.0

- a. ¿Hay alguna razón para creer que el proceso 1 tiene un rendimiento medio mayor?
 - b. Suponiendo que para adoptar el proceso 1 debe producirse un rendimiento medio que es al menos 5% mayor que el proceso 2, ¿cuáles son sus recomendaciones?
- 4.20 Un nuevo dispositivo de filtrado se instala en una unidad química. Antes de su instalación, una muestra aleatoria produce la siguiente información acerca del porcentaje de impurezas: $\bar{x}_1 = 12.5$, $s_1^2 = 101.17$ y $n_1 = 8$. Después de la instalación, una muestra aleatoria produce $\bar{x}_2 = 10.2$, $s_2^2 = 94.73$ y $n_2 = 9$.
- a. ¿Puede usted concluir que las dos varianzas son iguales?
 - b. ¿El dispositivo de filtrado ha reducido en forma significativa el porcentaje de impurezas?
- 4.21 Una compañía química produce cierta droga cuyo peso tiene una desviación estándar de 4 miligramos. Se ha propuesto un nuevo método de producción de esta droga, aunque están involucrados costos adicionales. La administración autorizará un cambio en la técnica de producción sólo si la desviación estándar del peso en el nuevo proceso es tan pequeña como 3 miligramos, a la compañía le gustaría cambiar los métodos de producción

con una probabilidad de por lo menos 0.90. Suponiendo que el peso se distribuye normalmente y que $\alpha=0.05$, ¿Cuántas observaciones deben efectuarse? Supóngase que los investigadores eligen $n=10$ y obtienen los siguientes datos. ¿Es ésta una buena elección para n ? ¿Cuál debe ser su decisión?.

Gramos	Gramos
16.628	16.630
16.622	16.631
16.627	16.624
16.623	16.622
16.618	16.626

- 4.22 Dos maquinas producen piezas metálicas. Interesa la varianza del peso de estas piezas. Se han colectado los siguientes datos:

Maquina 1	Maquina 2
$n_1=25$	$n_2=30$
$\bar{x}_1 = 0.984$	$\bar{x}_2 = 0.907$
$s_1^2 = 13.46$	$s_2^2 = 9.65$

- Pruebe la hipótesis de que las dos varianzas son iguales. Emplee $\alpha=0.05$
 - Pruebe la hipótesis de que las dos maquinas producen piezas que tienen el mismo peso medio. Use $\alpha=0.05$.
- 4.23 Un diseñador de aviones tiene evidencia teórica de que la pintura del avión reduce la velocidad del mismo a una potencia especificada y una colocación de alerón. Prueba seis aviones consecutivos de la línea de ensamble antes y después de pintarlos. Los resultados se muestran a continuación:

Velocidad Máxima (mph)		
Avión	Pintado	No pintado
1	286	289
2	285	286
3	279	283
4	283	288
5	281	283
6	286	289

¿Sustentan los datos la teoría del diseñador? Emplee $\alpha=0.05$

- 4.24 Un embotellador de refrescos está estudiando la resistencia a la presión interna de botellas no retornables de un litro. Se prueba una muestra aleatoria de 16 botellas y se obtiene la resistencia a la presión. Los datos se muestran adelante. Grafique estos datos en papel de probabilidad normal. ¿Es razonable concluir que la resistencia a la presión se distribuye normalmente?

226.16 psi	211.14 psi
202.20	203.62
219.54	188.12
193.73	224.39
208.15	221.31
195.45	204.55
193.71	202.21
200.81	201.63

- 4.25 Se tomaron diez muestras de un baño de platinado utilizado en un proceso de manufactura electrónica y se determinó el pH del baño. Los valores del pH de la muestra son los siguientes:

7.91, 7.85, 6.82, 8.01, 7.46, 6.95, 7.05, 7.35, 7.25, 7.42

El departamento de ingeniería de proceso cree que el pH tiene un valor medio de 7.0. ¿Indican los datos de la muestra que esta suposición es correcta?. Emplee la prueba de signo para investigar esta hipótesis.

- 4.26 El contenido de titanio en una aleación utilizada en aeronaves determina de manera importante la resistencia. Una muestra de 20 cupones de prueba revela los siguientes contenidos de titanio (en porcentajes):

8.32, 8.05, 8.93, 8.65, 8.25, 8.46, 8.52, 8.35, 8.36, 8.41,
8.42, 8.30, 8.71, 8.75, 8.60, 8.83, 8.50, 8.38, 8.29, 8.46

El contenido medio de titanio debe ser 8.5 por ciento. Emplee la prueba del signo para investigar esta hipótesis.

- 4.27 Considere los datos en el ejercicio 5.26. Suponga que la distribución del contenido de titanio es simétrica y continua. Emplee la prueba de Wilcoxon del rango con signo para probar la hipótesis $H_0: \mu = 8.5$ contra $H_1: \mu \neq 8.5$.
- 4.28 Considere los datos en el ejercicio 5.26. Emplee la aproximación de muestra grande en la prueba Wilcoxon del rango con signo para probar la hipótesis $H_0: \mu = 8.5$ contra $H_1: \mu \neq 8.5$. Suponga que la distribución del contenido de titanio es continua y simétrica.
- 4.29 Un Ingeniero químico debe diseñar una pila para dar la máxima corriente. Él a desarrollado dentro de las restricciones de diseño permisible, dos pilas candidatas y prueba los prototipos de cada una de ellas. Las candidatas resultantes (en microamperios) se muestran enseguida:

Pila 1	251	255	258	257	250	251	254	250	248
Pila 2	250	253	249	256	259	252	260	251	

Emplee la prueba Wilcoxon de la suma de rangos para probar $H_0: \mu_1 = \mu_2$ contra la alternativa $H_1: \mu_1 > \mu_2$.

- 4.30 Un artículo en el *Quality Control Handbook*, 3ª. Edición (McGraw-Hill, 1962) presenta los resultados de un experimento realizado para investigar el efecto de tres métodos diferentes de humidificación relativos a la resistencia a la fractura de bloques de cemento. Los datos se muestran a continuación. ¿Hay una indicación de que el método de humidificación afecte la resistencia a la fractura?

Método de humidificación	Resistencia a la fractura (lbf/pulg ²)				
1	553	550	568	541	537
2	553	599	579	545	540
3	492	530	528	510	571

CAPÍTULO 5

ANÁLISIS DE REGRESIÓN

El análisis de regresión es una técnica estadística para modelar e investigar la relación entre dos o más variables. Por ejemplo, en un proceso químico, supóngase que el rendimiento del producto o el grado de conversión de un proceso, se relaciona con la temperatura de operación del proceso. El análisis de regresión puede emplearse para construir un modelo que exprese el rendimiento como función de la temperatura. El modelo puede luego emplearse para producir el rendimiento en un nivel determinado de temperatura (interpolación o extrapolar un nivel de rendimiento). También puede emplearse en propósitos de optimización o control de procesos.

En el resto del capítulo se hará un estudio minucioso, para modelos de regresión lineal simple (relación de dos variables, con una variable independiente y una variable dependiente), y modelos de regresión lineal múltiple (relación de tres o más variables (una variable respuesta con o más variables causas o independientes)).

5.1 IMPORTANCIAS DE LOS MODELOS DE REGRESIÓN

El análisis de regresión permite analizar la posible relación existente entre la pauta de variabilidad de una variable aleatoria y los valores de una o más variables (aleatorias o no) de las que la primera depende, o puede depender.

El recurso a los modelos de regresión resulta indispensable cuando no es posible fijar previamente los valores a adoptar por las variables explicativas en un determinado experimento de interés, por ejemplo el efecto de la temperatura diaria en el consumo de energía en una instalación. Dado que en estos casos no es posible diseñar un experimento que garantice los efectos de la variable. También

en algunos casos es necesario recurrir a técnicas de regresión en el análisis de información histórica que no fue obtenida a partir de un diseño experimental (por ejemplo, los datos o resultados procedentes del control estadístico de cierto proceso industrial recopilado el último año).

Ya que el objetivo principal de muchas investigaciones estadísticas es hacer predicciones de preferencia con base en ecuaciones matemáticas. Usualmente tales predicciones requieren que se encuentre una fórmula que relacione la variable dependiente (cuyo valor quiere predecirse), con una o más variables independientes. En problemas, donde la o las variables independientes no son controlables. En otros casos puede permitir sustituir un ensayo destructivo por otro no destructivo.

Dentro de las utilidades más comunes de los modelos de regresión, se encuentran:

1. Proyecto y predicción
2. Descripción cuali y cuantitativa entre las variables que intervienen en el fenómeno.
3. Interpolación y extrapolación entre valores de una función
4. Determinación de variables independientes importantes
5. Descubrimiento de las condiciones de funcionamiento óptimas o control del proceso
6. Selección entre modelos alternativos
7. Estimación de coeficientes de regresión particulares.

5.2 EN QUE CONSISTE UN ANÁLISIS DE REGRESIÓN

El análisis de regresión es una técnica estadística para estimar los parámetros de una ecuación matemática que relaciona una determinada variable con un conjunto

de variables. El análisis de lleva a cabo mediante el establecimiento de un modelo cuyos parámetros recogen y cuantifican los efectos que se pretende estudiar.

Dichos parámetros se estiman a partir de los datos disponibles. Algunos autores se refieren al análisis de regresión como TECNICA DE LOS MINIMOS CUADRADOS O DE AJUSTE DE CURVAS.

En general supóngase que hay una sola variable o respuesta , dependiendo que se relaciona con k variables independientes o regresivas (causas), digamos x_1, x_2, \dots, x_k . La variable de respuesta "y" es una variable aleatoria, en tanto que las variables regresivas x_1, x_2, \dots, x_k se miden con error despreciable. Las x_i se llaman variables matemáticas y con frecuencia son controladas por el experimentador. El análisis de regresión también puede utilizarse en situaciones en las que y, x_1, x_2, \dots, x_k , son variables aleatorias distribuidas conjuntamente, tal es el caso cuando los datos se recaban como mediciones diferentes en una unidad experimental común. La relación entre estas variables se caracterizan por medio de un modelo matemático llamado ecuación de regresión.

5.3 FASES DE UN ESTUDIO MEDIANTE MODELOS DE REGRESIÓN

Las fases que deben seguirse en un estudio mediante modelos de regresión, se sintetizan en la Figura 5.1.

5.4 MODELOS DE REGRESIÓN

Los modelos de regresión por la cantidad de variables regresivas (causas), se clasifican en:

1. Regresión simple
2. Regresión múltiple

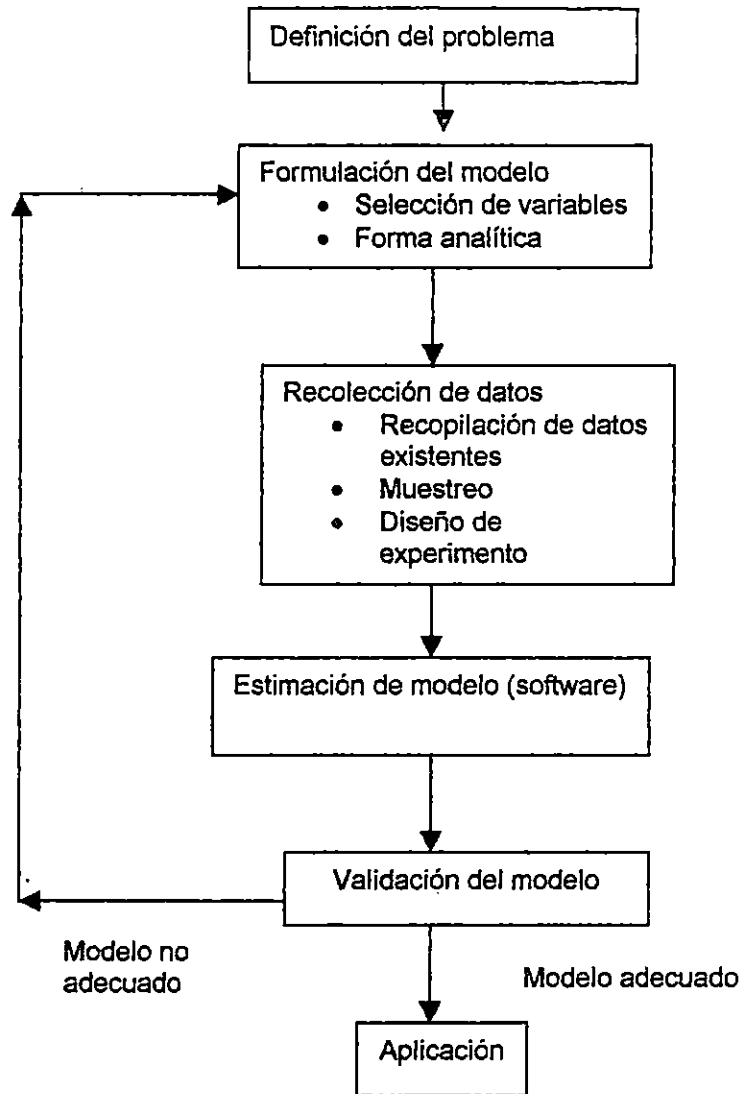


Figura 5.1. Uso de los modelos de regresión.

En la Tabla 5.1. se muestran los datos de un problema dado y el planteamiento de las variables que interactúan en el proceso o fenómenos de intereses

Tabla.5.2. Recolección de datos para el análisis de regresión.

	Variables a explicar (aleatoria)	Variable o factores explicativos (aleatorios o no).
Observaciones	y_1	$X_{11} \dots X_{i1} \dots X_{j1}$
	.	.
	y_j	$X_{1j} \dots X_{ij} \dots X_{jj}$
	.	.
	y_J	$X_{1J} \dots X_{iJ} \dots X_{jJ}$

El modelo fundamental, es:

$$E(Y_j) = f(x_{1j} \dots x_{ij} \dots x_{jj})$$

$$\sigma^2(Y_j) = \text{constante} = g(x_{1j} \dots x_{ij} \dots x_{jj}) \quad 5.1$$

6.5.1. REGRESIÓN LINEAL SIMPLE

Es el caso con en que se desea determinar la relación entre una sola variable regresiva (explicativa) "x" y una variable respuesta "y" (se supone a las demás constantes o se consideran efectos menores que las demás). La variable regresiva "x", se supone como una variable matemática continua, contable por el experimentador. Supóngase que la verdadera relación entre "y" y "x" es una línea recta.

Los pasos del análisis de regresión simple se mencionan a continuación:

1. Representación gráfica de datos. El primer paso para el estudio de relaciones entre variables consiste en trazar una gráfica de los datos (diagrama de dispersión). Con la finalidad siguiente:
 - a. Facilitar información sobre la relación existente entre variables

- b. Sugerir modelos posibles para los datos o transformaciones de datos (linealizar los datos de posibles modelos experimentales o potenciales)
 - c. Poder señalar la existencia de observaciones extemporáneas (anómalas).
 - d. Poder facilitar una indicación de "y" para "X" fija o que esa variabilidad permanece constante para todos los valores de "X".
2. Planteamiento del modelo: La ecuación que relaciona los datos es

$$y = \beta_0 + \beta_1 X + \varepsilon \quad 5.2$$

Donde: β_0 y β_1 son respectivamente la ordenada del origen y la pendiente, son constantes desconocidas de la recta de regresión. ε , es el error aleatorio con media cero y varianza σ^2 , que puede deberse a errores en la medición de y/o a efectos de variables no incluidas en el modelo. Se denota como error de la ecuación matemática de regresión.

3. Estimación de la ecuación de predicción: esto tiene la finalidad de encontrar los estimadores de mínimos cuadrados (b_0, b_1) de los parámetros desconocidos (β_0, β_1)

$$(b_0, b_1) \rightarrow (\beta_0, \beta_1)$$

Entonces:

$$b_0 = \bar{y} - b_1 \bar{x} \quad 5.3$$

y para b_1 , es:

$$b_1 = \frac{S_{xy}}{S_{xx}} \quad 5.4$$

donde:

$$\bar{y} = (1/n) \sum_{i=1}^n y_i$$

$$\bar{x} = (1/n) \sum_{i=1}^n x_i$$

$$S_{xx} = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}$$

$$S_{xy} = \sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n x_i\right) \left(\sum_{i=1}^n y_i\right)}{n}$$

El modelo de regresión simple ajustado es:

$$\hat{y} = \beta_0 + \beta_1 x$$

5.5

Nota: b_1 y b_1 , se denominan estimadores mínimos cuadrados (método de los mínimos cuadrados) que minimizan la suma de las desviaciones elevadas al cuadrado, entre los valores observados y los valores pronosticados para la variable respuesta. Siendo estos los mejores estimadores lineales no sesgados y útiles para evaluar la frecuencia del modelo, ya que son variables aleatorias, puesto que son justamente combinaciones lineales de las y_i

Condiciones que se deben cumplir

1. Las observaciones son independientes
2. La varianza de los errores es constante
3. El modelo lineal promulgado es correcto.

4. Examen de la ecuación del modelo lineal

Para realizar los exámenes del modelo lineal se hará de la siguiente manera:

1. Examen preliminar: Dibujar la ecuación. Aproximadamente la mitad de los puntos deben quedar por encima de la recta y la otra por debajo. Además la ecuación debe pasar exactamente por (\bar{x}, \bar{y}) .
2. Determinación de la significación global del ajuste.
(R^2) llamado coeficiente de determinación y es usado a menudo para juzgar la suficiencia de un modelo de regresión.

$$R^2 = \frac{SS_R}{S_{yy}} = 1 - \frac{SS_E}{S_{yy}} \quad 5.6$$

donde: SS_E se denomina la suma de cuadrados del error y SS_R la suma de regresión de cuadrados y S_{yy} la suma total de cuadrados corregida para "y". Y estas se definen de la siguiente manera:

$$SS_R = \hat{\beta}_1 S_{xy}$$

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}$$

$$SS_E = S_{yy} - SS_R$$

R^2 debe oscilar entre 0 y 1. Cuando más cercano a 1 mayor porcentaje de la variabilidad en los datos es explicada por el modelo.

Cuando una variable explicativa es aleatoria, la raíz cuadrada positiva de R^2 , es el coeficiente de correlación entre las variables "x" y "y".

Para estudiar la hipótesis de la significancia del efecto de la variable explicativa (a nivel poblacional), se utilizan los resultados siguientes:

Dado el modelo matemático supuesto:

$$y = \beta_0 + \beta_1 x$$

Se desea determinar que la variable explicativa influye en la variable dependiente (y), si influye $\beta_1 \neq 0$ y si no influye $\beta_1 = 0$, bajo la suposición que se distribuye normalmente $NID(0, \sigma^2)$.

Si se desea probar que :

$H_0: \beta_1 = 0$, la variable explicativa no influye en la respuesta

$H_1: \beta_1 \neq 0$, la variable explicativa influye en la respuesta.

Test 1.

Prueba estadística

$$t_0 = \frac{\beta_1}{\sqrt{MS_E / S_{xx}}} \quad 5.7$$

Criterio de rechazo:

$$t_0 > t_{\alpha/2, n-2} \quad 5.8$$

Donde: en la tabla 5.2. se muestra el análisis de varianza para la prueba de significancia de la regresión.

Tabla 5.2. Análisis de varianza para probar la significación de la regresión.

Fuente de variación	Suma de cuadrados	Grados de libertad	Media cuadrática	t_0
Regresión	$SS_R = b_1 S_{xy}$	1	MS_R	$\beta_1 / \sqrt{MS_E / S_{xx}}$
Error residual	$SS_R = S_{yy} - b_1 S_{xy}$	n-2	MS_E	
Total de grados	S_{yy}	n-1		

Tes 2. Para $H_0: \beta_1 = 0$, es verdadera, la estadística para corroborar que SSe y SSR son independiente es:

La prueba estadística:

$$F_0 = \frac{MS_R}{MS_E} \quad 5.8$$

Criterio de rechazo:

$$F_0 > F_{\alpha, 1, n-2} \quad 5.9$$

En las Figuras 5.2 y 5.3, se muestra el comportamiento de datos para hipótesis nula aceptada y rechazada respectivamente:

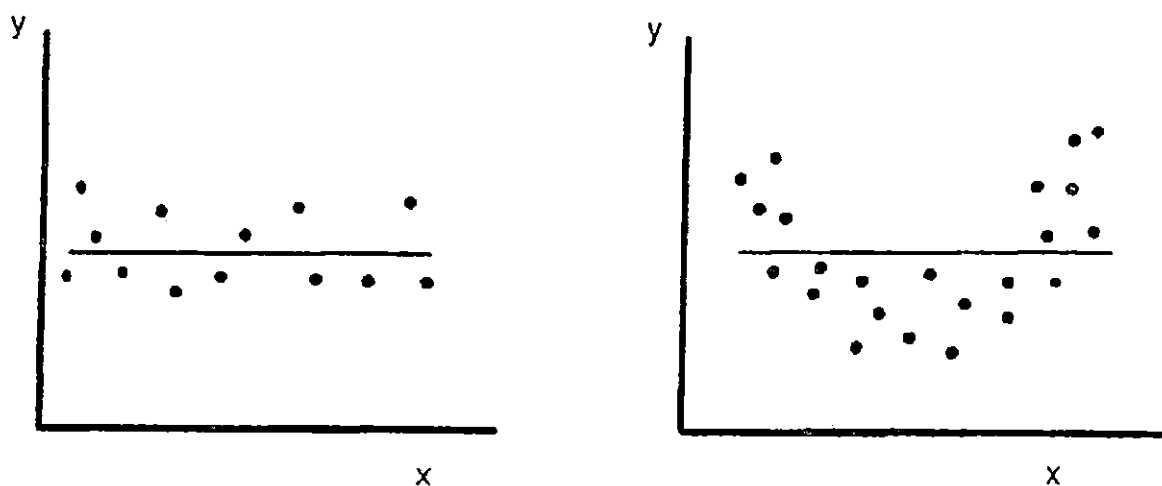


Figura 5.2. La hipótesis $H: \beta_1 = 0$, no se rechaza

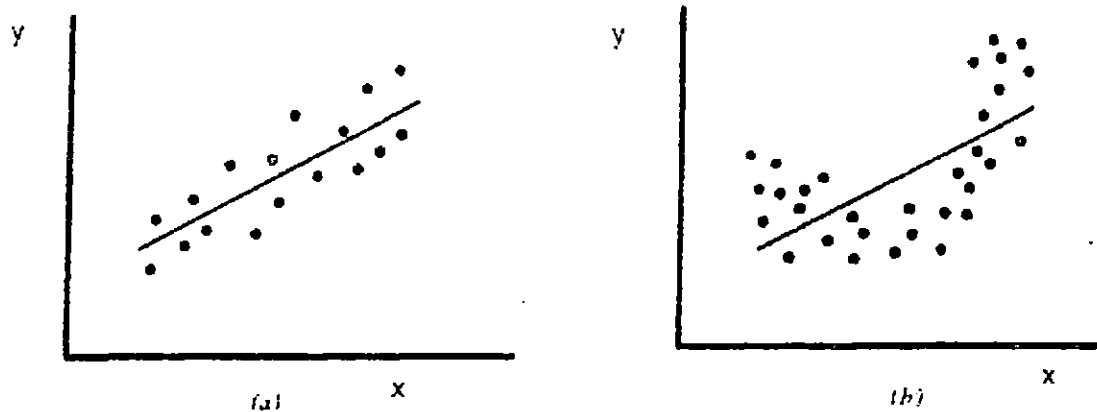


Figura 5.3. La hipótesis $H_0: \beta_1 = 0$ se rechaza.

3. análisis residual: Es con frecuencia útil en la confirmación de la suposición de que los errores son NID $(0, \sigma^2)$ y en la determinación de si los términos adicionales en el modelo serán útiles.

$$e_i = y_i - \hat{Y}_i, i = 1, 2, \dots, n \quad 5.9$$

A menudo es útil graficar los residuos (1), en secuencia de tiempo (si se conoce), (2) contra \hat{y}_i , y (3) contra la variable independiente x .

Estas gráficas suelen verse como una de los cuatro patrones generales de la Figura 5.4. El patrón (a) representa la normalidad en tanto que los patrones (b), (c) y (d), representan anomalías. Además la gráfica (b) la varianza puede incrementarse en el tiempo con la magnitud, la (c) demuestra también desigualdad de varianza y la (d) demuestra insuficiencia del modelo.

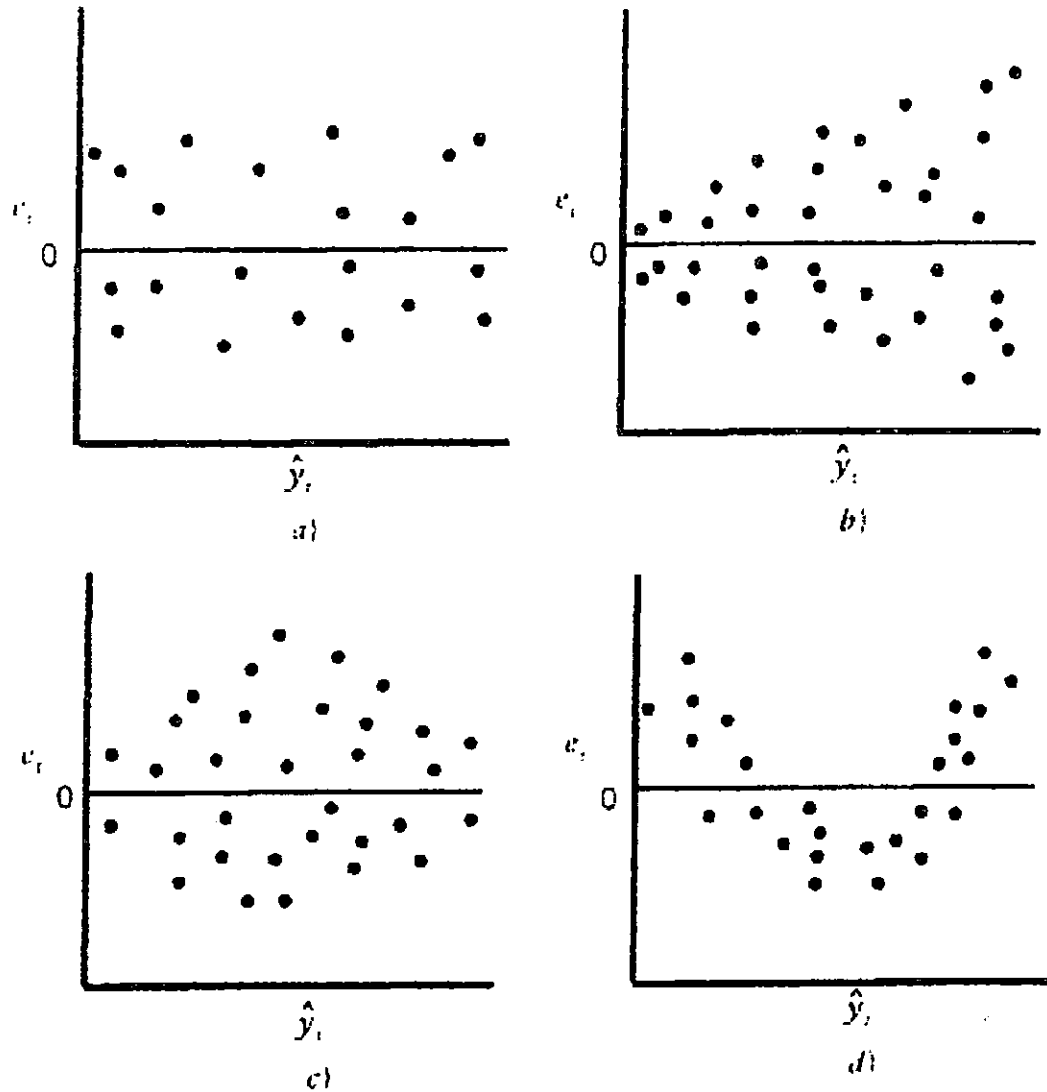


Figura 5.4. Patrones para las gráficas de los residuos. a) Satisfactorio
b) Embudo, c) doble arco, d) no lineal.

4. Prueba de la falta de ajuste: Es una prueba para la validez del modelo asumido tentativo.

Las pruebas de hipótesis a probar son:

H_0 : El modelo ajusta adecuadamente los datos

H_1 : El modelo no ajusta los datos

La prueba consiste en dividir la suma de cuadrados del error o del residuo de los siguientes dos componentes:

$$SS_E = SS_{PE} + SS_{LOF} \quad 5.10$$

Donde: SS_{PE} , es la suma cuadrados atribuibles al error "puro" y SS_{LOF} , es la suma de cuadrado atribuible a la falta de ajuste del modelo. Para calcular SS_{PE} , se debe contar con observaciones repetidas en y para al menos un nivel x. Supóngase que se tienen n observaciones totales que:

Y_{11}, Y_{12}	y_{1n}	Observaciones repetidas en X_1
Y_{12}, Y_{22}	y_{2n}	Observaciones repetidas en X_2
.....
Y_{m1}, Y_{m2}	y_{mn}	Observaciones repetidas en X_m

Donde M, SM los distintos niveles de X. Entonces:

$$SS_{PE} = \sum_{L=1}^m \sum_{U=1}^{n_L} (y_{LU} - \bar{y}_L)^2 \quad 5.11$$

$$SS_{LOF} = SS_E - SS_{PE} \quad 5.12$$

Con grados de libertad de n-m para la suma de cuadrados del error puro y m-2 asociados a la suma de cuadrados para la falta de ajuste.

La estadística de prueba es:

$$F_0 = \frac{SS_{LOF} / m - 2}{SS_{PE} / (n - m)} = \frac{MS_{LOF}}{MS_{PE}} \quad 5.13$$

Criterio de rechazo, es, si

$$F_0 > F_{\alpha, m-2, n-m}$$

5. Intervalos de confianzas

Los intervalos de confianza proporcionan otra medida de fiabilidad de los estimadores. Dado el caso, que en una regresión lineal se obtiene la estimación puntual de la pendiente y la ordenada al origen, para esto es posible obtener estimaciones del intervalo de confianza de estos parámetros. El ancho de estos intervalos de confianza es una medida de la calidad total de la línea de regresión. Si E_i se distribuye normal e independiente, se tiene:

$$(\hat{\beta}_1 - \beta_1) / \sqrt{MS_E / S_{xx}} \text{ y } (\hat{\beta}_0 - \beta_0) / \sqrt{MS_E \left[\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}} \right]} \quad 5.14$$

Se distribuye como t con $n-2$ grados de libertad. En consecuencia, un intervalo de confianza del $100(1-\alpha)$, por ciento en la pendiente b_1 está dado por:

$$\hat{\beta}_1 - t_{\alpha/2, n-2} \sqrt{\frac{MS_E}{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2, n-2} \sqrt{\frac{MS_E}{S_{xx}}} \quad 5.15$$

y para la ordenada al origen es:

$$\hat{\beta}_0 - t_{\alpha/2, n-2} \sqrt{MS_E \left[\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}} \right]} \leq \beta_0 \leq \hat{\beta}_0 + t_{\alpha/2, n-2} \sqrt{MS_E \left[\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}} \right]} \quad 5.16$$

5.5.1 TRANSFORMACIONES A UNA LÍNEA RECTA.

En muchas ocasiones se encuentra que el modelo de regresión de línea recta $y = \beta_0 + \beta_1 x + \varepsilon$, es inapropiado porque la función de regresión verdadera no es lineal. Algunos veces esto se determina en forma visual a partir del diagrama de dispersión, y algunas veces se sabe en principio que el modelo es no lineal debido a las experiencias previas a la teoría básica. En ciertas situaciones una función no lineal puede exponerse como una línea recta. Utilizando una transformación

apropiada. A estos modelos lineales se les denominan lineales intrínsecamente. Estos modelos no lineales pueden ser:

1. Función exponenciales
2. Función potenciales
3. Función recíproca

1. Función exponencial: Esta es una función intrínsecamente lineal, ya que puede transformarse en una línea recta mediante una transformación logarítmica.

Función exponencial:

$$y = \beta_0 e^{x\beta_1} \varepsilon \quad 5.17$$

Transformada queda:

$$\ln y = \ln \beta_0 + \beta_1 X + \ln \varepsilon \quad 5.18$$

2. Función exponencial: También es una función intrínsecamente lineal, ya que puede transformando en una línea recta mediante una transformación logarítmica.

Función potencial:

$$y = \beta_0 X^{\beta_1} \varepsilon \quad 5.19$$

Transformada

$$\ln y = \ln \beta_0 + \beta_1 \ln X + \ln \varepsilon \quad 5.20$$

3. Función recíproca: La transformada de esta función intrínseca, se realiza haciendo un artificio algebraico.

Función potencial:

$$y = \beta_0 + \beta_1 (1/x) + \varepsilon \quad \text{Si } Z = (1/x) \quad 5.21$$

Función transformada

$$y = \beta_0 + \beta_1 z + \varepsilon \quad 5.22$$

Existen particularidades de funciones intrínsecamente lineales, pero se han incrementado las más comunes en razón de ejemplos.

5.6 REGRESIÓN LINEAL MÚLTIPLE

Las áreas de Ingeniería Química e Ingeniería de Alimentos se realizan muchas investigaciones en donde involucran una variable respuesta en función de dos o más variables causas ó independientes (variables regresivas). Tales modelos se denominan *regresión múltiple*. La regresión múltiple es una de las técnicas estadísticas más ampliamente utilizada para este tipo de casos.

En esta sección se presentará las técnicas básicas de la estimación de parámetros, estimación de intervalos de confianza y la suficiencia del modelo.

5.6.1 MODELOS DE REGRESIÓN MÚLTIPLE

El modelo de regresión que involucra más de una variable regresiva (explicativas), se llama modelo de regresión múltiple. Un modelo de regresión múltiple que podría describir esta relación, es:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad 5.23$$

Para un modelo de regresión lineal múltiple para dos variables regresivas. El término "lineal" se emplea debido a la ecuación es la función lineal de los parámetros desconocidos β_0 , β_1 y β_2 . Nótese que el modelo describe un plano en

el espacio bidimensional x_1, x_2 . El parámetro β_0 define la ordenada al origen del plano. Algunas veces se le denomina a β_1 y β_2 coeficiente de regresión parciales, porque β_1 mide el cambio esperado en "y" por cambio unitario en x_1 , cuando x_2 se mantiene constante y β_2 mide el cambio esperado en "y" por cambio unitario en x_2 cuando x_1 se mantiene constante.

En general, la variable dependiente o respuesta "y" puede relacionarse con k variables independientes. En el modelo:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad 5.24$$

Las variables $x_1, x_2, \dots, x_{k-1}, x_k$ puede ser:

- a. La variable original cuantitativas.
- b. Transformaciones de la variable original: polinómicas, logarítmica, etc. Ejemplo : $y = a + bx_1 + cx_1^2$
Donde puede definirse una variable $X_2 = X_1^2$ y se tiene el siguiente modelo lineal: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$
- c. Variables cualitativas con P variantes, deben incluirse en este caso lo que se conoce como variable "dummy" (p-1) cuyos únicos valores posibles son 0 y 1: $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 Z_1 + \beta_3 Z_2$.
- d. Inclusión de interacciones: Es posible que se den efectos interactivos entre las variables explicativas. Hay que incluir interactivo como una nueva variable: por ejemplo: $X_1 * X_2 \equiv X_3$.

En general, en el modelo lineal los residuos ε se distribuye normalmente con media cero y σ^2 constante, es decir $\varepsilon \sim N(0, \sigma^2)$ e independientes. La interpretación de los parámetros en el modelo será:

β_0 = Valor medio de "y" cuando $X_1 = X_2 = \dots X_k = 0$

β_i = Incremento en el valor medio de "y" cuando x_i aumenta en una unidad manteniéndose constantes las restantes variables explicativas.

5.6.2 ESTIMACIÓN DE PARÁMETROS EN MODELOS DE REGRESIÓN.

El método de mínimos cuadrados puede utilizarse para estimar los coeficientes de regresión en la ecuación (5.24). Supóngase que se disponen $n > k$ observaciones, y déjese que x_{ij} denote la observación i -ésima o el nivel de la variable x_i . Los datos aparecerán en la tabla 5.3. Se supone que el término de error en el modelo tiene $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma^2$, y que las ε_i son variables aleatorias no correlacionadas.

Tabla 5.3. Datos para la regresión lineal múltiple.

y	X_1	X_2	X_k
y_1	X_{11}	X_{12}	X_{1k}
y_2	X_{21}	X_{22}	X_{2k}
.	.	.	.
y_n	X_{n1}	X_{n2}	X_{nk}

Se puede escribir el modelo, 6. , en términos de las observaciones como:

$$y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i$$

$$y = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i \quad i = 1, 2, \dots, n \quad 5.25$$

La función de mínimos cuadrados es:

$$L = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 \quad 5.26$$

La función L se minimizará con respecto a $\beta_0, \beta_1, \dots, \beta_k$. Los estimadores de los mínimos cuadrados de $\beta_0, \beta_1, \dots, \beta_k$ debe satisfacer.

$$\left. \frac{\partial L}{\partial \beta_0} \right|_{b_0, b_1, \dots, b_k} = -2 \sum_{i=1}^n \left(y_i - b_0 - \sum_{j=1}^k b_j x_{ij} \right) = 0 \quad 5.27$$

y

$$\left. \frac{\partial L}{\partial \beta_j} \right|_{b_0, b_1, \dots, b_k} = -2 \sum_{i=1}^n \left(y_i - b_0 - \sum_{j=1}^k b_j x_{ij} \right) x_{ij} = 0 \quad j = 1, 2, \dots, k \quad 5.28$$

Al simplificar la ecuación , se obtienen las ecuaciones normales de mínimos cuadrados.

$$\begin{aligned} n\beta_0 + \beta_1 \sum_{i=1}^n x_{i1} + \beta_2 \sum_{i=1}^n x_{i2} + \dots + \beta_k \sum_{i=1}^n x_{ik} &= \sum_{i=1}^n y_i \\ \beta_0 \sum_{i=1}^n x_{i1} + \beta_1 \sum_{i=1}^n x_{i1}^2 + \beta_2 \sum_{i=1}^n x_{i1} x_{i2} + \dots + \beta_k \sum_{i=1}^n x_{i1} x_{ik} &= \sum_{i=1}^n x_{i1} y_i \\ \beta_0 \sum_{i=1}^n x_{ik} + \beta_1 \sum_{i=1}^n x_{ik} x_{i1} + \beta_2 \sum_{i=1}^n x_{ik} x_{i2} + \dots + \beta_k \sum_{i=1}^n x_{ik}^2 &= \sum_{i=1}^n x_{ik} y_i \end{aligned}$$

5.29

Nótese que hay $p = k+1$, ecuaciones normales, una para cada uno de los coeficientes de regresión desconocidos. La solución para las ecuaciones normales serán los estimadores de mínimos cuadrados de los coeficientes de regresión $\beta_0, \beta_1, \dots, \beta_k$.

Es más simple resolver las ecuaciones normales si ellos se expresan en notación de matriz. Se dará ahora un desarrollo matricial de las ecuaciones normales que es afín al desarrollo de la ecuación (5.29).

El modelo en términos de las observaciones, ecuación 6 , puede escribirse en notación matricial como:

$$y = x\beta + \varepsilon \quad 5.30$$

donde:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \quad x = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdot & X_{1k} \\ 1 & X_{21} & X_{22} & \cdot & X_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & X_{n1} & X_{n2} & \cdot & X_{nk} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_k \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_3 \end{pmatrix}$$

Los estimadores de mínimos cuadrados deben satisfacer:

$$\frac{\partial L}{\partial \beta} \Big|_{\beta} = -2x'y + 2x'x\hat{\beta} = 0 \quad 5.31$$

Que se simplifica a:

$$x'x\hat{\beta} = x'y \quad 5.32$$

Las ecuaciones 5.29. Son las ecuaciones normales de los mínimos cuadrados. Ellos son idénticos a las ecuaciones 5.31. Para resolver las ecuaciones normales, multiplíquese ambos lados de la ecuación 5.32. Por la inversa de $X'X$. De modo, el estimador de mínimos cuadrados de β es:

$$\beta = (x'x)^{-1} x'y \quad 5.33$$

Es fácil que la forma matricial de las ecuaciones normales es idénticas a la forma escolar. Al escribir completa la ecuación 5.29, se obtiene:

$$\begin{vmatrix} n & \sum_{i=1}^n X_{i1} & \sum_{i=1}^n X_{i2} & \dots & \sum_{i=1}^n X_{ik} \\ \sum_{i=1}^n X_{i1} & \sum_{i=1}^n X_{i1}^2 & \sum_{i=1}^n X_{i1}X_{i2} & \dots & \sum_{i=1}^n X_{i1}X_{ik} \\ \dots & \dots & \dots & \dots & \dots \\ \sum_{i=1}^n X_{ik} & \sum_{i=1}^n X_{ik}X_{i1} & \sum_{i=1}^n X_{ik}X_{i2} & \dots & \sum_{i=1}^n X_{ik}^2 \end{vmatrix} \begin{vmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{vmatrix} = \begin{vmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n X_{i1}y_i \\ \dots \\ \sum_{i=1}^n X_{ik}y_i \end{vmatrix}$$

El modelo de regresión ajustado es:

$$\hat{y} = x\hat{\beta} \quad 5.34$$

En notación escolar, el modelo ajustado es:

$$\hat{y}_i = \beta_0 + \sum_{j=1}^k \beta_j X_{ij} \quad i = 1, 2, \dots, n \quad 5.35$$

La diferencia entre la observación y_i y el valor ajustado \hat{y}_i es un residuo, digamos $e_i = y_i - \hat{y}_i$. El vector $(n \times 1)$, de los residuos se denota mediante:

$$e = y - \hat{y} \quad 5.36$$

Suele ser necesario estimar σ^2 . Para desarrollar este estimador, considérese la suma de cuadrados de los residuos es:

$$SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e^2 = y'y - \hat{\beta}'x'y \quad 5.37$$

La media cuadrática media es :

$$MS_E = \frac{SS_E}{n - p} \quad 5.38$$

Puede mostrarse que el valor esperado de MS_E es σ^2 , por lo que el estimador neutral de σ^2 está dado por:

$$\sigma^2 = MS_E \quad 5.39$$

5.6.3 SIGNIFICACIÓN GLOBAL DEL AJUSTE

Es posible utilizar diversas técnicas para medir la adecuación de un modelo de regresión múltiple. La validación del modelo es una parte importante del proceso de construcción del modelo de regresión múltiple. Para ello se realiza la siguiente prueba de validación:

- a. Coeficiente de determinación múltiple (R^2)
- b. Pruebas de hipótesis.

A continuación se detallan cada una de las pruebas:

- a. Coeficiente de determinación múltiple (R^2).

El coeficiente de determinación múltiple R^2 se define como:

$$R^2 = \frac{SS_R}{S_{yy}} = 1 - \frac{SS_E}{S_{yy}} \quad \text{para } N - 1 \text{ grados de libertad}$$

donde:

SS_R es la suma de cuadrados de la regresión

S_{yy} es la suma total de los cuadrados

SS_E es la suma de los cuadrados del error.

En forma matricial se tiene:

$$SS_R = \hat{\beta}'x'y - \frac{(\sum_{i=1}^n y_i)^2}{n}$$

$$SS_E = S_{yy} - SS_R \quad 5.40$$

$$S_{yy} = y'y - \frac{(\sum_{i=1}^n y_i)^2}{n}$$

Lo que implica que R^2 es una medida del grado de reducción en la variabilidad de y obtenida mediante el empleo de las variables regresivas x_1, x_2, \dots, x_n . Como en el caso de la regresión lineal simple, se debe tener que $0 \leq R^2 \leq 1$. Sin embargo, un valor grande de R^2 no necesariamente implica que el modelo de regresión sea bueno. Como por ejemplo, añadir una variable al modelo siempre aumentará R^2 , independientemente de si la variable adicional es o no estadísticamente significativa. De tal modo, es posible en modelos que tienen grandes valores de R^2 producir predicciones pobres de nuevas observaciones o estimadores de la respuesta media.

La raíz cuadrado positiva de R^2 es el coeficiente de correlación múltiple entre y y el conjunto de variables regresoras x_1, x_2, \dots, x_k . Esto es, R , es una medida de la

asociación lineal entre "y" y X_1, X_2, \dots, X_k . Cuando $k = 1$, esto se vuelve la correlación simple entre "y" y "x".

b. Pruebas de hipótesis

En problemas de regresión lineal múltiple, si ciertos tipos de hipótesis respecto a los parámetros del modelo son útiles al medir la suficiencia del modelo.

b.1. Prueba de significación de regresión

La prueba de significación de regresión es para determinar si hay una relación lineal entre la variable respuesta (dependiente) "y" y un subconjunto de las variables explicativas (regresoras) x_1, x_2, \dots, x_k . Las hipótesis apropiadas son:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \beta_j \neq 0 \text{ por lo menos una } j \text{ será diferente}$$

El rechazo de $H_0: \beta_j = 0$ implica que al menos una de las variables independientes x_1, x_2, \dots, x_k contribuye significativamente al modelo.

El procedimiento de prueba para $H_0: \beta_j = 0$, es calcular:

$$F_0 = \frac{SS_R / k}{SS_E / (n - k - 1)} \quad 5.41$$

Criterio de rechazo, es cuando:

$$F_0 > F_{\alpha, k, n-k-1}$$

El procedimiento suele resumirse en una tabla de análisis de varianza; como el que muestra en la tabla 5.4.

Tabla 5.4 . Análisis de varianza para la significación de la regresión múltiple.

Fuente de variación	Suma de cuadrados	Grados de libertad	Media cuadrática	F ₀
Regresión	SS _R	k	MS _R	MS _R /MS _E
Error o residuo	SS _E	n-k-1	MS _E	
Total	S _{yy}	n-1		

b.2. Pruebas de coeficientes individuales de regresión

Esto es para la prueba de significación del efecto de una variable regresiva o explicativa. Con frecuencia se está interesado en probar hipótesis respecto a los coeficientes individuales de regresión. Tales pruebas son útiles en la determinación del valor de cada una de las variables independientes en el modelo de regresión. Por ejemplo, el modelo podría ser más eficaz con la inclusión de variables adicionales, o quizá con la omisión de una o más variables, ya en el modelo .

1. Prueba de hipótesis parcial o marginal

La hipótesis para probar la significación de cualquier coeficiente de regresión individual es:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

Si $H_0: \beta_j = 0$, no se rechaza, entonces esto indica que X_j puede ser eliminada del modelo, en caso contrario no puede omitirse del modelo, porque x_j influye sobre "y". La estadística de prueba para hipótesis es:

$$t_0 = \frac{\beta_j}{\sqrt{MS_E C_{ij}}} \quad 5.42$$

Donde : C_{ij} es el elemento de la diagonal de $(x'x)^{-1}$ correspondiente a β_j . La hipótesis nula se rechaza cuando:

$$|t_0| > t_{\alpha/2, n-k-1}$$

Esta se le denomina prueba de hipótesis parcial ó marginal. Debido que β_j depende de todas las variables regresivas x_i ($i \neq j$) que están en el modelo.

2. Prueba de hipótesis para un subconjunto de las variables regresivas.

Se realiza examinando la contribución a la suma de cuadrados de la regresión de una variable, por ejemplo x_j , dado que las otras variables x_i ($i \neq j$), están incluidas en el modelo. El procedimiento empleado para hacer esto se denomina la prueba general de significación de la regresión, o el método de la "suma de cuadrados extra". Considérese el modelo de regresión con k variables regresivas.

$$y = x\beta + \varepsilon \quad 5.43$$

Donde "y" es $(n \times 1)$, x es $(n \times p)$, β $(p \times 1)$, ε es $(n \times 1)$, y $p = k+1$. Se interesa determinar si el subconjunto de variables regresivas x_1, x_2, \dots, x_r ($r < k$) contribuye de manera significativa al modelo de regresión. Se desea probar las hipótesis.

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

La estadística de prueba es:

$$F_0 = \frac{SS_R(\beta_1 | \beta_2) / r}{MS_E} \quad 5.44$$

donde:

$$SS_R\left(\frac{\beta_1}{\beta_2}\right) = SS_R(\beta) - SS(\beta_2). \text{ Suma de cuadrados con } r \text{ grados de}$$

libertad "Suma de cuadrados extra" debido a β_1 .

$$SS_R = \hat{\beta}'X'y. \text{ (} p \text{ grados de libertad). Suma de cuadrados de la}$$

regresión debida a β

$$SS_R(\beta_2) = \hat{\beta}_2'X_2'y. \text{ (} p-r \text{ grados de libertad). Suma de cuadrados de la}$$

regresión debida β_1 dado que β_2 ya esta en el modelo.

El criterio de rechazo utilizado es:

$$F_0 > F_{\alpha, r, n-p}$$

La hipótesis nula se rechaza. Llamada prueba F parcial, la cual mide el efecto conjunto de variables en el modelo de regresión.

6.5.2.1 Intervalos de Confianza en Regresión Lineal Múltiple.

Con frecuencia es necesario construir estimadores del intervalo de confianza para los coeficientes de regresión (β_j). Suponiendo que los errores (ε_i) se distribuyen normal e independientemente con media $\beta_0 + \sum_{i=1}^p \beta_j x_{ij}$ y varianza σ^2 . Puesto que el estimador de mínimos cuadrados β es una combinación lineal de las observaciones resulta que b se distribuye normalmente con la media β y matriz de covarianza $\sigma^2 (X'X)^{-1}$. Entonces una de las estadística es:

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 C_{jj}}} \quad j=0, 1, \dots, k \quad 5.45$$

Se distribuye como t con $n-p$ grados de libertad, donde C_{ij} es el elemento ij ésimo de la matriz $(X'X)^{-1}$, y σ^2 es la estimación de la varianza del error, obtenida de por MS_E . En consecuencia, un intervalo de confianza de $100(1-\alpha)$ por ciento para el coeficiente de regresión β_j , $j=1, 2, \dots, k$, es:

$$\hat{\beta}_j - t_{\alpha/2, n-p} \sqrt{\sigma^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2, n-p} \sqrt{\sigma^2 C_{jj}} \quad 5.46$$

5.6.5 PREDICCIÓN DE NUEVAS OBSERVACIONES.

El modelo de regresión puede utilizarse para predecir observaciones futuras respecto a "y"; entonces una estimación puntual de las observaciones futuras, y_0 en el punto $x_{01}, x_{02}, \dots, x_{0k}$, es:

$$\hat{y}_0 = x'_0 \beta \quad 5.47$$

Un incremento de predicción del 100 (1-), por ciento para esta observación futura es:

$$\hat{y}_0 - t_{\alpha/2, n-p} \sqrt{(1 + x'_0 (x'x)^{-1} x_0)} \leq y_0 \leq \hat{y}_0 + t_{\alpha/2, n-p} \sqrt{(1 + x'_0 (x'x)^{-1} x_0)} \quad 5.48$$

5.6.6 VALIDACIÓN POR ANÁLISIS DE RESIDUOS.

Los residuos del modelo de regresión múltiple estimado, definido por $y_i - \hat{y}_i$, desempeñan un importante papel al juzgar la suficiencia del modelo del mismo modo que lo hacen en regresión lineal.

Cuestionamiento:

- ¿Es admisible que $\varepsilon_t \sim \text{NID}(0, \sigma^2)$?
- ¿Hay algún dato claramente anómalo?
- ¿Es admisible que la varianza de ε_t no depende de las variables explicativas (regresivas)?
- ¿Es realmente lineal la relación entre \hat{y} y x_i ?

Para la comprobación se hace necesario revisar las gráficas de residuos (análisis de residuos).

5.7 REGRESIÓN POLINOMIAL

El modelo $y = X\beta + \varepsilon$, es un modelo general que puede emplearse para ajustar cualquier relación que sea lineal en los parámetros desconocidas B . Esto incluye la importante clase de los modelos de regresión polinomial, por ejemplo:

$$y_0 = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k \quad 5.49$$

Se puede resolver por métodos algebraicos, a partir de las ecuaciones normales o métodos matriciales descritos en el análisis de regresión lineal múltiple.

Las ecuaciones normales para el análisis de regresión polinómicas son:

$$\begin{aligned} \sum y &= n\beta_0 + \beta_1 \sum x + \dots + \beta_k \sum x^k \\ \sum xy &= \beta_0 \sum x + \beta_1 \sum x^2 + \dots + \beta_k \sum x^{k+1} \\ \sum x^k y &= \beta_0 \sum x^k + \beta_1 \sum x^{k+1} + \dots + \beta_k \sum x^{2k} \end{aligned}$$

5.8 EJEMPLOS DEL CAPÍTULO

Ejemplo 5.1

Una empresa de productos químicos desea estudiar la influencia del tiempo de extracción en la eficiencia de una operación de extracción, obteniendo los datos mostrados en la tabla 5.4.

Tabla 5.4. Resultados obtenidos de los tiempos de extracción (X) (min) y eficiencia (y) (%).

Min.	15	19	19	27	31	35	39	41	45	49
%	57	46	52	57	68	62	72	80	64	77

- Realice en diagrama de dispersión para verificar que una línea recta proporcionará un buen ajuste de los datos.
- Ajuste un modelo de regresión lineal simple a los datos.
- Pruebe la significancia de la regresión.
- Construya un intervalo de confianza del 95 por ciento para β_0 y β_1 .
- Construya un intervalo de confianza sobre la línea de regresión real en $x=41$.

Solución:

- En la figura 5.5. se muestra los datos graficados de la eficiencia de extracción (Y) (%) vrs. Tiempo de extracción (min) (X).

Según se visualiza el diagrama de dispersión de los datos (Figura 6.) los datos son ajustable por un modelo de regresión simple lineal sin la necesidad de transformar las variables de origen.

- El modelo para ajustar los datos es el siguiente:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Para determinar los coeficiente de la ordenada y lineal, se harán los siguientes cálculos previos:

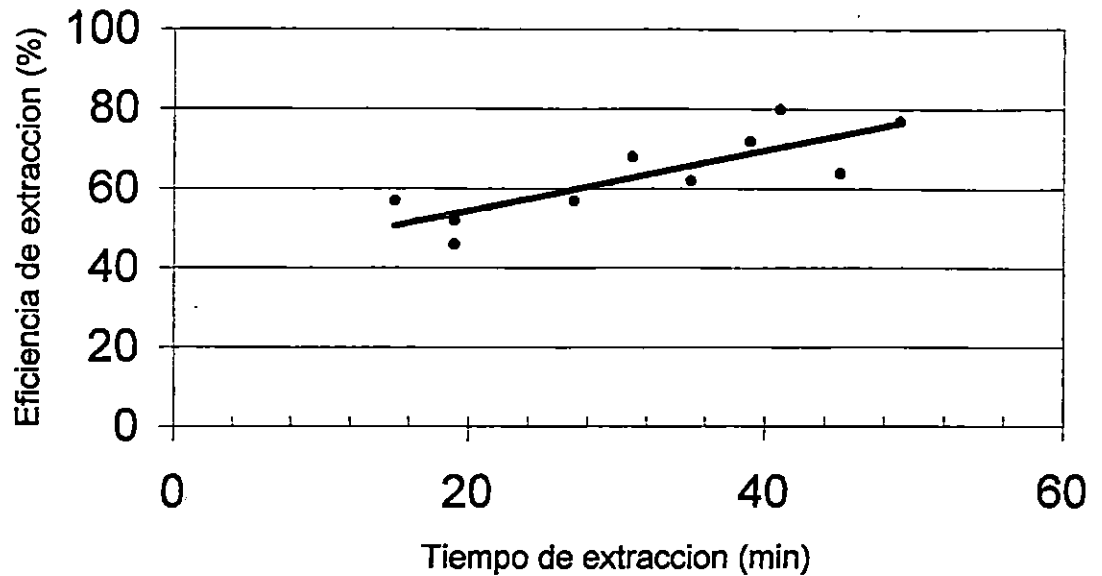


Figura 5.4. Diagrama de dispersión de los datos

$$\begin{array}{ll}
 n=10 \text{ observaciones} & \sum_{i=1}^{10} x_i = 320 \\
 \sum_{i=1}^{10} y_i = 635 & \sum_{i=1}^{10} x_i^2 = 11490 \\
 \sum_{i=1}^{10} y_i^2 = 41395 & \sum_{i=1}^{10} x_i y_i = 21275 \\
 \bar{x} = 32 & \bar{y} = 63.5 \\
 S_{xx} = \sum_{i=1}^{10} x_i^2 - \frac{\left(\sum_{i=1}^{10} x_i\right)^2}{10} = 1250 & S_{xy} = 955
 \end{array}$$

Entonces:

β_0 es:

$$B_1=0.764$$

$$B_0=39.052$$

Y R^2 es:

$$R^2=0.6803$$

Lo que significa es que el modelo representa la variabilidad de los datos en un 68.03 %. Y el modelo ajustado es:

$$\hat{y} = 39.052 + 0.764x$$

- c. Para la significancia del modelo se seguirá el siguiente procedimiento:
1. Hipótesis nula: $\beta_1=0$
Hipótesis alternativa: $\beta_1 \neq 0$
 2. Nivel de significancia: $\alpha=0.05$
 3. Criterio de rechazo: $F_0 > 5.32$, con 1 y 8 grados de libertad.
 4. Análisis de varianza, como se resume en la tabla 5.5.

Tabla 5.5. Resumen del análisis de varianza para el modelo ajustado

Fuente de variación	Suma de cuadrados	Grados de libertad	Media Cuadrática	F_0	$F_{\alpha/2,1,8}$
Regresión	729.62	1	729.62	17.023	5.32
E. residual	342.88	8	42.86		
Total	1072.5	9			

En conclusión la hipótesis nula se rechaza y con ello se dice que β_1 es diferente de cero de manera significativa.

- d. El intervalo de confianza para β_1 con $t_{0.025,8}=2.306$, es:

$$0.764 - 2.306 \sqrt{\frac{42.86}{1250}} \leq \beta_1 \leq 0.764 + 2.306 \sqrt{\frac{42.86}{1250}}$$

$$0.337 \leq \beta_1 \leq 1.191$$

el valor de β_1 fluctuará entre 0.337 como mínimo y como máximo de 1.191.

y para β_0 es:

$$39.052 - 2.306 \sqrt{42.86 \left[\frac{1}{10} + \frac{32^2}{1250} \right]} \leq \beta_0 \leq 39.052 + 2.306 \sqrt{42.86 \left[\frac{1}{10} + \frac{32^2}{1250} \right]}$$

$$36.841 \leq \beta_0 \leq 41.263$$

Lo que significa que β_0 fluctuará entre 36.841 como mínimo y como en máximo de 41.263.

- e. Para el intervalo de confianza sobre la línea de regresión con $x=41$, con y ajustado 70.376 (de Tabla 6...) es:

$$70.376 \pm 2.306 \sqrt{42.86 \left[\frac{1}{10} + \frac{(80 - 32)^2}{1250} \right]}$$

$$49.331 \leq E(y / x_0 = 41) \leq 91.421$$

En conclusión los intervalos de fluctuación tanto para los coeficiente y sobre la línea de regresión es bastante significativo. Esto se debe por que el coeficiente de determinación es bien bajo y el grado de dispersión de los datos son grandes.

- f. Para hacer un examen de idealidad del modelo se determinará lo siguiente:
- Los valores predichos para el modelo ajustados
 - Los residuos de los valores observados reales y predichos por el modelo ajustado.
 - El calculo de las frecuencia acumulada.

Tabla 5.6. Tabla resumen de los valores predichos, residuos y frecuencia acumulada

X	Y	Ajustado	Residuos	Frecuencia acumulada (%)	Residuos ascendente
15	57	50.512	6.488	5	-9.435
19	46	53.568	-7.568	15	-7.568
19	52	53.568	-1.568	25	-3.792
27	57	59.68	-2.68	35	-2.68
31	68	62.736	5.264	45	-1.568
35	62	65.792	-3.792	55	0.512
39	72	68.848	3.152	65	3.152
41	80	70.376	9.624	75	5.264
45	64	73.435	-9.435	85	6.488
49	77	76.488	0.512	95	9.624

Todos estos se presenta en forma resumida en la tabla 5.5.

Con lo antes determinado se realiza el análisis de residuos, graficando:

- Frecuencia acumulada vrs. Residuos en forma ascendente (Figura 5.6)
- Residuos correspondiente vrs. Valores ajustados (Figura 5.7)
- Residuos correspondiente vrs. X (Figura 5.8)

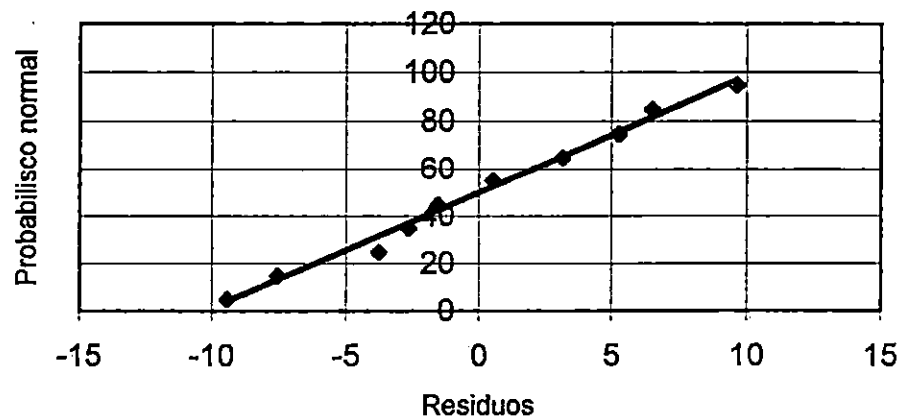


Figura 5.6. Grafica en papel probabilístico contra residuos

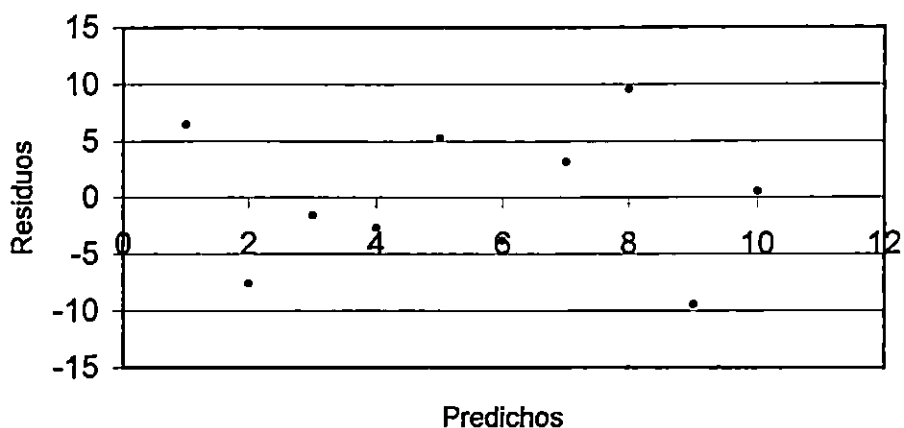


Figura 5.7. Grafica de residuos contra los valores predichos

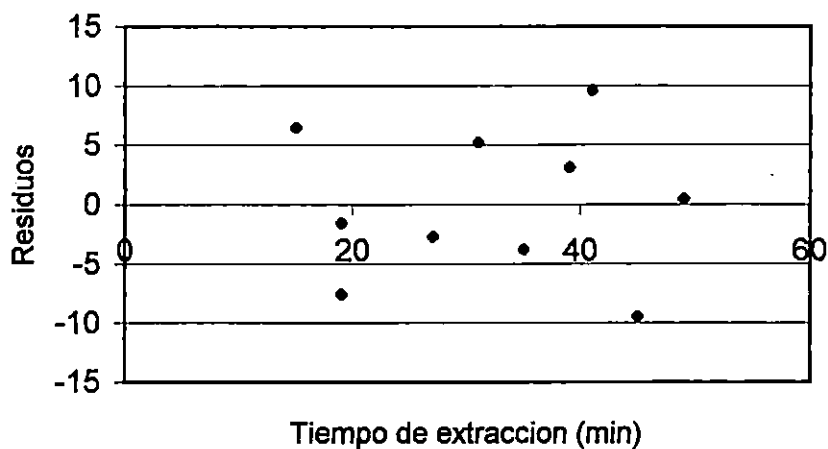


Figura 5.8. Grafica de residuos contra el tiempo de extracción (min)

En conclusión los datos presentan normalidad y el análisis de residuos tiene un buen comportamiento.

Ejemplo 5.2

A partir de los resultados retomados del trabajo de graduación para optar al título de Ingeniero Químico "*Recuperación de Grasas Residuales de Tenerías y Uso Industrial*"; presentado por: García. Et al, 1994. En el que se extrajo grasa residual de la camaza de tenerías. Se obtuvieron los siguientes datos de rendimiento (y)

Como función de las variables: Tiempo (horas) (x_1) y Temperatura ($^{\circ}\text{C}$) (x_2). Los resultados se presentan en la Tabla 5.7

Tabla 5.7. Resultados de rendimiento obtenido en la recuperación de grasa de la carnaza por tratamiento térmico en función de la temperatura y tiempo de procesamiento (para el 24% de grasa en la carnaza)

Tiempo (X_1) (horas)	Temperatura (X_2) ($^{\circ}\text{C}$)	Rendimiento (Y) (%P)
1	54	9.12
1	74	13.16
1	94	29.06
1.5	54	40.53
1.5	74	46.09
1.5	94	46.5
2.0	54	53.06
2.0	74	65.06
2.0	94	73.87

Determinar:

- Un modelo de regresión doble para ajustar los datos
- La prueba de significancia de la regresión
- Los residuos de este modelo y analizados en forma grafica

Solución

- El modelo de regresión doble propuesto sería:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Se comprobó que los datos no resultaron lineales en sus variables naturales, por ello se tuvo que transformar, tanto las variables independientes como dependiente (aplicando logaritmos). Entonces se procede a lo siguiente:

Cálculos previos:



$$\sum_{i=1}^9 \ln x_{i1} = 3.296$$

$$\sum_{i=1}^9 \ln x_{i2} = 38.509$$

$$\sum_{i=1}^9 \ln x_{i1}^2 = 1.934$$

$$\sum_{i=1}^9 \ln x_{i2}^2 = 165.235$$

$$\sum_{i=1}^9 y_i = 376.45$$

$$\sum_{i=1}^9 y_i^2 = 19535.0067$$

$$\sum_{i=1}^9 x_{i1}x_{i2} = 14.102$$

$$\sum_{i=1}^9 x_{i1}y_i = 187.379$$

$$\sum_{i=1}^9 x_{i2}y_i = 1625.674$$

El sistema de ecuaciones en forma matricial queda de la siguiente manera:

$$X'y = \begin{vmatrix} 9 & 3.296 & 38.509 \\ 3.296 & 1.934 & 14.102 \\ 38.509 & 14.102 & 165.235 \end{vmatrix} \begin{vmatrix} 376.92 \\ 187.379 \\ 1625.674 \end{vmatrix}$$

Resolviéndose se tiene:

$$\hat{\beta}_0 = -102.756$$

$$\hat{\beta}_1 = 67.906$$

$$\hat{\beta}_2 = 27.9905$$

En consecuencia el modelo de regresión múltiple es:

$$\hat{y} = -102.756 + 67.906x_1 + 27.995x_2$$

b. Para la significación de lo modelo se seguirá de la siguiente manera:

1. planteamiento de hipótesis:

para la hipótesis nula se tiene:

$$H_0 : \beta_0 = \beta_1 = \beta_2 = 0$$

Para la hipótesis alternativa se tiene:

$$\beta_i \neq 0 \text{ por lo menos para un } i$$

2. Cálculos previos (Suma de cuadrados):

$$SS_{1y} = \sum_{i=1}^9 y_i X_{i1} - \frac{\left(\sum_{i=1}^9 y_i\right)\left(\sum_{i=1}^9 X_{i1}\right)}{9} = 187.379 - \frac{376.45 * 3.296}{9} = 49.515$$

$$SS_{2y} = \sum_{i=1}^9 y_i X_{i2} - \frac{\left(\sum_{i=1}^9 y_i\right)\left(\sum_{i=1}^9 X_{i2}\right)}{9} = 1625.674 - \frac{376.45 * 38.509}{9} = 14.928$$

$$SS_R = \hat{\beta}_1 * SS_{1y} + \hat{\beta}_2 * SS_{12} = 3780.21$$

$$S_{yy} = \sum_{i=1}^9 y_i^2 - \frac{\left(\sum_{i=1}^9 y_i\right)^2}{9} = 19535.0067 - \frac{376.45^2}{9} = 3788.939756$$

$$SS_E = S_{yy} - SS_R = 3788.939756 - 3780.21 = 8.73$$

3. Resumen de análisis de varianza, en la Tabla 5.8 se muestra el análisis de varianza.

Tabla 5.8. Análisis de varianza de los datos

Fuente de variación	Suma de cuadrados	Grados de libertad	Media cuadrática	F ₀
Regresión	3780.21	2	1890.105	1298.41659
Error	8.73	6	1.4557	
total	3788.939756	8		

Y coeficiente de determinación múltiple de:

$$R^2 = \frac{SS_R}{S_{yy}} = \frac{3780.21}{3788.94} = 0.9976$$

Lo que significa que el modelo propuesto describe la variabilidad de los datos casi en un 100%.

- c Análisis de residuos. Es semejante al análisis realizado en el ejemplo 5.1, y consiste en graficar los residuos en probabilístico normal (Figura 5.8), con los valores predichos (Figura 5.9), con los valores de los tiempo de extracción transformada (Ln X1) (Figura 5.10) y con la temperatura de extracción transformada (Ln X2) (Figura 5.11).

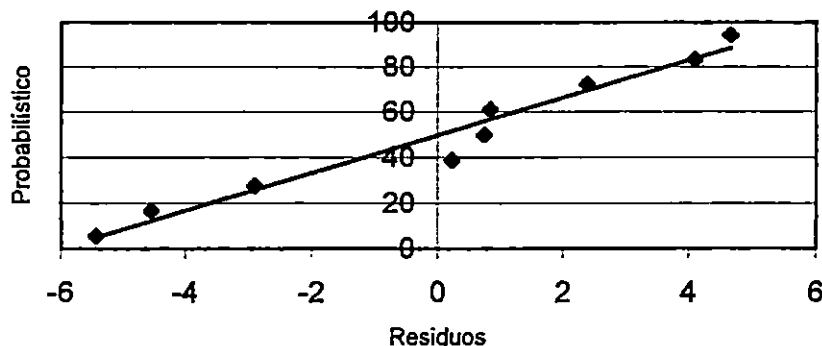


Figura 5.8. Gráfica de los residuos en probabilístico normal

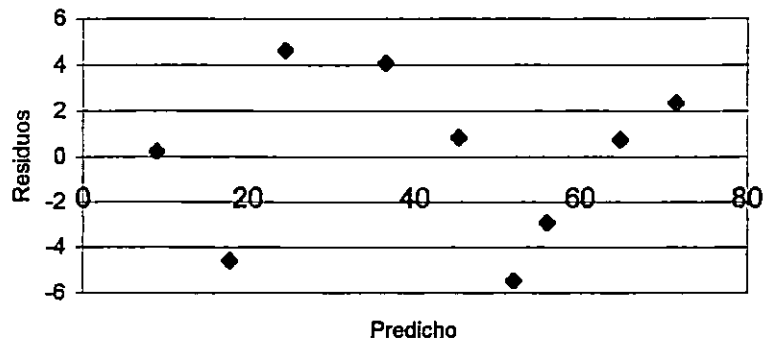


Figura 5.9. Grafica de residuos vs. los valores predichos

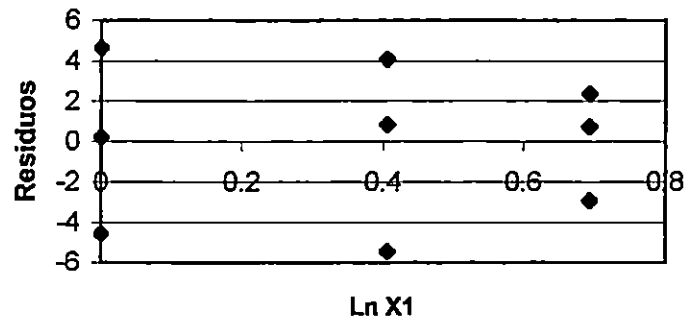


Figura 5.10. Grafica de residuos vs. Ln de los tiempos de extracción

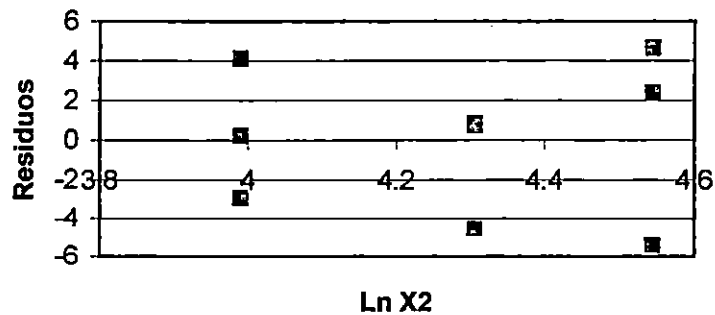


Figura 5.11. Grafica de residuos vs. Ln de la temperatura

En base de las figura 5.9, 5.10 y 5.11 y a los resultados de la Tabla 5.8 y el coeficiente de determinación múltiple puede concluirse que el modelo sobre las variables transformadas es significativo indicando una conveniente solución lineal entre el neperiano de rendimiento con el neperiano del tiempo y temperatura.

5.9 EJERCICIOS PROPUESTOS

- 5.1. Un ingeniero químico está investigando el efecto de la temperatura de operación de proceso en el rendimiento del producto. El estudio da como resultado los siguientes resultados (Montgomery, 1991):

Temperatura °C (x)	100	110	120	130	140	150	160	170	180	190
Rendimiento % (y)	45	51	54	61	66	70	74	78	85	89

Determinar:

- Un modelo de regresión lineal simple a los resultados
 - La prueba de hipótesis de los estimadores y del modelo.
 - La idoneidad del modelo.
- 5.2. La resistencia del papel utilizado en la manufactura de caja de papel (y) se relaciona con el porcentaje de la concentración de madera dura en la pulpa original (x). En condiciones controladas, una planta piloto manufactura 16 muestras, cada una de diferentes lotes de pulpa, y se mide la resistencia a la tensión. Los datos son los siguientes (Hines. Et al, 1997):

X	Y	X	Y	X	Y
1.0	101.4	2.2	146.8	2.8	145.2
1.5	117.4	2.4	133.9	3.0	134.3
1.5	117.1	2.5	113.3	3.0	144.5
2.0	131.9	2.5	123.0	3.2	143.7
2.0	146.9	2.8	125.1	3.3	146.9

Determinar:

- Un modelo de regresión lineal simple para los datos.
- La significación del modelo de la regresión.

- c. Construya intervalos de confianza del 90 por ciento de los estimadores y del modelo.

5.3. Se considera que el número de litros de vapor utilizados al mes por una planta química se relaciona con la temperatura ambiente para ese mes. El consumo y la temperatura del año pasado se muestran a continuación (Johnson, 1997):

Mes	Temperatura	Consumo/1000	Mes	Temperatura	Consumo/1000
Enero	21	185.79	Julio	68	621.55
Febrero	24	214.47	Agosto	74	675.06
Marzo	32	288.03	Septiembre	62	562.03
Abril	47	424.84	Octubre	50	452.93
Mayo	50	454.58	Noviembre	41	369.95
Junio	59	539.03	Diciembre	30	273.98

Determinar:

- Un modelo de regresión simple lineal que ajuste los datos.
 - La significación del modelo del modelo de regresión.
 - Las hipótesis de: β_0 , β_1 y del modelo de regresión.
 - Los intervalos de confianza del 95 por ciento del modelo de regresión.
- 5.4. A doce especímenes de láminas de acero reducido en frío con contenido diferentes de cobre y diferentes temperaturas de recocido se les mide su dureza con los siguientes resultados (Montgomery, 1991):

Dureza (Rockwell 30-T)	Contenido de cobre (%)	Temperatura de cocido (°F)
78.9	0.02	1000
65.1	0.02	1100
55.2	0.02	1200

56.4	0.02	1300
80.9	0.10	1000
69.7	0.10	1100
57.4	0.10	1200
55.4	0.10	1300
85.3	0.18	1000
71.8	0.18	1100
60.7	0.18	1200
58.9	0.18	1300

Para los resultados presentados en la tabla anterior, ajuste una ecuación de la forma $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, donde x_1 representa el contenido de cobre, x_2 la temperatura de recocido y " y " es la dureza. Además, realizar las pruebas de significación del modelo y comprobación de la idoneidad del modelo.

CAPÍTULO 6

DISEÑOS DE EXPERIMENTOS

Los investigadores en el campo de Ingeniería Química e Ingeniería de alimentos, realizan experimentos virtualmente en cada etapa de su desempeño, por lo general, para descubrir algo sobre un proceso particular. Un diseño de experimentos, es literalmente el diseño de una prueba o ensayo. Un experimento diseñado es la prueba o serie de pruebas, en las cuales se inducen cambios deliberados en las variables de entrada (explicativas o regresivas) de un proceso o sistema, de manera que sea posible observar e identificar las causas de los cambios en las respuestas de salida.

Entre los objetivos del experimento pueden incluirse:

1. Determinar cuáles variables tienen mayor influencia en la respuesta, "y".
2. Determinar el mejor valor de las "x" que influyen en "y" de modo que "y" tenga casi siempre un valor cercano al valor nominal deseado.
3. Determinar el mejor valor de las "x" que influyen en "y" de modo que la variabilidad de "y" sea pequeña.
4. Determinar el mejor valor de las "x" que influyen en "y", de modo que se minimicen los efectos de las variables incontables.

Los métodos de diseño experimental tienen un cometido importante en el desarrollo de procesos y en la depuración de procesos para mejorar el rendimiento.

En muchos casos, el objetivo puede ser desarrollar un proceso consistente o robusto; esto es, un proceso afectado mínimamente, por las fuentes de variación externas.

6.1 APLICACIONES DE LOS DISEÑOS EXPERIMENTALES

Es posible considerar a la experimentación parte del proceso científico y una de las formas en que se aprende acerca de la forma en que funcionan los sistemas o procesos. Con el fin de generar datos a partir del proceso, y luego se puede usar la información para establecer nuevas conjeturas, que llevan a realizar nuevos experimentos. La aplicación de técnicas de diseño experimental en una fase temprana del desarrollo de un proceso puede dar por resultado:

1. Mejora en el ordenamiento y planificación del proceso investigativo
2. Menor variabilidad y mayor apego a los factores que afectan la variable respuesta, haciéndolas más objetivas.
3. Menor tiempo de desarrollo y ejecución
4. Menores costos globales.

Los métodos de diseño experimental también tienen un cometido importante en las actividades de diseño técnico (o diseño de ingeniería), en las cuales se desarrollan nuevos productos y mejoras de otros ya existentes. Algunas aplicaciones del diseño experimental en el diseño técnico son:

1. Evaluación y comparación de configuración de diseño básicos
2. Evaluación de materiales alternativos
3. Selección de parámetros de diseño de modo que el producto funciones bien en una amplia variedad de condiciones de campo (de uso real); esto es, de modo que el producto sea consistente (robusto).

6.2 PRINCIPIOS BÁSICOS

Para que un experimento se realice en la forma más eficiente, es necesario emplear métodos científicos en su planeación. El diseño estadístico de experimentos es el proceso de planear un experimento para lograr obtener datos apropiados, que pueden ser analizados mediante métodos estadísticos, con objetos de producir conclusiones válidas y objetivas. Así que hay dos aspectos en cualquier problema experimental; estos son:

1. El diseño del experimento
2. El análisis estadístico.

Los tres principios básicos en el diseño de experimentos son:

- a. La obtención de réplicas
- b. Aleatorización
- c. Análisis por bloques.

a. La obtención de réplicas, permite:

1. Obtener una estimación del error experimental. Usualmente la media de la muestra.
2. Obtener un alto grado de confianza, cuando n es suficientemente grande, obteniendo un error experimental suficientemente pequeño.

b. La aleatorización. Es el hecho de que tanto la asignación del material experimental como el orden en que se realizan las pruebas individuales o ensayos se determinan aleatoriamente, permitiendo que los diseños de experimentos posea:

1. Fundamento del uso de los métodos estadísticos
 2. Observaciones (o los errores) que sean variables aleatorias independientes.
 3. Disminución de los factores extraños que pudieran estar presentes.
- c. El análisis por bloques. Es una técnica que se usa para incrementar la precisión del experimento. Siendo una porción del material experimental que sea más homogénea que el total del material, haciendo comparaciones entre las condiciones de interés del experimento dentro de cada bloque.

Para usar un enfoque estadístico al diseñar y analizar un experimento se requiere tener una idea clara de qué es exactamente lo que se va a estudiar, como se van a recopilar los datos y, al menos, una idea cualitativa de cómo se van a analizar. A continuación, una guía del procedimiento recomendado:

1. Comprensión y planteamiento del problema. Desarrollar todas las ideas sobre los objetivos del experimento, solicitando la opinión de todas las partes implicadas. Ayudando a un mejor conocimiento del fenómeno y de la solución final del problema.
2. Elección de factores y niveles. Elegir los factores que variarán en el experimento, los intervalos de dicha variación y los niveles específicos y la forma de cómo se controlarán en el experimento.
3. Selección de la variable de respuesta. Seleccionar la variable dependiente, se debe estar seguro de que la respuesta que se va a medir realmente provea información útil
4. Elección del diseño experimental. Para elegir el diseño es necesario considerar el tamaño muestral (número de repeticiones), seleccionar un orden adecuado para los ensayos y determinar si hay implicado bloqueo u otras restricciones de aleatorización.

5. **Realización del experimento.** Es poner en marcha el experimento en forma física y es necesario vigilar el proceso cuidadosamente para asegurar que todo se haga conforme a lo planeado. Es la fase donde se obtiene las observaciones, la cual, los errores en el procedimiento suelen anular la validez experimental.
6. **Análisis de datos.** Es la fase donde se emplea métodos estadísticos para analizar los datos, de modo que los resultados y conclusiones sean objetivas más que apreciativas. Si el experimento se diseñó correctamente y si se ha realizado conforme al diseño, los métodos estadísticos que se requieren no son complicados.
7. **Conclusiones y recomendaciones.** Es la fase en donde el experimentador deba extraer conclusiones prácticas de los resultados y recomendar un curso de acción. En esta fase es necesario a menudo utilizar métodos gráficos. También deben realizarse corridas de seguimiento y pruebas de confirmación para validar las conclusiones del experimento.

Los métodos estadísticos pueden incrementar grandemente la eficiencia de los experimentos, y, a menudo, reforzar las conclusiones obtenidas, para ello el investigador requiere que tenga en mente los siguientes puntos.

1. **Uso del conocimiento no estadístico del problema.** El investigador debe conocer a fondo su campo de especialidad. En algunos campos pueden utilizarse una gran cantidad de teoría para explicar las relaciones que hay entre los factores y respuestas. El tipo de conocimiento no estadístico es invaluable al elegir los factores y sus niveles.
2. **Mantener el diseño y el análisis tan simple como sea posible.** No se debe exagerar el uso de técnicas estadísticas complejas y muy

refinadas. Por lo general, lo más adecuado en los métodos de diseño y análisis relativamente simples.

3. Reconocer la diferencia entre la significación práctica y la estadística. Se debe tener en cuenta que no existe seguridad de que una diferencia sea suficientemente grande, desde el punto de vista práctico, por solo el hecho de que dos condiciones experimentales producen respuestas medias, estadísticamente diferentes.
4. Usualmente los experimentos son iterativos. Un diseño exitoso requiere que se conozca los factores importantes, los intervalos en los que estos factores van a ser investigados, el número apropiado de niveles para cada factor y las unidades de medición adecuadas a cada factor y a la respuesta. Generalmente, al inicio de un experimento no se está en condiciones de responder adecuadamente a estas preguntas, pero es posible conocer las respuestas a medida que avanza la experimentación. Esto favorece el proceso iterativo o secuencial, ya tratado anteriormente.

6.3 DISEÑO Y ANALISIS DE EXPERIMETNO DE UN FACTOR.

Un experimento no es más que una prueba o una serie de pruebas y son una parte fundamental del proceso de descubrimiento y aprendizaje. Los diseños de experimentos de un factor, es cuando se diseña y analiza el experimento por el efecto de un solo factor que es la variable independiente o causa que tendrá niveles o tratamientos.

Se llaman factores controlados a aquellos parámetros o características del producto o proceso para los que se prueban distintas variables o valores, con el fin de estudiar como influyen sobre los resultados. Pueden ser: cuantitativas, por ejemplo: Temperatura, presión, etc; que se prueban a diferentes niveles (ejemplo: temperatura a 115 °C, 125 °C, y 135 °C) y cualitativos, por ejemplo proveedor,

tipo de lubricante, técnica, método, etc; se prueban a diferentes variantes cualitativamente diferentes (ejemplo: técnico A, B, Y C).

Tratamiento: Combinación de variantes y/o niveles de cada factor que se utiliza en una determinada prueba.

Los diseños experimentales de un solo factor se clasifican en:

1. Experimento de un factor, completamente aleatorio
2. Modelo de efectos aleatorios
3. Diseño de bloques aleatorio
4. Análisis de covarianza

En las secciones 6.3.1 a 6.3.3., se analizará los diseños usados con frecuencia de clasificación sencilla y doble; en la sección se presenta pruebas para comparar varias medias, así como el análisis de un experimento balanceado en presencia de una variable concomitante o covariada.

6.3.1 EXPERIMENTOS DE UN FACTOR COMPLETAMENTE ALEATORIO (CASO DE EFECTOS FIJOS)

El diseño completamente aleatorio, es muy utilizado cuando el material experimental es homogéneo, aun cuando tiene varios tratamientos. Consistiendo de que el experimento se diseña con resultados de "n" muestras aleatorias, independientes, cada una de tamaño n, de a poblaciones diferentes (esto es, dato que concierne a "a" métodos de producciones diferentes, etc). Tratando de probar la hipótesis de que las medias de esas "a" poblaciones son todas iguales. Los datos aparecerán como en la tabla 6.1.

Tabla 6.1. Datos típicos para el análisis de varianza de clasificación en un sentido.

Tratamiento	Observación	Totales	Promedio
1	$y_{11} \ y_{12} \ \dots \ y_{1n}$	$y_{1.}$	$\bar{y}_{1.}$
2	$y_{21} \ y_{22} \ \dots \ y_{2n}$	$y_{2.}$	$\bar{y}_{2.}$
.	.	.	.
.	.	.	.
a	$y_{a1} \ y_{a2} \ \dots \ y_{an}$	$y_{a.}$	$\bar{y}_{a.}$

6.3.1.1 Ventajas

Este diseño tiene varias ventajas, éstas son:

1. Permite flexibilidad completa, puede usarse cualquier número de tratamientos y de repeticiones. Puede variarse a voluntad el número de repeticiones de un tratamiento a otro (pero no es muy recomendable sin una buena razón). Todo el material disponible puede usarse, lo cual es una ventaja en experimentos preliminares pequeños donde el material experimental de que se dispone es escaso
2. El análisis estadístico es fácil, aun si el número de repeticiones no es el mismo para todos los tratamientos, o si los errores experimentales difieren de un tratamiento a otro.
3. Aun cuando los datos de algunos de las unidades o algunos tratamientos completos se hayan perdido, o se rechacen por alguna causa, el método de análisis sigue siendo sencillo. Por otra parte, la pérdida relativa de información debida a los datos faltantes, es de menos importancia que en cualquier otro diseño.

6.3.1.2 Modelo Estadístico Lineal

El modelo del análisis unifactorial, viene dado por :

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, 3, \dots, a \\ j = 1, 2, 3, \dots, n \end{cases} \quad 6.1$$

Donde:

y_{ij} = La observación ij ésima

μ = Es un parámetro común para todos los tratamientos, llamados media general.

τ_i = Es un parámetro asociado con el tratamiento i ésimo denominado efecto del tratamiento i ésimo.

ε_{ij} = Es el componente del error aleatorio.

Recibiendo el nombre de análisis de varianza de clasificación unidireccional, debido a que solo se investiga un factor, los errores del modelo se toman como variables aleatorias normal e independiente distribuidas con media cero y varianza σ^2 (NID $(0, \sigma^2)$). La varianza se considera constante para todos los niveles del factor.

6.3.1.3 Planteamiento de la Prueba de Hipótesis

El análisis de varianza para la clasificación unidireccional del modelo de efectos fijos de tratamientos τ_i se suelen definir como desviación de la media general, por lo que :

$$\sum_{i=1}^a \tau_i = 0 \quad 6.2$$

Se está interesado en probar la igualdad de los "a" efectos de tratamientos. Al emplear la ecuación (7.2), las hipótesis aprobadas son:

$$\begin{aligned} H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0 \\ H_1: \tau_i \neq 0 \text{ para al menos una } i. \end{aligned} \quad 6.3$$

Esto es, si la hipótesis nula es verdadera, entonces cada observación está integrada por la media general μ , más una realización del error aleatorio ϵ_{ij} .

6.3.1.4 Tratamiento Estadístico

Sea y_i la representación del total de las observaciones bajo el tratamiento y $y_{i.}$ la representación del promedio de las observaciones bajo el tratamiento i -ésimo. De modo similar, considere que $y_{..}$ representa el gran total de todas las observaciones y $y_{..}$ la gran media de todas las observaciones. Expresado matemáticamente de la siguiente forma:

$$\begin{aligned} y_i &= \sum_{j=1}^n y_{ij} & \bar{y}_i &= \frac{y_{i.}}{n} & i &= 1, 2, 3, \dots, a \\ y_{i..} &= \sum_{i=1}^a \sum_{j=1}^n y_{ij} & y_{..} &= \frac{y_{..}}{N} \end{aligned} \quad 6.4$$

Donde: $N = an$, es el número total de observaciones.

De tal modo la notación del subíndice "punto" implica la sumatoria sobre el subíndice que él reemplaza.

El Procedimiento de prueba para la hipótesis se llama análisis de Varianza. El término "análisis de varianza", resulta de partir la variabilidad total de los datos, en

sus partes componentes. La suma corregida total de los cuadrados, que es una medida de la variabilidad total en los datos, puede escribirse como:

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^n \left[\left(\bar{y}_{i.} - \bar{y}_{..} \right) + \left(y_{ij} - \bar{y}_{i.} \right) \right]^2$$

Por tanto, tenemos:

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 \quad 6.5$$

Mostrando la variabilidad total en los datos, en consecuencia se escribe simbólicamente como:

$$SS_T = SS_{\text{TRATAMIENTOS}} + SS_E \quad 6.6$$

Donde:

SS_T = Es la suma total de los cuadrados

$SS_{\text{TRATAMIENTOS}}$ = Es la suma de cuadrados debida a tratamientos (es decir, entre tratamientos).

SS_E = Es la suma de los cuadrados debida al error (esto es, dentro de los tratamientos).

En la tabla 6.2. Se muestra el análisis de varianza para el modelo de efectos fijos, teniendo en cuenta lo siguiente:

$$MS = \frac{SS_{\text{TRATAMIENTO}}}{(a-1)}$$

$$MS_E = \frac{SS_E}{(N-a)}$$

$$SS_E = SS_T - SS_{\text{TRATAMIENTO}}$$

6.7

$$SS_T = \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - \frac{y_{..}^2}{N}$$

$$SS_{\text{TRATAMIENTO}} = \sum_{i=1}^a \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N}$$

Tabla 6.2. El análisis de varianza para el modelo de efectos fijos de clasificación en un sentido.

Fuente de Variación	Suma de cuadrado	Grados de libertad	Media Cuadrática	F _o
Entre tratamientos	SS tratamiento	a-1	MS _{TRATAMIENTO}	MS _{TRATAMIENTO} /MS _E
Error dentro de los tratamientos	SS _E	N-a	MS _E	
Total	SS _T	N-1		

En consecuencia, se debe rechazar H₀ si la estadística de prueba es grande, de tal modo, rechazar H₀ Si :

$$F_o > F_{\alpha, a-1, N-a}$$

6.8

6.3.1.5 Estimación de Parámetros

Es posible obtener estimadores para los parámetros en el análisis en un sentido del modelo de la varianza.

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad 6.9$$

Estimar μ y τ_i , tales que la suma de los cuadrados de los errores o desviaciones ε_{ij} sea mínima. Esta estimación de parámetros recibe el nombre mínimos cuadrados, no es necesario suponer que ε_{ij} es normal teniendo:

$$L = \sum_{i=1}^a \sum_{j=1}^n \varepsilon_{ij}^2 = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \mu - \tau_i)^2 \quad 6.10$$

Esta ecuación que minimizan L . Los valores $\bar{\mu}$ y $\bar{\tau}_i$ con las soluciones a las $a+1$, ecuaciones simultáneas, denominadas ecuaciones normales de mínimos cuadrados. Resultando:

$$\begin{aligned} N \hat{\mu} + n \hat{\tau}_1 + n \hat{\tau}_2 + \dots + n \hat{\tau}_a &= y_{..} \\ n \hat{\mu} + n \hat{\tau}_1 &= y_{1.} \\ n \hat{\mu} + n \hat{\tau}_2 &= y_{2.} \\ &\dots \\ n \hat{\mu} + n \hat{\tau}_a &= y_{a.} \end{aligned} \quad 6.11$$

Considerando la restricción de los efectos de tratamientos como desviaciones de la media general, se tiene :

$$\sum_{i=1}^a \hat{\tau}_i = 0 \quad 6.12$$

La solución de para las ecuaciones normales son:

$$\begin{aligned}\hat{\mu} &= \hat{y}_{..} \\ \hat{\tau}_i &= \hat{y}_{i.} - \hat{y}_{..} \quad i=1,2,3,\dots, a\end{aligned}\tag{6.13}$$

Es evidente que esta solución no es la única, porque depende de la restricción que se elige. Restringiendo a $\mu + \tau_i$, puede estimarse por medio :

$$\hat{\mu} + \hat{\tau}_i = \hat{y}_{i.}\tag{6.14}$$

Ahora si se desea contribuir con intervalo de confianza para la media de tratamientos i ésima. La media de tratamiento i ésima es:

$$\mu_i = \mu + \tau_i \quad i = 1,2,3,\dots, a\tag{6.15}$$

Un estimador puntual de μ_i es: $\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \hat{y}_{i.}$. Luego se supone que los errores se distribuyen normalmente, cada $\hat{y}_{i.}$ es: $NID(\mu_i, \sigma^2/n)$. En consecuencia, un intervalo de confianza del $100(1-\alpha)$ por ciento en la media de tratamiento i ésima μ_i se basa en la distribución t, de la manera siguiente:

$$\left[\bar{y}_{i.} \pm t_{\alpha/2, N-a} \sqrt{MS_E / n} \right]\tag{6.16}$$

Y para la diferencia en cualesquiera dos medios de tratamiento, dígame $\mu_i - \mu_j$ sería:

$$\left[\bar{y}_{i.} - \bar{y}_{j.} \pm t_{\alpha/2, N-a} \sqrt{2MS_E / n} \right]\tag{6.17}$$

6.3.1.6 Análisis de Residuos y Validación del Modelo

El análisis de varianza del modelo supone que las observaciones se distribuyen normal e independientemente con la misma varianza en cada tratamiento o nivel de factor. Estas suposiciones deben unificarse examinando los residuos graficándolos en papel probabilístico normal.

Para comprobar la suposición de varianza iguales en cada nivel de factor, se grafican los residuos contra los niveles de factor y se compara la dispersión de los residuos. También es útil graficar los residuos contra \hat{y}_i (llamado algunas veces el Valor de Ajuste), la variabilidad de los residuos no debe depender de ningún modo del valor de se compara la dispersión de los residuos. También es útil graficar los residuos contra \hat{y}_i . Cuando compruebe aparece un patrón en estos gráficos suele indicar la necesidad de una transformación. Estos gráficos pueden ser:

- a. Gráfica de probabilidad normal de los residuos del experimento
- b. Gráfica de residuos contra tratamiento
- c. Gráfica de residuos contra \hat{y}_i

Una forma para comprender la suposición de normalidad consiste en hacer un histograma de los residuos. Si la suposición de que los errores son $NID(0, \sigma^2)$ satisface, esta gráfica debe ser semejante a la de una muestra extraída de una distribución normal centrada en cero.

- a. *Gráfica de Probabilidad normal de los Residuos del Experimento*
Consiste en construir una gráfica de probabilidad normal de los residuos. Una gráfica de este tipo es la representación de la distribución acumulada de los residuos sobre el papel de probabilidad normal, en otras palabras, es papel para gráficas cuya escala de ordenada es tal que la distribución acumulada sea una recta. Si la

distribución de los errores es normal, esta gráfica parecerá una línea recta.

Un defecto que a menudo se muestra en los gráficos de probabilidad normal es un residuo que es mucho mayor a los otros. Este residuo se denomina comúnmente residuo inusitado. Siendo una forma informal para detectar residuos distanciados es analizar los residuos estandarizados.

$$d_{ij} = \frac{\ell_{ij}}{\sqrt{MS_E}} \quad 6.18$$

Si los errores $\varepsilon_{ij} \sim N(0, \sigma^2)$, los residuos estandarizados deben ser aproximadamente normales con media cero y varianza a uno. Por lo tanto, aproximadamente el 68% de los residuos estandarizados deben encontrarse entre los límites de ± 1 , alrededor del 95% entre ± 2 , y prácticamente todos deben estar entre ± 3 . Un residuo a una distancia mayor que 304 desviaciones estándar del origen es potencialmente un residuo distanciado.

b. *Gráfica de residuos contra tratamientos*

La suposición de independencia puede verificarse graficando los residuos contra el tiempo y orden de la serie que se ejecutó el experimento. Una tendencia a tener residuos positivos y negativos indica una correlación positiva. Esto implica que la suposición de independencia ha sido violada. Un paso importante para asegurar la independencia es realizar un procedimiento apropiado de aleatorización. La gráfica debe presentar una disposición semejante en todos los tratamientos.

c. *Gráfica de Residuos contra \bar{y}_i .*

Si el modelo es correcto y las suposiciones se satisfacen los residuos no deben tener ningún patrón, ni deben estar relacionados con alguna otra variable, incluyendo la respuesta y_{ij} . Consiste en graficar los residuos contra los valores ajustados \bar{y}_{ij} (debe de recordarse que para el modelo en un sentido $\bar{y}_{ij} = \bar{y}_i$, el promedio del tratamiento i ésimo). En esta gráfica no debe revelarse ningún patrón obvio. Un defecto que en ocasiones revela la gráfica es el de una varianza variable. Debido a la escala de los instrumentos de medición, resultando una gráfica de los residuos contra \bar{y}_{ij} parecerá un embudo. El enfoque usual para tratarlo consiste en aplicar una transformación para igualar las varianzas, y volver a aplicar el análisis de varianza a los datos transformados.

6.3.1.7 Diseños Balanceados

En algunos experimentos de un solo factor el número de observaciones tomadas bajo cada tratamiento puede ser diferente. Entonces se dice, en ese caso que el diseño está desbalanceado. El análisis de varianza descrito antes sigue siendo, pero debe efectuarse ligeras modificaciones en las sumas de las fórmulas cuadradas. Sean n_i las observaciones tomadas bajo el tratamiento i ($i=1,2,\dots,a$) y

el número total de observaciones $N = \sum_{i=1}^a n_i$. Las relaciones para SS_T y

$SS_{\text{TRATAMIENTO}}$ se vuelve de la siguiente manera:

$$SS_T = \sum_{i=1}^a \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N} \quad 6.19$$

$$SS_{\text{TRATAMIENTO}} = \sum_{i=1}^a \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N}$$

Al resolver las ecuaciones normales se emplean las restricciones $\sum_{i=1}^a n_i \hat{\tau}_i = 0$. Ningún otro cambio se requiere en el análisis de varianza.

VENTAJAS AL ELEGIR UN DISEÑO BALANCEADO

- La estadística de prueba es relativamente insensible a las pequeñas desviaciones de la suposición de igualdades de varianza, si los tamaños de muestras son iguales.
- La potencia de la prueba se maximiza si las muestras son de igual tamaño.

6.3.1.8 Pruebas Estadísticas para Igualdad de Varianza.

Aparte de los gráficos de residuos que frecuentemente se usan para diagnosticar la igualdad en la varianza, se cuenta con algunas pruebas estadísticas. Estas pruebas formales para la hipótesis:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2 \quad 6.20$$

H_1 : lo anterior no es cierto al menos para una σ_i^2 .

Una técnica ampliamente usada es la prueba de Bartlett. Este procedimiento consiste en calcular una estadística cuya distribución muestral es aproximadamente Ji cuadrada con $a-1$ grados de libertad cuando las a muestras aleatorias provienen de poblaciones normales independientes. La estadística de prueba es:

$$\chi_0^2 = 2.3026 \frac{q}{c}$$

$$q = (N-1) \log_{10} Sp^2 - \sum_{i=1}^a (n_i - 1) \text{Log}_{10} S_i^2$$

$$c = 1 + \frac{1}{3(a-1)} \left(\sum_{i=1}^a (n_i - 1)^{-1} - (N-a)^{-1} \right) \quad 6.21$$

$$Sp^2 = \frac{\sum_{i=1}^a (n_i - 1) S_i^2}{N - a}$$

y S_i^2 es la varianza muestral de la i ésima población.

El criterio de rechazo de la hipótesis nula H_0 es:

$$\chi_0^2 > \chi_{\alpha, a-1}^2 \quad 6.22$$

6.3.2 MODELOS DE EFECTOS ALEATORIOS

En muchas situaciones, el factor de interés tiene una gran número de niveles posibles y se está interesado en extraer conclusiones acerca de la población completa d niveles del factor.

Si el experimentador selecciona al azar "a" de los niveles de la población de niveles del factor, entonces se afirma que el factor es aleatorio. Debido a la suposición de que los niveles del factor utilizados realmente en el experimento se eligieron en forma aleatoria, las conclusiones alcanzadas serán válidas a la población completa de los niveles del factor. Entonces se tiene que la población de niveles del factor es de tamaño grande como para considerarse infinito.

6.3.2.1 Modelo Estadístico Lineal

El modelo del análisis unifactorial viene dado por:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases} \quad 6.23$$

Donde: τ_i y ε_{ij} son variables aleatorias independientes.

Nótese que el modelo es idéntico en estructura al caso de efectos fijos, pero los parámetros tienen una interpretación diferentes.

Si τ_i tiene una varianza a σ_i^2 , y es independiente de ε_{ij} , la varianza de cualquier observación es:

$$V(y_{ij}) = \sigma_i^2 + \sigma^2 \quad 6.24$$

Las varianzas $\sigma_i^2 + \sigma^2$, se llaman componentes de varianza, y el modelo, recibe el nombre de componentes de varianza o modelo de efectos aleatorios.

6.3.2.2 Planteamiento de la Prueba de Hipótesis

Para probar la hipótesis en este modelo, se requiere que el ε_{ij} , sea $NID(0, \sigma^2)$, que el $\{\tau_i\}$ sea $NID(0, \sigma_i^2)$ y que τ_i y ε_{ij} sean independientes.

La identidad de la suma de los cuadrados es:

$$SS_T = SS_{\text{TRATAMIENTOS}} + SS_E \quad 6.24$$

En lugar de probar la hipótesis para efectos individuales, se prueba de la siguiente manera:

$$\begin{aligned} H_0: \sigma_r^2 &= 0 \\ H_1: \sigma_r^2 &> 0 \end{aligned} \quad 6.25$$

Si $\sigma_r^2 = 0$, todos los tratamientos son idénticos; pero si $\sigma_r^2 > 0$, entonces hay variabilidad entre tratamientos. SS_E / σ^2 se distribuyen como ji cuadrado con N-a grados de libertad, y bajo la hipótesis nula, $SS_{TRATAMIENTOS} / \sigma^2$ se distribuye como una ji cuadrado con a-1 grados de libertad. Ambas variables aleatorias son independientes.

6.3.2.3 Tratamiento Estadístico

Del modo que bajo la hipótesis nula, el cociente resulta:

$$F_0 = \frac{\frac{SS_{TRATAMIENTOS}}{a-1}}{\frac{SS_E}{N-a}} = \frac{MS_{TRATAMIENTO}}{MS_E}$$

Se distribuye como F con a-1 y N-a grados de libertad. Al examinar los cuadrados de media esperados, se puede determinar la región crítica para esta estadística es:

$$\begin{aligned} E(MS_{TRATAMIENTO}) &= \sigma^2 + n\sigma_r^2 \\ y \\ E(MS_E) &= \sigma^2 \end{aligned} \quad 6.26$$

De los dos cuadrados medios esperados, se ve que si H_0 es verdadera tanto el numerador como el denominador de la estadística de prueba, son estimadores insesgados de σ^2 , en tanto que si H_1 es verdadera, el valor esperado del numerador es mayor que el valor esperado del denominador. Entonces el criterio de rechazo de la hipótesis nula es:

$$F_0 > F_{\alpha, b-1, N-1} \quad 6.26$$

El procedimiento de cálculo y análisis de la tabla de varianza para el modelo de efectos aleatorios son idénticos al caso de efectos fijos, Las conclusiones, sin embargo, son bastante diferentes, debido a que ellos se aplican a la población completa de tratamientos

6.3.2.4 Estimación de Parámetros

El procedimiento utilizado para estimar $\sigma^2 y \sigma_r^2$, se llama el “método del análisis de varianza”, debido a que utiliza las líneas en el análisis de la tabla de varianza. No requiere la composición de normalidad con las observaciones. El procedimiento consiste en igualar los cuadrados de la media esperados con su valor observado en el análisis de la tabla de varianza, y en resolver con respecto a los componentes de la varianza, obteniendo:

$$MS_{TRATAMIENTO} = \sigma^2 + n\sigma_r^2$$

$$y \quad 6.27$$

$$MS_E = \sigma^2$$

En consecuencia, los estimadores de los componentes de varianza son:

$$\hat{\sigma}^2 = MS_E$$

$$y \quad 6.28$$

$$\hat{\sigma}_r^2 = \frac{MS_{TRATAMIENTO} - MS_E}{n}$$

6.3.2.5 Selección de Tamaño de Muestra.

En cualquier tipo de investigación del área de ingeniería química e ingeniería de alimentos, se encuentra con el problema en el diseño del experimento, con la selección del tamaño de muestras; en otras palabras, la determinación del número de réplicas que deben hacerse. En esta sección se estudiará varios métodos para determinar el tamaño apropiado de la muestra.

A pesar de que este análisis se hace para diseños de un factor, la metodología generada puede aplicarse a situaciones más complejas.

6.3.2.5.1 Curvas Características de Operación

Como una curva característica de operación es una gráfica de la probabilidad del error tipo II de una prueba estadística. Estas curvas pueden ser guías para el experimentador selecciones el número de réplicas con el objeto de que el diseño sea sensible a diferencias potenciales importantes entre tratamientos.

a. Efectos fijos

Para el caso de muestra del mismo tamaño en cada tratamiento, es:

$$\begin{aligned} \beta &= 1 - p \quad \{ \text{se rechaza } H_0 / H_0 \text{ es falsa} \} \\ \beta &= 1 - p \quad \{ F_0 > F_{\alpha, a-1, N-a} / H_0 \text{ es falsa} \} \end{aligned} \quad 6.29$$

Para tamaño de muestra desiguales, se reemplaza n en la ecuación (6.29) por:

$$n_0 = \frac{1}{a-1} \left[\sum_{i=1}^a n_i - \frac{\sum_{i=1}^a n_i^2}{\sum_{i=1}^a n_i} \right] \quad 6.30$$

Los estimadores que proporciona son los mejores, ya que son insesgados para σ^2 y σ_r^2 y (en otras palabras, de todas las funciones cuadráticas insesgadas de las observaciones, estos estimadores poseen la mínima varianza).

Las curvas características que se proporcionan en la tabla del apéndice I, son usadas para evaluar la ecuación anterior. En estas curvas, se indica la probabilidad del error tipo II (β) contra el parámetro ϕ , donde:

$$\phi_2 = \frac{n \sum_{i=1}^n r_i^2}{2\sigma^2} \quad 6.31$$

ϕ_2 está relacionado con el parámetro de descentralización. Las curvas proporcionan para $\alpha = 0.01$ y $\alpha = 0.05$, y para diversos valores de grados de libertad del numerador y del denominador.

b. *Para el modelo de efectos aleatorios.*

La probabilidad del error tipo II para este modelo es:

$$\begin{aligned} \beta &= 1 - p \quad \{ \text{se rechaza } H_0 / H_0 \text{ es falsa} \} \\ \beta &= 1 - p \quad \{ F_0 > F_{\alpha, a-1, N-a} / \sigma_r^2 > 0 \} \end{aligned} \quad 6.32$$

Dado que la probabilidad del error tipo II del modelo de efectos aleatorios está basado en la usual distribución centrada o proporcionando en el Apéndice " ", un conjunto de curvas para diversos valores de grados de libertad del numerador y del denominador, y para valores de $\alpha = 0.01$ y $\alpha = 0.05$. Las curvas graficará la probabilidad del error tipo II contra el parámetro ϕ , teniéndose:

$$\lambda = \sqrt{1 + \frac{n\sigma_\tau^2}{\sigma^2}} \quad 6.33$$

Donde: λ comprende los dos parámetros desarrollados σ^2 y σ_τ^2 .

En el diseño de bloques aleatorios, se reemplazan n por b , el número de bloques.

6.3.2.6 Método de Estimación por Intervalos de Confianza

En este método se supone que los resultados finales que el experimentador desea reportar aparecen en términos de intervalos de confianza y que está dispuestos a especificar la varianza de estos intervalos.

La ecuación que encuentra la precisión del intervalo es:

$$\pm t_{\alpha/2, N-a} \sqrt{\frac{2MS_E}{n}} \quad 6.34$$

6.3.3 DISEÑO DE BLOQUES ALEATORIOS

En muchos problemas de investigación en las Ingeniería Química e Ingeniería de Alimentos, es necesario diseñar experimentos en los que se pueda controlarse sistemáticamente la variabilidad producidas por diversas fuentes extrañas (variables perjudiciales) Esto dividiendo las observaciones en cada clasificación en bloques o grupos. Asumiendo que las fuentes conocidas de variabilidad (o sea variables extrañas) son constantes en cada bloque, pero varían de bloque a bloque.

El objetivo es de dividir en grupos el diseño del experimento es para reducir los errores experimentales y es aplicado cuando el material experimental no es

homogéneo, y es posible dividirlo. Esto debido que en el área de Ingeniería química e Ingeniería de alimentos se encuentran problemas en que se deberán diseñar experimentos para poblaciones no homogéneas.

6.3.3.1 Ventajas de su uso.

1. Por medio de la agrupación comúnmente se obtienen resultados más exactos que cuando se usan diseños completamente al azar.
2. Pueden incluirse cualquier número de tratamientos y repeticiones. Con lo cual cada tratamiento tendrá el mismo número de repeticiones y si desea hacer repeticiones adicionales para algunos tratamientos, cada uno de estos puede aplicarse a dos unidades dentro del grupo.
3. El análisis estadístico es el acostumbrado. Los cuidados que hagan necesario remisión de un grupo completo o de todos los datos de uno o más tratamientos no originan ninguna complicación.
4. Si la varianza del error experimental es mayor para algunos tratamientos que para otros puede obtenerse un error insesgado para probar cualquier combinación específica de las medias de los tratamientos.

6.3.3.2 Diseño Experimental

El diseño de bloques aleatorios es:

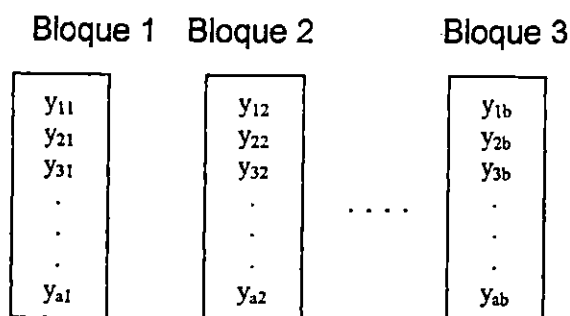


Figura. 6.1 . Diseño de bloque completamente aleatorio.

6.3.3.3 Modelo Estadístico Lineal

El modelo estadístico lineal es:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \end{cases} \quad 6.35$$

Donde:

μ : = es una media general

τ_i = es el efecto del tratamiento i ésimo

β_j = es el efecto del bloque j ésimo

ε_{ij} = el error aleatorio usado $NID(0, \sigma^2)$

6.3.3.4 Planteamiento de la Prueba de Hipótesis

Los tratamientos y los bloques se consideran al inicio como factores fijos. Además los efectos del bloque se definen como desviaciones respecto a la media general,

por lo que $\sum_{i=1}^a \tau_i = 0$ y $\sum_{j=1}^b \beta_j = 0$. Se está interesado en probar la igualdad de los

efectos del tratamiento. De la siguiente manera:

$$H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0$$

$$H_1: \tau_i \neq 0 \text{ al menos una } i$$

6.36

$$\text{Donde: } \mu_i = (1/b) \sum_{j=1}^b (\mu + \tau_i + \beta_j) = \mu + \tau_i$$

6.3.3.5 Tratamiento Estadístico

Sea y_i , el total de todas las observaciones tomadas bajo el tratamiento i , y_j , el total de todas las observaciones en el bloque j , $y_{..}$ el gran total de todas las observaciones, $y := ab$ el número total de observaciones. De modo similar \hat{y}_i , es el promedio de las observaciones tomadas bajo el tratamiento i , \hat{y}_j , es el promedio de las observaciones en el bloque j , $\hat{y}_{..}$ es el gran promedio de todas las observaciones. La suma corregida total al cuadrado es:

$$\sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_i - \bar{y}_{..}) + (\bar{y}_j - \bar{y}_{..}) + (y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}_{..})^2$$

ó

$$\sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2 = b \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2 + a \sum_{j=1}^b (\bar{y}_j - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}_{..})^2 \quad 6.37$$

ó simbólicamente:

$$SS_T = SS_{\text{TRATAMIENTO}} + SS_{\text{BLOQUE}} + SS_E \quad 6.38$$

El análisis de varianza se presenta en la tabla 6.3, la cual resume las fórmulas de cálculo para las sumas de cuadrados. El mismo procedimiento de prueba se utiliza

en caso en los que los tratamientos y/o bloques son aleatorios. El cambio surge en la interpretación de los resultados.

Tabla 6.3. Análisis de varianza para el diseño de bloques completamente aleatorio.

Fuente de Variación	Suma de Cuadrados	Grado de Libertad	Media Cuadrática	F ₀
Tratamiento	$\sum_{i=1}^a \frac{y_{i.}^2}{b} - y_{..}^2$	a - 1	$\frac{SS_{TRATAMIENTO}}{a - 1}$	$\frac{SS_{TRATAMIENTO}}{MS_E}$
Bloque	$\sum_{j=1}^b \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab}$	b - 1	$\frac{SS_{BLOQUE}}{b - 1}$	
Error	SS _E por sustracción	(a-1)(b-1)	$\frac{SS_E}{(a-1)(b-1)}$	
Total	$\sum_{i=1}^a \sum_{j=1}^b \left(y_{ij}^2 - \frac{y_{i.}^2}{b} - \frac{y_{.j}^2}{a} + \frac{y_{..}^2}{ab} \right)$	ab - 1		

El criterio de rechazo de la hipótesis nula (H₀), de la estadística utilizada es:

$$F_0 > F_{\alpha, a-1, (a-1)(b-1)} \quad 6.39$$

La región crítica es del extremo superior de la distribución F y se deberá rechazar la hipótesis nula (H₀).

También puede ser de interés la comparación entre las medias de los bloques, puede parecer que la hipótesis $H_0: \beta_{ij} = 0$ puede provocarse conformando la estadística $F_0 = MS_{\text{bloque}}/MS_E$ con $F_{\alpha, b-1, (a-1)(b-1)}$. Sin embargo, debe recordarse que la aleatorización fue aplicada sólo a los tratamientos dentro de los bloques

(representa una restricción para la aleatorización), pero que si los errores son NID $(0, \sigma^2)$, la estadística MS_{bloque}/MS_E puede hacerse para comparar las medias de bloque.

6.3.3.6 Pruebas Sobre Medias de Tratamientos Individual.

Cuando el análisis de la varianza indica que existe una diferencia entre medias de tratamientos, usualmente se necesita realizar algunas pruebas de seguimiento para aislar la diferencia específica. Cualquier método de comparación múltiple (prueba de intervalos múltiples de Duncan), puede utilizarse para esta.

El error estándar de una media de tratamiento es:

$$S_{\bar{y}_i} = \sqrt{\frac{MS_E}{b}} \quad 6.40$$

Si debe utilizar $(a-1)(b-1)$ grados de libertad en la tabla del apéndice .., para determinar los intervalos significativos, donde se obtienen valores $R_\alpha(P, f)$, para $p = 2, 3 \dots a$, donde α es el nivel de significancia y f es el número de grados de libertad para el error. Convirtiéndose estos intervalos en un conjunto de $a-1$, intervalos significativos menores (esto es, R_p), para $p = 2, 3 \dots a$, calculando se tiene :

$$R_p = r_\alpha(p, f) S_{\bar{y}_i}, \text{ para } p = 2, 3 \dots a \quad 6.41$$

6.3.3.7 Análisis Residual y Verificación.

En cualquier experimento diseñado es importante examinar los residuos y revisar las violaciones de las suposiciones básicas que podrán invalidar los resultados.

Los residuos para el diseño de bloque aleatorio son justo la diferencia entre los valores observados y los ajustados.

$$e_{ij} = y_{ij} - \bar{y}_{ij} \quad 6.42$$

Y los valores ajustados son:

$$\bar{y}_{ij} = \bar{y}_i - \bar{y}_j - \bar{y}_{..} \quad 6.43$$

El análisis de residuos se puede representar un gráfico, estas gráficas pueden ser:

- a. De probabilidad normal de residuos
- b. De residuos por tratamiento
- c. De residuos por bloque
- d. De residuos contra y_{ij} ,

6.3.3.8 Estimación de Parámetros y Prueba de Significancia

La estimación de parámetros de este modelo lineal, es:

$$\mu = \bar{y}_{..}$$

$$t_i = \bar{y}_i - \bar{y}_{..} \quad i = 1, 2, 3, \dots, a$$

$$\beta_j = \bar{y}_j - \bar{y}_{..} \quad j = 1, 2, 3, \dots, b$$

Al sustituir las ecuaciones normales se encuentra que los valores estimados o ajustados de y_{ij} son:

$$\hat{y}_{ij} = \mu + \tau_i + \beta_j \quad 6.44$$

$$\hat{y}_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..})$$

$$\hat{y}_{ij} = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}$$

6.3.4 ANÁLISIS DE COVARIANZA

El propósito de los métodos de las secciones 6.3.1 a 6.3.3, fue liberar el error experimental de la variabilidad, debida a causas identificables y controlables. En este método llamado análisis de varianza, que es aplicable cuando tales valores extraños o concomitantes no pueden mantenerse constantes, pero que, sin embargo, pueden medirse.

El método con el cual, se analiza los datos de esta clase es una combinación de regresión lineal y el análisis de varianza (diseños completamente aleatorizados).

La variable "x", está constituida por cada una de las observaciones tomada en cada unidad experimental antes de aplicar los tratamientos. Por ejemplo, en el experimento de mejorar una reacción estabilizada (x), sería la conversión inicial y la conversión final utilizando un catalizador, sería la observación de la variable dependiente "y".

6.3.4.1 Fundamentos del Análisis

Con este análisis se elimina la variabilidad, debido a la variación simultánea de las 2 variables consideradas, no solo en la varianza del error experimental, sino en las varianzas de los demás factores separados al hacer el análisis de varianza.

El análisis de varianza permite reducir el error experimental siempre que exista una relación de regresión entre la variación de la variable "x" y la variable "y", ya que mediante una ecuación de regresión se puede obtener los valores corregidos de las medias de los tratamientos en estudio.

6.3.4.2 Finalidad del Análisis

Las finalidades del análisis de covarianza son :

1. Ajustar los promedios cuya variación ha sido producida por la diferencia entre los promedios de la variable independiente.
2. Disminuir el error experimental, permitiendo de esta manera una mayor precisión del experimento.
3. Permite hacer una mejor interpretación de los resultados de los experimentos, dando a esto una mayor confiabilidad.

6.3.4.3 Asunciones de la Covarianza

El análisis de covarianza asume una serie de requisitos que le dan validez al análisis, estos son:

1. Los valores de la variable independiente no son afectados por los tratamientos asegurando lo anterior cuando la prueba de "f" para "x" es significativa.
2. Tanto la variable "x" la variable "y" tengan varianzas homogéneas.
3. La variable "x" y "y", deben tener distribución normal o sea que el error experimental debe ser normal e independiente con media cero, igual varianzas.
4. La regresión de "y" sobre "x" debe ser lineal

6.3.4.4 Método de Regresiones de Grupo.

Con este método consiste en la aplicación de la regresión a los diferentes tratamientos, con el objeto de:

- Evaluarse los tratamientos tienen diferentes o igual coeficiente de regresión: $b_1 = b_2 = b_3 = \dots = b_n$.
- Evaluar si los tratamientos tienen diferentes o igual ecuación de regresión y analizar si las rectas son paralelas.
- Comprobar que las varianzas para cada uno de los tratamientos es igual o diferente (varianza común).

6.3.4.5 Diseño Unifactorial con una Covariable

Si se supone que para un diseño experimental de un solo factor (completamente aleatorio), existe una relación lineal entre la respuesta y la covariable, el modelo estadístico apropiado es:

$$y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, 3, \dots, a \\ j = 1, 2, 3, \dots, n \end{cases} \quad 6.45$$

En donde:

y_{ij} = jésima observación de la variable respuesta formada bajo el i-ésimo tratamiento o nivel del factor único

x_{ij} = El valor de la covariación o variable concomitante correspondiente a y_{ij}

\bar{x} = Es el promedio de la s_{xj}

μ = La media general

τ_i = El efecto del i-ésimo tratamiento

β = El coeficiente de regresión lineal que indica la dependencia entre y_{ij} y x_{ij}

ε_{ij} = Un componente del error aleatorio

El modelo puede expresarse de la siguiente manera:

$$y_{ij} = \mu' + \tau_i + \beta x_{ij} + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, 3, \dots, a \\ j = 1, 2, 3, \dots, n \end{cases} \quad 6.46$$

Donde: $\mu' = \mu + \beta \bar{x}_{..}$

Con el objeto de descubrir el análisis se introduce la siguiente notación:

$$S_{yy} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - \frac{y_{..}^2}{an} \quad 6.47$$

$$S_{xx} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^n x_{ij}^2 - \frac{x_{..}^2}{an} \quad 6.48$$

$$S_{xy} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..}) = \sum_{i=1}^a \sum_{j=1}^n x_{ij} y_{ij} - \frac{x_{..} y_{..}}{an} \quad 6.49$$

$$T_{yy} = \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2 = \sum_{i=1}^a \frac{y_i^2}{n} - \frac{y_{..}^2}{an} \quad 6.50$$

$$T_{xx} = \sum_{i=1}^a (\bar{x}_i - \bar{x}_{..})^2 = \sum_{i=1}^a \frac{x_i^2}{n} - \frac{x_{..}^2}{an} \quad 6.51$$

$$T_{yy} = \sum_{i=1}^a (\bar{x}_i - \bar{x}_{..})(\bar{y}_i - \bar{y}_{..}) = \sum_{i=1}^a \frac{x_i y_i}{n} - \frac{x_{..} y_{..}}{an} \quad 6.52$$

$$E_{yy} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2 = S_{yy} - T_{yy} \quad 6.53$$

$$E_{xx} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 = S_{xx} - T_{xx} \quad 6.54$$

$$E_{xy} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i) = S_{xy} - T_{xy} \quad 6.55$$

En general, $S = T + E$, en donde S, T, y E sirven para representar las sumas de cuadrados y de productos cruzados del total, los tratamientos y el error, respectivamente. A diferencia de las sumas de cuadrados para x y y, que son positivas, las sumas de productos cruzados (xy) pueden ser negativas.

6.3.4.5.1 Planteamiento de la Hipótesis

Para probar la hipótesis de que no hay efecto de tratamiento, es:

$$H_0 : \tau_i = 0 \quad 6.56$$

6.3.4.5.2 Análisis de Covarianza

En la tabla 6.4, se presenta un análisis de covarianza como un análisis de varianza "ajustado".

Tabla 6.4. El análisis de covarianza como un análisis de varianza ajustado.

Fuente de Variación	Suma de cuadrados	Grados de libertad	Media del cuadrado	F ₀
Regresión	$(S_{xy})^2 / S_{xx}$	1		
Tratamiento	$SS'_E - SS_E = S_{yy} - (S_{xy})^2 / S_{xx} - [E_{yy} - (E_{xy})^2 / E_{xx}]$	a-1	$\frac{SS'_E - SS_E}{a-1}$	$\frac{(SS'_E - SS_E)/(a-1)}{MS_E}$
Error	$SS_E = E_{yy} - (E_{xy})^2 / E_{xx}$	a(n-1)-1	$MS_E = \frac{SS_E}{a(n-1)-1}$	
Total	S_{yy}	a n - 1		

Donde: $SS'_E = S_{yy} - (S_{xy})^2 / S_{xx}$ 6.57

Si no hubiese variable concomitante, $S_{xy} = S_{yx} = E_{xy} = E_{yx} = 0$. En este caso, la suma de cuadrados del error simplemente sería E_{yy} , y la suma de cuadrados de tratamientos sería $S_{yy} = E_{yy} = T_{yy}$.

Generalmente, los cálculos se presentan en una tabla de análisis de covarianza como la que aparece en la tabla 6.5. Se usa este esquema, porque muestra convenientemente las sumas requeridas de cuadrados y de productos cruzados, así como las sumas de cuadrados necesarios para probar la hipótesis de los efectos de tratamiento. Suele útil presentar las medias de tratamientos ajustados, estos se calculan de la siguiente manera:

$$\bar{y}_{i,\text{ajustado}} = \bar{y}_i - \bar{\beta}(\bar{x}_i - \bar{x}_{..}) \quad i = 1, 2, \dots, a \quad 6.58$$

Tabla 6.5. Análisis de covarianza para un experimento de un factor con una variable.

Fuente de Variación	Grados de libertad	Suma de cuadrados y productos			Ajustados para la regresión		
		x	xy	y	Y	Grados de libertad	Media de cuadrado
Tratamiento	a-1	Txx	Txy	Tyy			
Error	a(n-1)	Exx	Exy	Eyy	SSE=Eyy-(Ex) ² /Exx	a(n-1)-1	MSE= SSE/a(n-1)-1
Total	an-1	Sxx	Sxy	Syy	SSE=Syy -(Sxy) ² /Sxx	an-2	
Tratamiento ajustado					SSE - SSE	a-1	SSE-SSE/a-1

6.3.4.5.3 Estadística de Prueba

Si hipótesis nula tiene una distribución $F_{a-1, a(n-1)-1}$. Entonces, se realiza cuando $F_0 > F_{a-1, a(n-1)-1}$. El cálculo de F_0 viene dado por la relación siguiente:

$$F_0 = \frac{(SS'_E - SS_E)/(a-1)}{SS_E/[a(n-1)-1]} \quad 6.59$$

Esto para probar la hipótesis para el efecto de tratamiento.

Pero, para probar la hipótesis para el coeficiente de regresión del modelo $H_0: \beta=0$, se lleva a cabo usando la estadística siguiente:

$$F_0 = \frac{(E_{xy})^2 / E_{xx}}{MS_E} \quad 6.60$$

Donde : $\hat{\beta} = E_{xy} / E_{xx}$

La hipótesis nula se rechaza, si :

$$F_0 > F_{\alpha,1,a(n-1)-1}$$

6.3.4.5.4 Prueba de Idoneidad del Modelo

Las pruebas de diagnóstico para el modelo de covarianza se basa en el análisis de residuos y estos son:

$$\begin{aligned} e_{ij} &= y_{ij} - \hat{y}_{ij} \\ &= y_{ij} - \bar{y}_{i.} - \hat{\beta}(x_{ij} - \bar{x}_{i.}) \end{aligned} \quad 6.61$$

Las gráficas para el análisis de residuos puede ser:

- a. residuos contra valores ajustados \bar{y}_{ij}
- b. residuos contra la covarianza x_{ij}
- c. residuos contra los tratamientos
- d. probabilidad normal.

6.3.4.5.5 Determinación de Parámetros

De las ecuaciones normales son:

$$\mu_i: an \hat{\mu} + n \sum_{i=1}^a \hat{\tau}_i = y_{..}$$

$$\tau_i: n \hat{\mu} + n \hat{\tau}_i + \hat{\beta} \sum_{j=1}^n (x_{ij} - \bar{x}_{..}) = y_i, \quad i = 1, 2, 3, \dots, a$$

$$\beta: \sum_{i=1}^a \hat{\tau}_i \sum_{j=1}^n (x_{ij} - \bar{x}_{..}) + \hat{\beta} S_{xx} = S_{xy}$$

Asumiendo :

$$\sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{..}) = 0 \quad y \quad \sum_{i=1}^a \hat{\tau}_i = 0,$$

6.62

Se tiene :

$$\hat{\mu} = \bar{y}_{..}$$

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})$$

6.63

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}}$$

6.3.4.6 Otros Modelos de Covarianza

Considere el diseño aleatorizado por bloques completos con una covariable.

$$y_{ij} = \mu + \tau_i + \gamma_j + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, 3, \dots, a \\ j = 1, 2, 3, \dots, b \end{cases} \quad 6.64$$

- Donde:
- y_{ij} = La observación del tratamiento i y el bloque j
 - μ = la media general
 - β = es el efecto del i -ésimo tratamiento
 - γ_j = el efecto del j -ésimo valor observado de la covarianza
 - ε_{ij} = error aleatorio con NID $(0, \sigma^2)$.

Las definiciones para las cantidades presentadas en la tabla 6.6 son:

$$R_{xx} = \sum_{j=1}^a \frac{x_{.j}^2}{a} - \frac{x_{..}^2}{ab} \quad 6.65$$

$$R_{yy} = \sum_{i=1}^b \frac{y_{i.}^2}{a} - \frac{y_{..}^2}{ab} \quad 6.66$$

$$R_{xy} = \sum_{ij} \frac{x_{ij}y_{ij}}{a} - \frac{x_{..}y_{..}}{ab} \quad 6.67$$

$$T_{xx} = \sum \frac{x_{i.}^2}{b} - \frac{x_{..}^2}{ab} \quad 6.68$$

$$T_{yy} = \sum \frac{y_{.i}^2}{b} - \frac{y_{..}^2}{ab} \quad 6.69$$

$$T_{xy} = \sum \frac{x_{.i}y_{.i}}{b} - \frac{x_{..}y_{..}}{ab} \quad 6.70$$

$$E_{xx} = S_{xx} - R_{xx} - T_{xx} \quad 6.71$$

$$E_{yy} = S_{yy} - R_{yy} - T_{yy} \quad 6.72$$

$$E_{xy} = S_{xy} - R_{xy} - T_{xy} \quad 6.73$$

El cálculo del parámetro de regresión B se estima mediante:

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}} \quad 6.74$$

Y la estadística de prueba de hipótesis $H_0 : \beta = 0$ es :

$$F_0 = \frac{(E_{xy})^2/E_{xx}}{MS_E} \quad 6.75$$

Cuando, $F_0 > F_{\alpha, 1, ab-a-b}$, se rechaza la hipótesis nula.

El resumen del análisis de covarianza se presenta en la tabla 6.6

Tabla 6.6. Análisis de covarianza para el diseño aleatorizado por bloques completos

Fuente de variación	Grados de libertad	Suma de cuadrados			Ajustado para regresión		F ₀	
		X	XY	Y	Y	Grados de libertad		Media de cuadrados
Bloque	b-1	R _{xx}	R _{xy}	R _{yy}	-	-	_____	_____
Tratamiento	a-1	T _{xx}	T _{xy}	T _{yy}	-	-	_____	_____
Error	(a-1)(b-1)	E _{xx}	E _{xy}	E _{yy}	SS _E =E _{yy} - (E _{xy}) ² /E _{xx}	ab-a-b	$MS_E = \frac{SS_E}{ab-a-b}$	_____
Total	ab-1	S _{xx}	S _{xy}	S _{yy}	-	ab-2	_____	_____
Tratamiento mas error	b(a-1)	S _{xx} ^t =T _{xx} + E _{xx}	S _{xy} ^t =T _{xy} + E _{xy}	S _{yy} ^t =T _{yy} + E _{yy}	SS _E ^t =S _{yy} ^t - (S _{xy} ^t) ² /S _{xx} ^t	b(a-1)- 1	_____	_____
Tratamiento ajustado	_____	_____	_____	_____	SS _E ^t -SS _E	a-1	$MS_A = \frac{SS_E^t - SS_E}{a-1}$	$\frac{MS_A}{MS_E}$

El análisis de covarianza puede ser generalizado a los bloques incompletos, con efectos al azar, factoriales, etc. Como también para comparar las pendientes de varias rectas de regresión. Finalmente, pueden existir situaciones en las que se requiere una estructura de covarianza múltiple, ya sea porque la relación entre x e y no es lineal.

6.4 ANÁLISIS Y DISEÑO DE EXPERIMENTO CON VARIOS FACTORES

En muchos experimentos en Ingeniería Química e Ingeniería de Alimentos, se llevan a cabo para estudiar los efectos producidos por dos o más factores. Los diseños factoriales son los más eficientes para este tipo de experimentos. Por diseño factorial se entiende aquel en el que se investigan todas las posibles combinaciones de los niveles de los factores en cada ensayo completo o réplica del experimento.

Esta sección presenta técnicas de diseños de experimentos útiles cuando están involucrados varios factores.

6.4.1 VENTAJAS

1. Facilita el estudio de varios factores simultáneamente
2. El número de observaciones necesarias para estudiar cada factor independiente es mayor en relación a cualquier diseño simple.
3. Proporciona una información más completa de cada factor en estudio; pues no solo se comprueban los efectos principales de cada uno, sino también las interacciones entre si.
4. Las conclusiones pueden tener mayor validez desde el punto de vista práctico, pues los distintos niveles de cada factor se estudian en presencia de los niveles del factor u otros factores.
5. Presenta un ahorro considerable de tiempo y material dedicado a la investigación

6.4.2 DESVENTAJAS:

1. Conforme aumentan las combinaciones de tratamientos, el error experimental también se aumenta, ya que pueden existir mayores problemas y causas de error en la ejecución de las actividades y en la toma de datos; por lo tanto, se necesita personal capacitado para su ejecución.
2. El análisis estadístico es más complicado y más difícil de interpretar.

6.4.3 MODELOS PARA DISEÑOS FACTORIAL

En esta sección, se hará un estudio y análisis para los modelos siguientes:

1. Experimentos de dos factoriales
2. Experimentos multifactoriales
3. Experimentos factoriales 2^n
4. Confusión en un experimento factorial 2^n
5. Réplica fraccional

En la tabla 6.6. se resumirá los modelos lineales de los diseños factoriales y una descripción de ellos.

6.4.3.1 Recolección de las Observaciones para Diseños Factoriales

La tabla 6.7, muestra la forma de recolectar las observaciones o mediciones hechas en el experimento, para diseños bifactoriales con efectos fijos, aleatorios y mixtos.

Tabla 6.6. Modelos de experimentos factoriales

Diseño	Modelos	Hipótesis nula	Descripción
<p>Experimento de dos factores completamente aleatorio (Efecto fijo)</p>	$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ij}$ <p>Con: $i=1,2,3,\dots,a$ $j=1,2,3,\dots,b$ $k=1,2,3,\dots,n$</p> <p>μ=Efecto de media general τ_i=Efecto del nivel i-esimo del factor A β_j=Efecto del nivel j-ésimo del factor B $(\tau\beta)_{ij}$=Efecto de interacción entre A y B ε_{ijk}=Error aleatorio con NID $(0, \sigma^2)$</p>	<p>$H_0: \tau_i = 0$ (ningún efecto del factor renglón)</p> <p>$H_0: \beta_j = 0$ (ningún efecto del factor columna)</p> <p>$H_0: (\tau\beta)_{ij} = 0$ (ningún efecto de interacción)</p>	<p>Es apropiado para experimento de dos efectos con media general cero y varianza, sin efecto del factor renglón, columna y de interacción. Existen niveles "A y B, niveles del factor B. Con n réplicas del experimento y todos ellos contienen las ab combinaciones de tratamiento.</p>
<p>Experimento de dos factores en efectos aleatorio</p>	$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ij}$ <p>Donde: $\tau_i, \beta_j, (\tau\beta)_{ij}$, y ε_{ijk} Son variables aleatorias. Con: NID $(0, \sigma^2_{\tau})$, NID $(0, \sigma^2_{\beta})$, NID $(0, \sigma^2_{\tau\beta})$ y NID $(0, \sigma^2)$ para ε_{ijk}. Llamadas componentes de varianza.</p> <p>Por lo tanto:</p>	<p>$H_0: \sigma^2_{\tau} = 0$</p> <p>$H_0: \sigma^2_{\beta} = 0$</p> <p>$H_0: \sigma^2_{\tau\beta} = 0$</p>	<p>Cuando los niveles de los factores A y B se eligen al azar de poblaciones más grandes llamados modelo de efectos aleatorios o de componentes de variación. Las inferencias pueden generalizarse a todos los niveles de las poblaciones bajo estudio, porque los niveles a niveles de A y B</p>

Continuación de la Tabla 6.6.

<p>Experimento factorial de dos factores mixtos</p>	$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ij}$ <p>Con: $i=1,2,3,\dots,a$ $j=1,2,3,\dots,b$ $k=1,2,3,\dots,n$</p> <p>Definido: $\sum_{i=1}^a \tau_i = 0$ β_j con NID $(0, \sigma^2_{\beta})$ ε_{ijk} con NID $(0, \sigma^2)$ $(\tau\beta)_{ij}$ con NID $(0, ((a-1)/a)\sigma^2_{\tau\beta})$</p>	<p>$H_0: \tau_i = 0$ $H_0: \sigma^2_{\beta} = 0$ $H_0: \sigma^2_{\tau\beta} = 0$</p>	<p>Cuando se considera la situación en la que uno de los factores A es fijo, mientras que el otro B es aleatorio.</p> <p>No siendo todos independientes, es decir que la suma del componente sobre el efecto es igual a cero.</p>
<p>Diseño factorial general (tres sectores) con efectos fijos</p>	$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijkl}$ <p>$k=1,2,3,\dots,c$ $l=1,2,3,\dots,n$</p> <p>con: γ_k = Efecto del nivel k-ésimo del factor C $(\tau\beta)_{ij}$ = Efecto de interacción de A y B $(\tau\gamma)_{ik}$ = Efecto de interacción de A y C</p>	<p>$H_0: \tau_i = 0$ $H_0: \beta_j = 0$ $H_0: \gamma_k = 0$ $H_0: (\tau\beta)_{ij} = 0$ $H_0: (\tau\gamma)_{ik} = 0$ $H_0: (\beta\gamma)_{jk} = 0$ $H_0: (\tau\beta\gamma)_{ijk} = 0$</p>	<p>Suponiendo que los factores A, B y C, son fijos, donde debe tener al menos dos réplicas para calcular una suma de cuadrados del error.</p> <p>Para el caso, se tiene a niveles del factor A, b niveles del factor B, c niveles del factor C, arreglados en un experimento. en general habrá abc...n observaciones totales, si hay n réplicas del experimento completo.</p>

Tabla 6.7. Disposición general para recolección de datos para un diseño bifactorial.

		FACTOR B			
		1	2	...	b
FACTOR A	1	$Y_{111}, Y_{112}, \dots, Y_{11n}$	$Y_{121}, Y_{122}, \dots, Y_{12n}$...	$Y_{1b1}, Y_{1b2}, \dots, Y_{1bn}$
	2	$Y_{211}, Y_{212}, \dots, Y_{21n}$	$Y_{221}, Y_{222}, \dots, Y_{22n}$...	$Y_{2b1}, Y_{2b2}, \dots, Y_{2bn}$

	a	$Y_{a11}, Y_{a12}, \dots, Y_{a1n}$	$Y_{a21}, Y_{a22}, \dots, Y_{a2n}$...	$Y_{ab1}, Y_{ab2}, \dots, Y_{abn}$

6.4.3.2 Análisis de datos (análisis de varianza)

El análisis de varianza para diseños bifactoriales; con efectos fijos, aleatorios y mixtos se presenta las tablas 6.8 y para diseñar factoriales generales, se hace el análisis de varianza en la tabla 6.10.

Donde:

$$SS_{Subtotales} = \sum_{i=1}^a \sum_{j=1}^b \frac{y_{ij}^2}{n} - \frac{y_{...}^2}{abn}$$

Tabla 6.8. Análisis de variancia del modelo bifactorial de efectos fijos y aleatorios.

FUENTE DE VARIACION	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	MEDIA DE CUADRADOS	F ₀
Tratamiento A	$SS_A = \sum_{i=1}^a \frac{y_i^2}{bn} - \frac{y_{...}^2}{abn}$	a-1	$MS_A = \frac{SS_A}{a-1}$	$\frac{MS_A}{MS_E}$
Tratamiento B	$SS_B = \sum_{j=1}^b \frac{y_j^2}{an} - \frac{y_{...}^2}{abn}$	b-1	$MS_B = \frac{SS_B}{b-1}$	$\frac{MS_B}{MS_E}$
Interacción	$SS_{AB} = SS_{Subtotales} - SS_A - SS_B$	(a-1)(b-1)	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_E}$
Error	$SS_E = SS_T - SS_{subtotales}$	ab(n-1)	$MS_E = \frac{SS_E}{ab(n-1)}$	
Total	$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}^2 - \frac{y_{...}^2}{abn}$	abn - 1		

Donde: F de tablas es:

$F_{(a-1)(a-1)(b-1)}$, $F_{(b-1)(a-1)(b-1)}$ y $F_{(a-1)(b-1)}$, $ab(n-1)$ respectivamente.

Donde:

$$SS_{Subtotales(AB)} = \sum_{i=1}^a \sum_{j=1}^b \frac{y_{ij..}^2}{cn} - \frac{y_{...}^2}{abcn}$$

$$SS_{Subtotales(AC)} = \sum_{i=1}^a \sum_{k=1}^c \frac{y_{i.k.}^2}{bn} - \frac{y_{...}^2}{abcn}$$

$$SS_{Subtotales(BC)} = \sum_{j=1}^b \sum_{k=1}^c \frac{y_{.jk.}^2}{an} - \frac{y_{...}^2}{abcn}$$

$$SS_{Subtotales(ABC)} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \frac{y_{ijk.}^2}{n} - \frac{y_{...}^2}{abcn}$$

y F₀ de tabla se calcula:

Tratamiento A : $F_{(a-1), (a-1)(b-1)(c-1)}$
 Tratamiento B : $F_{(b-1), (a-1)(b-1)(c-1)}$
 Tratamiento C : $F_{(c-1), (a-1)(b-1)(c-1)}$
 Interacción AB : $F_{(a-1)(b-1), (a-1)(b-1)(c-1)}$
 Interacción AC : $F_{(a-1)(c-1), (a-1)(b-1)(c-1)}$
 Interacción BC : $F_{(b-1)(c-1), (a-1)(b-1)(c-1)}$
 Interacción ABC : $F_{(a-1)(b-1)(c-1), abc(n-1)}$

Tabla 6.9. Resumen de análisis de varianza para el diseño trifactorial

Fuente de variación	Suma de cuadrados	Grados de libertad	Media de cuadrados	F ₀
A	$SS_A = \sum_{i=1}^a \frac{y_{i...}^2}{bcn} - \frac{y_{...}^2}{abcn}$	a - 1	$MS_A = \frac{SS_A}{a-1}$	$\frac{MS_A}{MS_E}$
B	$SS_B = \sum_{j=1}^b \frac{y_{.j..}^2}{acn} - \frac{y_{...}^2}{abcn}$	b - 1	$MS_B = \frac{SS_B}{b-1}$	$\frac{MS_B}{MS_E}$
C	$SS_C = \sum_{k=1}^c \frac{y_{..k.}^2}{abn} - \frac{y_{...}^2}{abcn}$	c - 1	$MS_C = \frac{SS_C}{c-1}$	$\frac{MS_C}{MS_E}$
AB	$SS_{AB} = SS_{subtotales(AB)} - SS_A - SS_B$	(a-1)(b-1)	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_E}$
AC	$SS_{AC} = SS_{subtotales(AC)} - SS_A - SS_C$	(a-1)(c-1)	$MS_{AC} = \frac{SS_{AC}}{(a-1)(c-1)}$	$\frac{MS_{AC}}{MS_E}$
BC	$SS_{BC} = SS_{subtotales(BC)} - SS_B - SS_C$	(b-1)(c-1)	$MS_{BC} = \frac{SS_{BC}}{(b-1)(c-1)}$	$\frac{MS_{BC}}{MS_E}$
ABC	$SS_{ABC} = SS_{subtotales(ABC)} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}$	(a-1)(b-1)(c-1)	$MS_{ABC} = \frac{SS_{ABC}}{(a-1)(b-1)(c-1)}$	$\frac{MS_{ABC}}{MS_E}$
ERROR	$SS_E = SS_T - SS_{subtotales(ABC)}$	abc(n-1)	$MS_E = \frac{SS_E}{abc(n-1)}$	
TOTAL	$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n y_{ijkl}^2 - \frac{y_{...}^2}{abcn}$	abcn - 1		

6.4.3.3 Estimación de la parámetros de los modelos factoriales

a. Diseño bifactorial con efecto fijo, los parámetros son:

$$\begin{aligned}
 \mu &= \bar{y}_{...} \\
 \tau_i &= \bar{y}_{i..} - \bar{y}_{...} & i = 1, 2, \dots, a \\
 \beta_j &= \bar{y}_{.j.} - \bar{y}_{...} & j = 1, 2, \dots, b \\
 (\tau\beta)_{ij} &= \bar{y}_{ij.} - \bar{y}_{...} & \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \end{cases}
 \end{aligned} \tag{6.76}$$

Para determinar el valor ajustado, se realiza de la siguiente forma:

$$\bar{y}_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} = \bar{y}_{ij}. \quad 6.77$$

- b. Diseño bifactorial con efectos aleatorio, los parámetros son: como para probar las hipótesis nulas es necesario determinar antes los valores esperados.

$$\sigma^2 = MS_E$$

$$\sigma_{\tau\beta}^2 = \frac{MS_{AB} - MS_E}{n}$$

$$\sigma_{\beta}^2 = \frac{MS_B - MS_E}{an}$$

$$\sigma_{\tau}^2 = \frac{MS_A - MS_E}{bn}$$

- c. Diseño bifactorial mixto, con efecto fijo para el factor A, se obtiene los parámetros mostrados en las ecuaciones (7.91). Para componentes de variancia $\sigma_{\beta}^2, \sigma_{\tau\beta}$ y σ^2

Nota: $\sigma_{\tau} = 0$ Por ser el efecto fijo y no independiente. Para los factores fijos es:

$$\mu = \bar{y}_{...}$$

$$\tau = \bar{y}_{i..} - \bar{y}_{...} \quad i = 1, 2, \dots, a \quad 6.78$$

- d. Diseño trifactorial con efectos fijo

Se obtiene los parámetros siguientes:

$$\mu = \bar{y}_{...}$$

$$\tau = \bar{y}_{i...} - \bar{y}_{...} \quad i = 1, 2, \dots, a$$

$$\beta_j = \bar{y}_{.j.} - \bar{y}_{...} \quad j = 1, 2, \dots, b$$

$$\gamma_k = \bar{y}_{.k.} - \bar{y}_{...} \quad k = 1, 2, \dots, c$$

$$(\tau\beta)_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \quad 6.79$$

$$(\tau\gamma)_{ik} = \bar{y}_{i.k.} - \bar{y}_{i..} - \bar{y}_{.k.} + \bar{y}_{...}$$

$$(\beta\gamma)_{jk} = \bar{y}_{.jk.} - \bar{y}_{.j.} - \bar{y}_{.k.} + \bar{y}_{...}$$

$$(\tau\beta\gamma)_{ijk} = \bar{y}_{ijk.} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{.k.} + \bar{y}_{...}$$

6.4.3.4 Elección de tamaño de muestra.

a. Para efectos (fijos bifactorial)

Es posible usar las curvas de operación característica V del Apéndice y una forma para usar estas curvas es determinar el valor mínimo de ϕ^2 que corresponde a una diferencia especificada entre dos medias de tratamiento (D), se tiene:

$$\phi_2 = \frac{nbD^2}{2a\sigma^2} \quad 6.80$$

Mientras que si la diferencia entre dos medias de columnas es D, el valor mínimo será:

$$\phi_2 = \frac{naD^2}{2b\sigma^2} \quad 6.81$$

Finalmente la diferencia es igual a D entre cualquier par de efectos de interacción; están será:

$$\phi_2 = \frac{nD^2}{2\sigma^2(a-1)(b-1)-1} \quad 6.82$$

En la tabla 6.10, se presenta los parámetros de la curvas características de operación, para el diagrama Anexo O.

Tabla 6.10. Parámetros de la curva característica para el diagrama O del Anexo. Modelo bifactorial de efectos fijos.

FACTOR	ϕ_2	GRADOS DE LIBERTAD DEL NUMERADOR	GRADOS DE LIBERTAD DEL DENOMINADOR
A	$\frac{bn \sum_{i=1}^a \tau_i^2}{a\sigma^2}$	a - 1	a b (n - 1)
B	$\frac{an \sum_{j=1}^b \beta_j^2}{b\sigma^2}$	b - 1	a b (n - 1)
AB	$\frac{n \sum_{i=1}^a \sum_{j=1}^b (\tau\beta)_{ij}^2}{\sigma^2 [(a-1)(b-1)+1]}$	(a - 1) (b - 1)	a b (n - 1)

- b. Para un modelo aleatorio de clasificación en dos sentidos y para el análisis de varianza del modelo mixto, se hace uso de las curvas de operación característica para el modelo aleatorio debe usarse el diagrama del Anexo Q. Y para el modelo bifactorial mixto debe usarse los diagramas del Anexo O y Q simultáneamente, tal como se muestra en la tabla 6.11.

Tabla 6.11. Parámetros de la curva característica de operación para las tablas O y Q del Anexo . Modelos bifactoriales mixtos y aleatorios.

Modelo de Efectos Aleatorios

FACTOR	λ	GRADOS DE LIBERTAD DEL NUMERADOR	GRADOS DE LIBERTAD DEL DENOMINADOR
A	$\sqrt{1 + \frac{bn\sigma_{\tau}^2}{\sigma^2 + n\sigma_{\tau\beta}^2}}$	a - 1	(a - 1)(b - 1)
B	$\sqrt{1 + \frac{an\sigma_{\beta}^2}{\sigma^2 + n\sigma_{\tau\beta}^2}}$	b - 1	(a - 1)(b - 1)
AB	$\sqrt{1 + \frac{n\sigma_{\tau\beta}^2}{\sigma^2}}$	(a - 1)(b - 1)	ab(n - 1)

Modelo de Efectos Mixos

FACTOR	λ	GRADOS DE LIBERTAD DEL NUMERADOR	GRADOS DE LIBERTAD DEL DENOMINADOR	DIAGRAMA DEL APENDICE
A (Fijo)	$\phi_2 = \frac{bn\sum_{i=1}^a \tau_i^2}{a(\sigma^2 + n\sigma_{\tau\beta}^2)}$	a - 1	(a - 1)(b - 1)	V
B (Aleatorio)	$\sqrt{1 + \frac{an\sigma_{\beta}^2}{\sigma^2}}$	b - 1	ab(n - 1)	VI
AB	$\sqrt{1 + \frac{n\sigma_{\tau\beta}^2}{\sigma^2}}$	(a - 1)(b - 1)	ab(n - 1)	VI

6.4.3.5 Identidad del modelo (para efectos fijos).

Antes de poder adoptar las conclusiones del análisis de varianza, debe probarse la adecuación del modelo supuesto. Para el caso la herramienta principal es el análisis de residuos. Los residuos para el caso de un modelo factorial de dos factores es:

$$e_{ijk} = y_{ijk} - \bar{y}_{ijk} \quad 6.83$$

Para ello es necesario graficar los residuos con diversos tipos de gráficas, estos son:

- a. Gráfica de probabilidad normal
- b. Gráfica de residuos contra los valores estimados (\hat{y}_{ijk})
- c. Gráfica de residuos contra cada uno de los factores del experimento.

6.4.4 DISEÑOS FACTORIAL 2^k (CADA FACTOR CON DOS NIVELES)

Hay ciertos criterios especiales de diseños factoriales que son muy útiles en la investigación y las áreas de investigación de la Ingeniería Química e Ingeniería de Alimentos, no es la excepción. Uno de estos casos especiales es un diseño factorial con k factores, cada uno en dos niveles.

El diseño 2^k es particularmente útil en los primeras fases del trabajo experimental, cuando es probable que haya muchos factores por investigar. Debido a que solo hay dos niveles para cada factor, debe suponerse que la respuesta es aproximadamente lineal en el intervalo de los niveles elegidos de los factores.

Además estos diseños tienen un análisis estadístico sumamente simplificado y forma la base de muchos otros diseños útiles.

6.4.4.1 Diseños 2^2

El tipo más simple de diseños 2^k es el 2^2 ; esto significa dos factores A y B, cada uno en dos niveles. Se suele considerar como los niveles "bajos" y "alto" del factor.

Adviertese que el diseño puede presentarse geoméricamente como un cuadrado con los $2^2 = 4$ ejecuciones formando las esquinas de un cuadrado. Tal como e muestra en la figura 6.2.

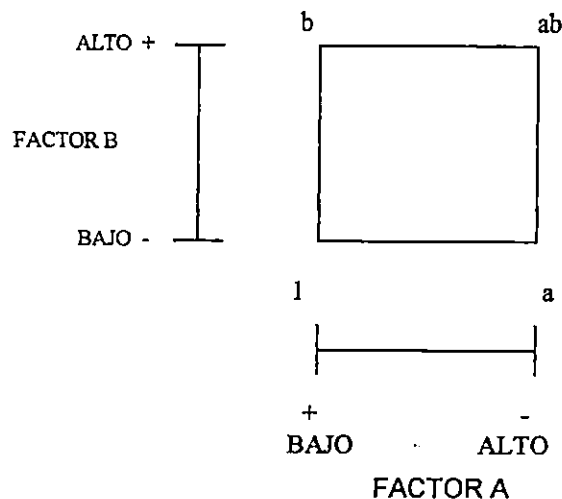


Figura. 6.2. Combinación de tratamientos en el diseño 2^2

En la tabla 6.12. se presenta una tabla de signos más o menos puede utilizarse para determinar el signo en cada combinación de tratamiento para un contraste particular.

En consecuencia, las sumas de cuadrados para A, B y AB es:

$$SS_A = \frac{[a + ab - b - (1)]^2}{4n}$$

$$SS_B = \frac{[b + ab - a - (1)]^2}{4n}$$

$$SS_{AB} = \frac{[ab + (1) - a - b]^2}{4n}$$

La suma total de cuadrados se determina de la siguiente forma:

$$SS_T = \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^n y_{ijk}^2 - \frac{y_{...}^2}{4n}$$

Tabla 6.12. Signos para los efectos en el diseño 2².

Combinación de Tratamiento	Efecto factorial			
	I	A	B	AB
(A)	+	-	-	+
a	+	+	+	-
b	+	-	-	-
c	+	+	-	+

Donde SST tiene 4n-1 grados de libertad. La suma de cuadrados del error, con 4(n-1) grados de libertad, se puede calcular por diferencia de la siguiente manera:

$$SS_E = SS_T - SS_A - SS_B - SS_{AB}$$

6.4.4.1.1 Recolección de datos

En la tabla 6.13, se ilustra la forma de recolectar las observaciones obtenidas del experimento.

Tabla 6.13. Tabla resumen para la observación o recolección de datos para un diseño factorial 2^2 .

No. De Corrida	Combinación de Tratamientos	Factor		Réplica		
		A	B	1	2
1	A baja, B baja (1)	-	-	n		
2	A alta, B baja (a)	+	-			
3	A baja, B alta (b)	-	+			
4	A alta, B alta (ab)	+	+			

6.4.4.1.2 Análisis de Varianza

En la tabla 6.14, se presenta una forma de presentar el análisis de varianza para un diseño factorial 2^2 .

Tabla 6.14. Análisis de varianza para un modelo factorial 2^2 .

Fuente de Variación	Suma de cuadrados	Grados de libertad	Media Cuadrática	F_0
A	SSA	1	$MS_A = \frac{SS_A}{1}$	$\frac{MS_A}{MS_E}$
B	SSB	1	$MS_B = \frac{SS_B}{1}$	$\frac{MS_B}{MS_E}$
AB	SSAB	1	$MS_{AB} = \frac{SS_{AB}}{1}$	$\frac{MS_{AB}}{MS_E}$
Error	SSE	$4(n-1)$	$MS_E = \frac{SS_E}{4(n-1)}$	
Total	SST	$4n-1$		

6.4.41.3 *Análisis de Residuos y de Identidad del Modelo*

Para un diseño 2^k , se obtienen los valores ajustados por medio de un modelo de regresión de los datos de la siguiente manera:

1. A través del análisis de varianza se puede visualizar los factores activos y pasivos. Entonces se toman los factores activos
2. Los niveles bajos se le asignan un valor (-1) y a los altos (+1)
3. La ordenada al origen B_0 se determina calculando el gran promedio total
4. Las pendientes del modelo de regresión se determina como un medio de las observaciones de los factores activos
5. Se calcula los valores ajustados por medio del modelo de regresión.

Para comprobar la idoneidad del modelo, se calcula los residuos de las observaciones menos el valor ajustado y luego se grafican los residuos en papel probabilístico y contra los valores ajustados y los factores A y B en forma individual.

6.4.4.2 **Diseño 2^K Para Factores ≥ 3 .**

Los métodos presentados en la sección previa para los diseños factoriales con $k = 2$ factores cada uno con dos niveles pueden extenderse con facilidad a más de dos factores. Por ejemplo 2^3 , y tiene ocho combinaciones de tratamientos. Geométricamente, el diseño es un cubo como se muestra en la figura 6.3.

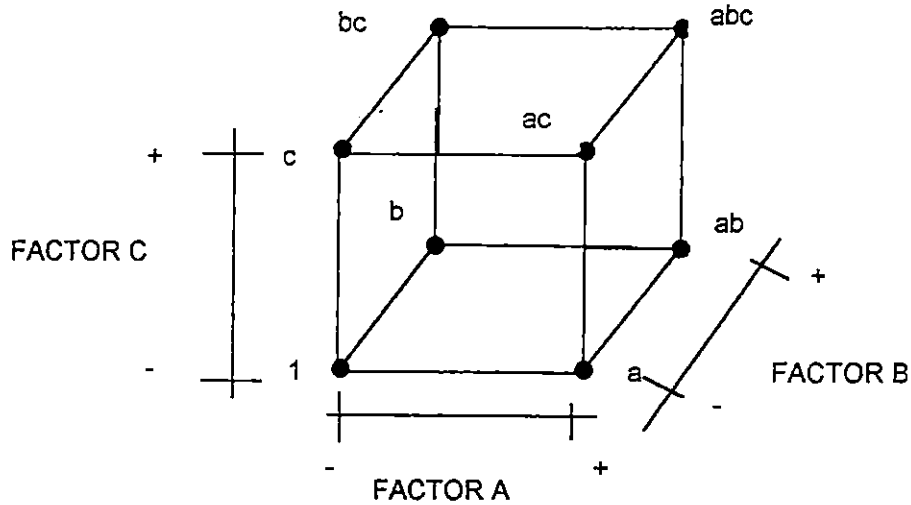


Figura 6.3. Diagrama de combinaciones en un diseño trifactorial 3^2

Las diferencia promedio de las 7 combinaciones, se determina de la siguiente manera:

Factor A:

$$A = \frac{1}{4n} [a + ab + ac + abc - b - c - bc - (1)] \quad (7.98)$$

Factor B:

$$B = \frac{1}{4n} [b + ab + bc + abc - a - c - ac - (1)] \quad (7.99)$$

Factor C:

$$C = \frac{1}{4n} [c + ac + bc + abc - a - b - ab - (1)] \quad (7.100)$$

Interacción AB:

$$AB(c \text{ baja}) = \frac{1}{2n} [ab - b] - \frac{1}{2b} [a - (1)]$$

$$AB(c \text{ alta}) = \frac{1}{2n} [(abc - bc) - \frac{1}{2n} (ac - (1))]$$

El promedio de estos dos componentes es la interacción AB

$$AB = \frac{1}{4n} [ab + (1) + abc + c - b - a - bc - ac] \quad (7.101)$$

Interacción AC. De forma similar

$$AC = \frac{1}{4n} [ac + (1) + abc + b - a - c - ab - bc] \quad (7.102)$$

Interacción BC

$$BC = \frac{1}{4n} [bc + (1) + abc + a - b - c - ab - ac] \quad (7.103)$$

Interacción ABC

$$ABC = \frac{1}{4n} [abc - bc - ac + c - ab + b + a - (1)] \quad (7.104)$$

Las cantidades entre corchetes se les denomina "Contrastes", en las ocho combinaciones. Estos contrastes pueden obtenerse a partir de una tabla de signos más y menos para el diseño 2^3 , mostrado en la tabla 6.15 y 6.16 para un diseño 2^3

6.4.4.2.1 *Recolección de Datos*

A través de la tabla 6.15 se muestra la forma de recolección de los datos, además muestra los signos algebraicos para calcular los efectos en con diseño 2^3 .

Tabla 6.15. Recolección de datos para un diseño

Replicas			Factor			Combinaciones Tratamientos	
1	2	...	A	B	C		
y_{11}	y_{21}	...	y_{n1}	-	-	-	(1)
y_{1a}	y_{2a}	...	y_{na}	+	-	-	a
.	.	.	.	-	+	-	b
.	.	.	.	+	+	-	ab
.	.	.	.	-	-	+	c
.	.	.	.	+	-	+	ac
.	.	.	.	-	+	+	bc
y_{1abc}	y_{2abc}	...	y_{nabc}	+	+	+	abc

Tabla 6.16. Signos para efectos en el diseño 2^3

Combinación de tratamientos	Efecto Factorial							
	I	A	B	AB	C	AC	BC	ABC
(1)	+	-	-	+	-	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	+	-	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	-	+	+	-	-
bc	+	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+	+

6.4.4.2.2 Análisis Estadístico.

Para los cálculos de cualquier efecto, se determina de la siguiente manera:

$$SS = \frac{(\text{Contraste})^2}{n2^k}$$

6.4.4.2.3 Análisis de Varianza

En la tabla 6.17. se muestra en forma general un resumen para el análisis de varianza:

Tabla 6.17. Análisis de varianza para un diseño 2^3

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Medía Cuadrática	F_0
A	$SS_A = \frac{(\text{Contraste } A)^2}{n2^3}$	1	$MS_A = SS_A$	$\frac{MS_A}{MS_E}$
B	$SS_B = \frac{(\text{Contraste } B)^2}{n2^3}$	1	$MS_B = SS_B$	$\frac{MS_B}{MS_E}$
C	$SS_C = \frac{(\text{Contraste } C)^2}{n2^3}$	1	$MS_C = SS_C$	$\frac{MS_C}{MS_E}$
AB	$SS_{AB} = \frac{(\text{Contraste } AB)^2}{n2^3}$	1	$MS_{AB} = SS_{AB}$	$\frac{MS_{AB}}{MS_E}$
AC	$SS_{AC} = \frac{(\text{Contraste } AC)^2}{n2^3}$	1	$MS_{AC} = SS_{AC}$	$\frac{MS_{AC}}{MS_E}$
BC	$SS_{BC} = \frac{(\text{Contraste } BC)^2}{n2^3}$	1	$MS_{BC} = SS_{BC}$	$\frac{MS_{BC}}{MS_E}$
ABC	$SS_{ABC} = \frac{(\text{Contraste } ABC)^2}{n2^3}$	1	$MS_{ABC} = SS_{ABC}$	$\frac{MS_{ABC}}{MS_E}$
ERROR	SS_E	$8(n-1)$	$MS_E = \frac{SS_E}{8(n-1)}$	
TOTAL	SS_T	$8n - 1$		

Donde:
$$SS_T = \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^2 \sum_{l=1}^n y_{ijkl}^2 - \frac{y_{\dots}^2}{2^3 n}$$

$$SS_E = SS_T - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} - SS_{ABC}$$

n = Número de Réplicas

6.4.4.2.4 Idoneidad del modelo

Como el caso del diseño factorial 2^2 , la comprobación de la idoneidad del modelo 2^3 , tiene los mismo procedimiento.

6.4.5 CONFUSIÓN EN EL DISEÑO 2^k .

En las áreas de ingeniería Química e Ingeniería de Alimentos, existen muchos problemas en los que es imposible realizar una réplica completa de un diseño factorial usando un solo bloque, para ello la característica es que el bloque sea homogéneo. La técnica de confusión es una técnica de diseño utilizado para acomodar un experimento factorial en bloque, donde el tamaño y bloque es menor que el número de combinaciones de tratamientos en una réplica.

6.4.5.1 Construcción de bloques.

Existen varios métodos para construir los bloques en un diseño confundido 2^k . Para el caso se ilustrará el método que se basa en la combinación lineal.

$$L = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k$$

Donde:

x_i = Es el nivel del i ésimo factor

α_i = Es el exponente que aparece en el i ésimo factor del efecto que será confundido.

Para el sistema 2^k , se tiene que: $\alpha_i = 0$, o bien 1, y que $x_i = 0$ (nivel inferior) ó 1 (nivel superior).

Para ilustrar este método, considere 2^3 , con el efecto ABC confundido con los bloques (tal como se muestra en la figura 6.4). En este caso, x_1 corresponde a A, x_2 a B, x_3 a C, y $\alpha_1 = \alpha_2 = \alpha_3 = 1$. El contraste de definición que corresponde a ABC es:

$$L = x_1 + x_2 + x_3$$

La combinación (1)

$$L = 1(0)+1(0)+1(0) = 0 = 0 \pmod{2}$$

De forma similar para los otros tratamientos combinados.

$$a: L=1(1)+1(0)+1(0)=1=1 \pmod{2}$$

$$b: L=1(0)+1(1)+1(0)=1=1 \pmod{2}$$

$$ab:L=1(1)+1(1)+1(0)=2=0 \pmod{2}$$

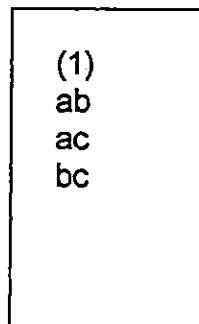
$$c: L=1(0)+1(1)+1(1)=1=1 \pmod{2}$$

$$ac:L=1(1)+1(0)+1(1)=2=0 \pmod{2}$$

$$bc:L=1(0)+1(1)+1(1)=2=0 \pmod{2}$$

$$abc:L=1(1)+1(1)+1(1)=3=1 \pmod{2}$$

BLOQUE 1



BLOQUE 2

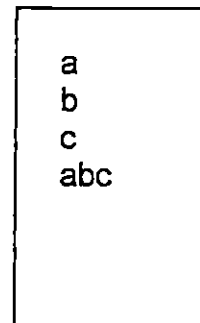
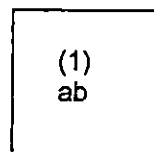


Figura 6.4. Construcción de bloques para un modelo trifactorial

De forma similar para diseños 2^2 y 2^4 como se muestra en las figuras 6.5 y 6.6

BLOQUE 1



BLOQUE 2

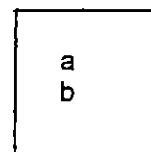


Figura 6.5. Construcción de bloques para un diseño confundido 2^2

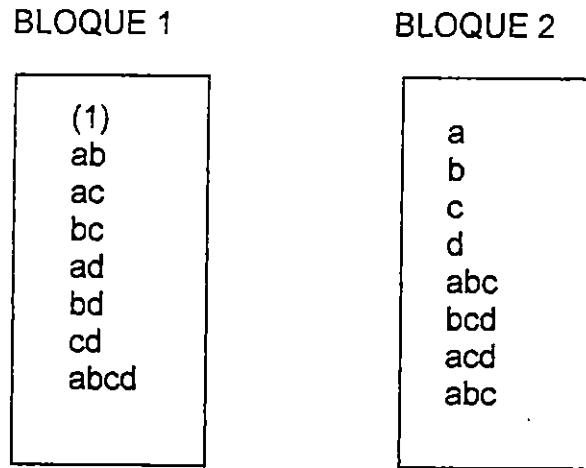


Figura 6.6. Construcción de bloques para un diseño 2^4

Para construir 4 bloques, se sigue el mismo procedimiento, con la variante que se seleccionan dos efectos que un a confundirse. Por ejemplo, para un diseño confundido 2^4 sería:

$$L_1 = x_1 + x_3 \quad (\text{confundido AC})$$

$$L_2 = x_2 + x_4 \quad (\text{confundido BD})$$

Al verificar los cuatros bloques se muestran similares a la Figura 6.7:

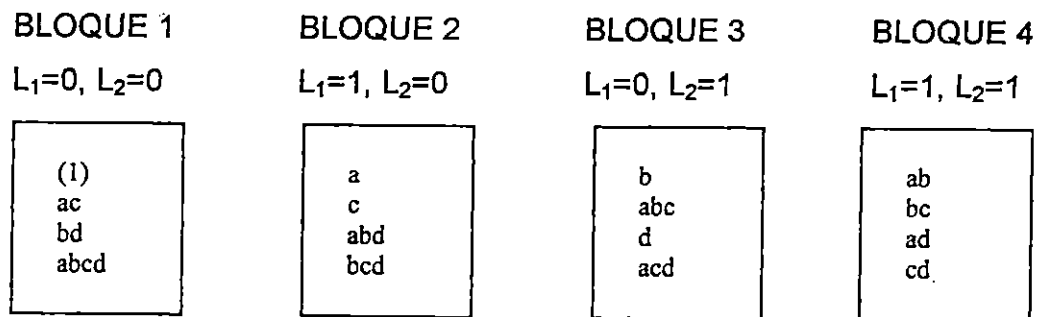


Figura 6.7. Construcción de cuatros bloques confundido 2^3

Para un diseño confundido 2^5 sería:

$$L_1 = x_1 + x_4 + x_5 \quad (\text{confundido ADE})$$

$$L_2 = x_2 + x_3 + x_5 \quad (\text{confundido BCE})$$

$L_1=0, L_2=0$ para (1), ad, bc, abcd, abe, ace, cde, bde

$L_1=1, L_2=0$ para a, d, abc, bcd, be, abde, ce, acde

$L_1=0, L_2=1$ para b, abd, c, acd, ae, de, abce, bcde

$L_1=1, L_2=1$ para e, ade, bce, abcde, ab, bd, ac, cd

El análisis de varianza y suma de cuadrados se determinará de la misma forma que se realizó previamente.

6.4.6 RÉPLICA FRACCIONAL DEL DISEÑO 2^K

En estudios complejos, estudios en las áreas de Ingeniería Química e Ingeniería de Alimento (Línea de producción, procesos químicos como se encuentran en la industria petrolera, plástico o metales , o procesos físico químico como los encontrados en las industrias electroquímico como los encontrados que requieran hacer usos de la operación unitarias, etc). Se encuentran a menudo con una serie enorme de variables interrelacionadas.

Cuando el número de factores en un diseño 2^K aumenta el número de ejecuciones que se requieren aumento rápidamente. Por ejemplo un diseño 2^5 , requiere 32 ejecuciones. En este diseño, solo cinco grados de libertad corresponden a los efectos principales y 10 grados de libertad corresponden a interacciones de dos factores.

6.4.6.1 Fracción Media del Diseño 2^K

Una fracción media del diseño 2^K contiene 2^{K-1} ejecuciones y suele llamarse diseño factorial fraccionario 2^{K-1} .

Como ejemplo, considere el diseño 2^{3-1} ; esto es, una fracción media del diseño 2^3 por que tiene $2^{3-1} = 4$ combinaciones de tratamientos.

En la tabla 6.19 aparecen los signos positivos y negativos del diseño 2^3 . Nótese que el diseño 2^{3-1} se forma al seleccionar solo las combinaciones de tratamientos que producen un signo positivo sobre la columna ABC por eso ABC se denomina generador de esta fracción en particular. Además, la columna identidad I siempre es positiva, por lo cual:

$$I = ABC$$

Se denomina relación definitiva del diseño. En general todas las columnas que son iguales a la columna identidad I.

Tabla 6.18. Signos positivos para el diseño factorial 2^3

Combinación de tratamientos	Efecto Factorial							
	I	A	B	AB	C	AC	BC	ABC
+	+	-	-	-	-	+	+	
+	-	+	-	-	+	-	+	
+	-	-	+	+	-	-	+	
abc	+	+	+	+	+	+	+	+
ab	+	+	+	+	-	-	-	-
ac	+	+	-	-	+	+	-	-
bc	+	-	+	-	+	-	+	-
(1)	+	-	-	+	-	+	+	-

Las combinaciones de tratamientos del diseño 2^{3-1} , producen tres grados de libertad que se pueden usarse para estimar los efectos principales. En la tabla 6.19, se muestra que las combinaciones lineales de las observaciones que se utilizan para estimar los efectos principales A, B y C, son:

$$L_A = 1/2(a-b-c+abc)$$

$$L_B = 1/2(-a+b-c+abc)$$

$$L_c = 1/2(-a-b+c+abc)$$

La estimación de dos factores son:

$$L_{BC} = 1/2(a-b-c+abc)$$

$$L_{AC} = 1/2(-a+b-c+abc)$$

$$L_{AB} = 1/2(-a-b+c+abc)$$

Por lo tanto, $L_A = L_{BC}$, $L_B = L_{AC}$ y $L_c = L_{AB}$. Dos o más efectos que tienen esta propiedad se conocen como "alias". Por ejemplo, A y BC, B AC, y C y AB son alias. Esto es : $L_A \rightarrow A+BC$, $L_B \rightarrow B+CA$ y $L_c \rightarrow C+BA$

Usando la fracción al tema, las combinaciones lineales de las observaciones:

$$L_A' \rightarrow A+BC$$

$$L_B' \rightarrow B+CA$$

$$L_c' \rightarrow C+BA$$

Por lo tanto, usando los tres pares de combinaciones lineales se obtienen la siguiente:

i	De $\frac{1}{2}(I_i+I_j)$	De $\frac{1}{2}(I_i-I_j)$
A	A	BC
B	B	AC
C	C	AB

La forma de encontrar los seudónimos es de la siguiente manera:

$$A=A.I=A.ABC=A^2BC=BC$$

Puesto que: $A.I=A$ y $A^2=I$. Los seudónimos de B y C son:

$$B=B.ABC=AB^2C=AC$$

Y

$$C=C.ABC=ABC^2=AB$$

6.4.6.2 Resolución del Diseño

El concepto de resolución del diseño es una manera útil para catalogar diseños factoriales fraccionarios de acuerdo con los patrones de seudónimos que ellos producen. Los diseños de resolución III, IV y V, son particularmente importantes. Las definiciones de estos términos y un ejemplo de cada uno de ellos se encuentran a continuación.

Diseños de resolución III. Estos son diseños en los cuales ningún efecto principal se hace seudónimo con cualquier otro efecto principal, pero los efectos principales se hacen seudónimos con interacciones de dos factores y las interacciones pueden hacerse seudónimos entre sí. El diseño 2^{3-1} con $I = ABC$, es de resolución III. Se suele emplear con subíndice de número romano para indicar la resolución del diseño, de tal manera esta fracción media es un diseño 2^{3-1}_{III} .

2. Diseño de resolución IV. Estos son diseños en los cuales ningún efecto principal es seudónimo con cualquier otro efecto principal o interacción de dos factores, pero las interacciones de dos factores se hacen seudónimo entre sí. El diseño 2^{4-1} , en $I = ABCD$, es de resolución IV (2^{4-1}_{IV}).
3. Diseño de resolución V. Estos son diseños en los cuales ningún efecto principal o interacción de dos factores se hacen seudónimos con cualquier otro efecto principal o interacción de dos factores, pero las

interacciones de dos factores se hacen seudónimo con interacciones de tres factores. El diseño 2^{5-1} en I =ABCDE es de resolución V (2^{5-1}_V).

Los diseños de resolución III y IV son particularmente útiles en los experimentos de eliminación de factores. El diseño de resolución IV brindada muy buena información acerca de los efectos principales y proporcionará cierta información acerca de las interacciones de dos factores.

6.4.6.3 Fracción Menores. El Factorial Fraccionario 2^{K-P} .

Se encuentra a menudo la necesidad de reducir el número de ejecuciones que se requieren para un experimento, entonces los diseños con fracciones más pequeñas brindarán información casi tan útil incluso con mayor economía (tiempo y recursos). Los diseños generalmente se llaman diseños factoriales fraccionario 2^{K-P} . Estos pueden ser :

- a. Fracción $\frac{1}{4}$ denominado diseño factorial fraccionario 2^{K-P}
- b. Fracción $\frac{1}{8}$ denominado diseño factorial fraccionario 2^{K-P}
- c. Fracción $\frac{1}{16}$ denominado diseño fraccionario 2^{K-P} .

En el presente trabajo se describe el modelo factorial fraccionario 2^{K-2} . Este diseño se puede construir escribiendo primero las combinaciones de tratamiento asociados con el factorial completo con $k-2$ factores, después se asocian las dos columnas adicionales con las interacciones, elegidas apropiadamente, que incluyen los primeros $k-2$ factores. La cuarta parte de un diseño 2^k , tiene dos generadores o generatrices, es decir P y Q; entonces las relaciones $I = P$ y $I = Q$, se conocen como relaciones generadoras del diseño. Además deben pertenecer a la misma familia. Cuando P y Q son ambas positivas se conoce como fracción principal.

En otras palabras, la relación que define el diseño es : $I = P = Q = PQ$.

Los seudónimos para los factores y de interacciones de dos o más factores se determinan de la misma forma que realizó previamente en el diseño con medio fraccional $2k-1$.

Nota: Para realizar el análisis de varianza y las pruebas de idoneidad del modelo de los diseños de réplicas fraccionaria es de la misma manera de cómo se vio previamente

6.5 EJEMPLOS DEL CAPÍTULO

Ejemplo 6.1

Se está investigando la falta de consistencia en las técnicas de medición de la cantidad de estaño depositado electrolíticamente sobre una larga lamina de acero. Se asignaron a cuatros laboratorios al azar. Cuyos resultados del contenido de estaño electrodepositado se muestra en la Tabla 6.19.

Tabla 6.19. Contenido de estaño electrodepositado (gramos).

Laboratorio	Contenido de estaño											
	Repeticiones											
	1	2	3	4	5	6	7	8	9	10	11	12
A	0.25	0.27	0.22	0.30	0.27	0.28	0.32	0.24	0.31	0.26	0.21	0.28
B	0.18	0.28	0.21	0.23	0.25	0.20	0.27	0.19	0.24	0.22	0.29	0.16
C	0.19	0.25	0.27	0.24	0.18	0.26	0.28	0.24	0.25	0.20	0.21	0.19
D	0.23	0.30	0.28	0.28	0.24	0.34	0.20	0.18	0.24	0.28	0.22	0.21

Asumiendo que es un diseño unifactorial completamente al azar, determinar:

- Si los laboratorios están obteniendo resultados consistentes.
- Los parámetros del modelo
- Idoneidad del modelo

Solución:

El modelo lineal es:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

- a. Para determinar si los laboratorios en sus análisis son consistentes, se seguirá los siguientes pasos:

1. Hipótesis nula: $\tau_A = \tau_B = \tau_C = \tau_D$
Hipótesis alternativa: $\tau_i \neq 0$ para al menos una i
2. Nivel de significación: $\alpha = 0.05$
3. Criterio: Rechazo de la hipótesis nula si $F_0 > F_{0.05, 3, 44}$, es decir si $F_0 > 2.82$
Con:

$a = 4$ tratamientos

$n = 12$ repeticiones

$N = an = 48$ observaciones.

4. Análisis de varianza: Para realizar el análisis de varianza se procederá de la siguiente forma: Primero ordenando los datos por tratamiento para calcular $y_{..}$, Y_A , Y_B , Y_C , y Y_D .

	Laboratorio A	Laboratorio B	Laboratorio C	Laboratorio D
	0.25	0.18	0.19	0.23
	0.27	0.28	0.25	0.30
	0.22	0.21	0.27	0.28
	0.30	0.23	0.24	0.28
	0.27	0.25	0.18	0.24
	0.28	0.20	0.26	0.34
	0.32	0.27	0.28	0.20
	0.24	0.19	0.24	0.18
	0.31	0.24	0.25	0.24
	0.26	0.22	0.20	0.28
	0.21	0.29	0.21	0.22
	0.28	0.16	0.19	0.21
$y_{i.}$	3.21	2.72	2.76	3.0
$\bar{y}_{i.}$	0.2675	0.2267	0.23	0.25

Ahora $y_{..} = 11.69$ y $\bar{y}_{..} = 0.2435$

Calculando la suma de cuadrados de la siguiente manera:

$$SS_T = \sum_{i=1}^4 \sum_{j=1}^{12} y_{ij}^2 - \frac{y_{..}^2}{48} = 2.9279 - 2.847 = 0.080898$$

$$SS_{TRAT} = \sum_{i=1}^4 \frac{y_{i.}^2}{12} - \frac{y_{..}^2}{48} = 0.01301$$

$$SS_E = SS_T - SS_{TRAT} = 0.0679$$

Con los cálculos previos realizados se hace una tabla resumen del análisis de varianza (Tabla 6.20).

Tabla 6.20. Análisis de varianza para los datos de contenido de estaño electrodepositado reportado por los laboratorios.

Fuente de variación	Suma de cuadrados	Grados de Libertad	Media cuadrática	F ₀	F _α
Laboratorios	0.01301	3	0.0043	2.87	2.82
Error	0.0679	44	0.0015		
Total	0.0809	47			

Decisión: La hipótesis nula se puede rechazar dado que F calculado es ligeramente mayor que F de tabla y se concluye que los laboratorios no están obteniendo resultados consistentes.

- b. Para calcular los parámetros del modelo se procede de la siguiente manera:

$$\bar{\mu} = \bar{y}_{..} = \frac{11.69}{48} = 0.2435 \text{ gramos}$$

$$\tau_A = \bar{y}_A - \bar{y}_{..} = 0.2675 - 0.2435 = 0.02396 \text{ gramos}$$

$$\tau_B = 0.2267 - 0.2435 = -0.0168 \text{ gramos}$$

$$\tau_C = 0.23 - 0.2435 = -0.0135 \text{ gramos}$$

$$\tau_D = 0.25 - 0.2435 = 0.0065 \text{ gramos}$$

El modelo es:

$$\bar{y}_j = 0.2435 + \tau_j$$

c. Para demostrar la idoneidad del modelo en la Tabla 6.21. se muestra el resumen de los residuos dando como resultados las siguientes graficas:

1. Grafica de probabilístico normal contra los residuos ordenados en forma ascendente (Figura 6.8.).
2. Grafica de residuos contra los tratamientos (Figura 6.9).
3. Grafica de residuos contra los valores \hat{y}_i (Figura 6.10)

Tabla 6.21. Tabla resumen de los residuos por tratamiento.

Residuo A	Residuos B	Residuos C	Residuos D
-0.01746	-0.0467	-0.04	-0.02
0.00254	0.0533	0.02	0.05
-0.04746	-0.0167	0.04	0.03
0.03254	0.0033	0.01	0.03
0.00254	0.0233	-0.05	-0.01
0.01254	-0.0267	0.03	0.09
0.05254	0.0433	0.05	-0.05
-0.02746	-0.0367	0.01	-0.07
0.04254	0.0133	0.02	-0.01
-0.00746	-0.0067	-0.03	0.03
-0.05746	0.0633	-0.02	-0.03
0.01254	-0.0667	-0.04	-0.04

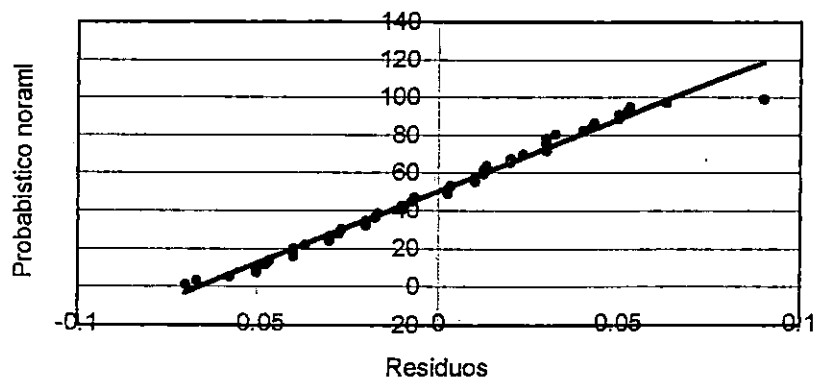


Figura 6.8. Grafica de probabilístico normal contra los valores de residuos

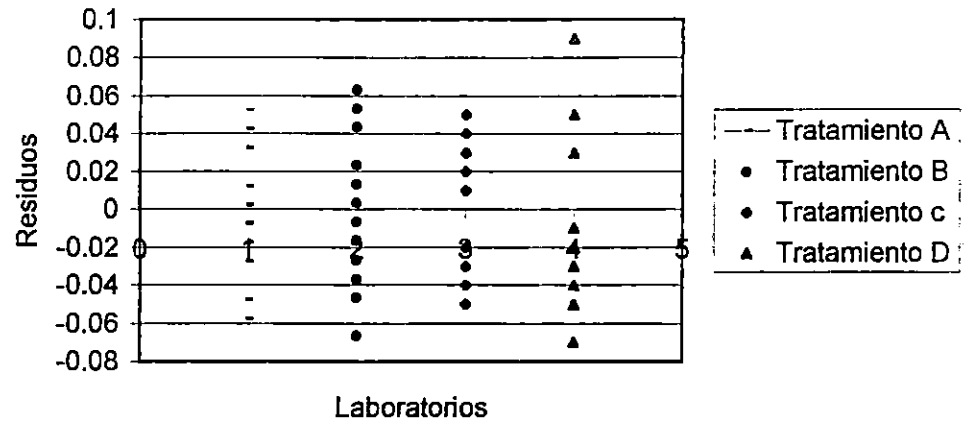


Figura 6.9 . Grafica de residuos contra tratamientos

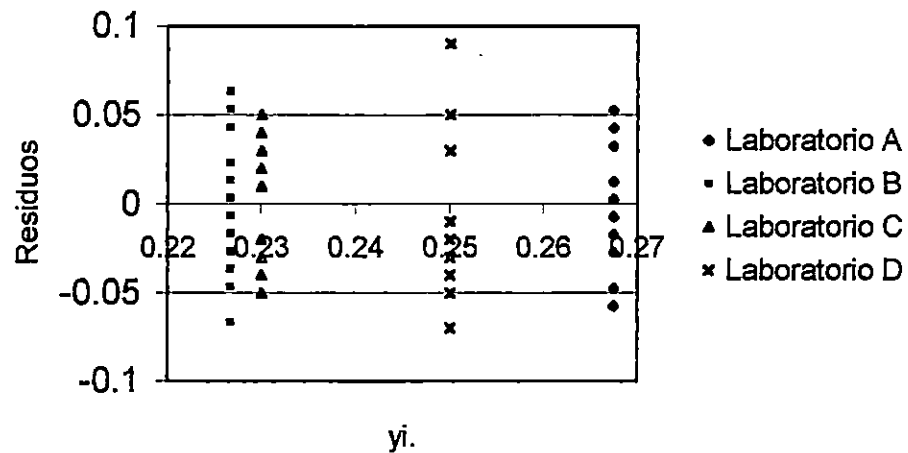


Figura 6.10. Grafica de residuos contra \bar{y}_i

Según el examen de idoneidad del modelo describe significativamente la relación existente entre los laboratorios y la cantidad de estaño en las láminas.

Ejemplo 6.2

Una fabrica de hilados tienen un gran numero de telares. Se supone que cada uno de los telares proporciona la misma salida de tela por minuto. Para investigar esta suposición, se eligen cinco telares al azar y su salida se medien en diferentes tiempos, obteniendo los siguientes resultados que presentan en la Tabla 6.22:

Tabla 6.22. Resultados obtenidos de los telares (libras de telas por minutos)

Telar	Salida (Lb/min)				
1	4.0	4.1	4.2	4.0	4.1
2	3.9	3.8	3.9	4.0	4.0
3	4.1	4.2	4.1	4.0	3.9
4	3.6	3.8	4.0	3.9	3.7
5	3.8	3.6	3.9	3.8	4.0

Determinar:

- Si los telares proporcionan la misma salida de tela por minuto
- Los componentes de varianza
- La idoneidad del modelo.

Solución:

- El modelo lineal es:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

- Hipótesis nula: $H_0: \sigma_r^2 = 0$

Hipótesis alternativa: $H_1: \sigma_r^2 > 0$

- Nivel de significación: $\alpha=0.05$
- Criterio: se rechaza la hipótesis nula cuando: $F_0 > F_{0.05,4,20}$ (2.87)

Donde:

$$a=4$$

$$n=4$$

$$N=25$$

- Elaborar una tabla de observaciones totales y promedio

Telar	Totales	Promedio
1	20.4	4.08
2	19.6	3.92
3	20.3	4.06
4	19	3.8
5	19.1	3.82
Totales	98.4	3.936

- Cálculos previos:

$$SS_T = \sum_{i=1}^5 \sum_{j=1}^5 y_{ij}^2 - \frac{y_{..}^2}{N} = 387.94 - 387.3024 = 0.6376$$

$$SS_{\text{Telares}} = \frac{\sum_{i=1}^5 y_{i.}^2}{5} - \frac{y_{..}^2}{25} = 387.644 - 387.3024 = 0.3416$$

$$SS_E = SS_T - SS_{\text{Telares}} = 0.6376 - 0.3416 = 0.296$$

7. Análisis de varianza del ejemplo 6.2 se muestra en la Tabla 6.23 a continuación.

Tabla 6.23. Análisis de varianza para el ejemplo 6.2

Fuente de Variación	Suma de cuadrados	Grados de libertad	Media cuadrática	Fo	F _{Tabla}
Telares	0.3416	4	0.0854	5.77027	2.87
Error	0.296	20	0.0148		
Total	0.6376	24			

8. Conclusión: Como Fo es mayor que F de tabla la hipótesis nula se rechaza, concluyendo que los telares no proporcionan la misma salida.

b. Los componentes de varianza son:

$$MS_E = \sigma^2 = 0.0148$$

y

$$\sigma_r^2 = \frac{MS_{\text{Telares}} - MS_E}{5} = 0.01412$$

Cuya $V(y_{ij}) = 0.0148 + 0.01412 = 0.2892$ y una desviación estándar estimada de $\sigma_y = 0.17$ lbs por minutos

c. Para la idoneidad del modelo es necesario estimar los parámetros del modelo lineal para luego determinar los valores predichos (ajustados) y sus respectivos residuos, como se muestra a continuación:

$$\mu = \bar{y}_{..} = 3.936$$

para los telares son:

$$t_1 = \bar{y}_1 - 3.936 = 4.08 - 3.936 = 0.117,$$

$$t_2 = 3.92 - 3.936 = -0.043,$$

$$t_3 = 4.06 - 3.936 = 0.097,$$

$$t_4 = 3.8 - 3.963 = -0.163 \text{ y}$$

$$t_5 = 3.82 - 3.963 = -0.143$$

siendo el modelo siguiente:

$$\hat{y}_{ij} = 3.963 + t_i \quad \text{para } i = 1, 2, \dots, 5$$

Este modelo se necesita para estimar los valores predichos por el modelo y los residuos, cuyas estimaciones se muestran en las Tablas 6.24 y 6.25. y posteriormente graficarlos, tal como se muestra en las figuras 6.11, 6.12 y 6.13.

Tabla 6.24. Tabla resumen de valores predichos por el modelo

Telar	Valores predichos
1	4.053
2	3.893
3	4.033
4	3.773
5	3.793

Tabla 6.25. Tabla resumen de los valores de los residuos

Valores de residuos				
Telar 1	Telar 2	Telar 3	Telar 4	Telar 5
-0.053	0.007	0.067	-0.173	0.007
0.047	-0.093	0.167	0.027	-0.193
0.147	0.007	0.067	0.227	0.107
-0.053	0.107	-0.033	0.127	0.007
0.047	0.107	-0.133	-0.073	0.207

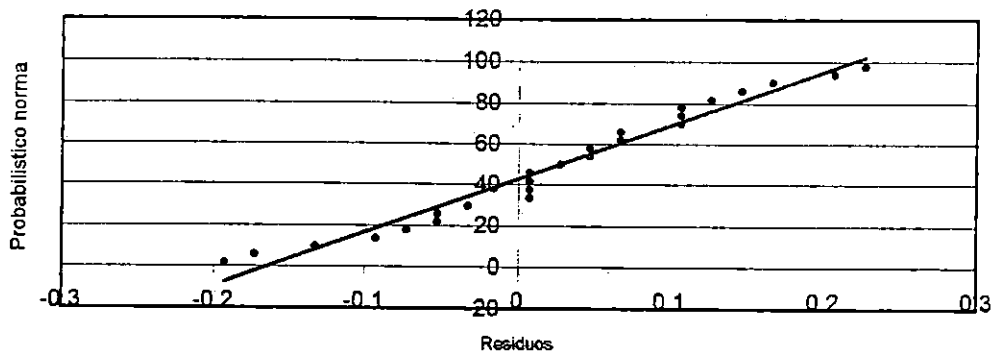


Figura 6.11. Grafica de probabilístico normal vs los residuos

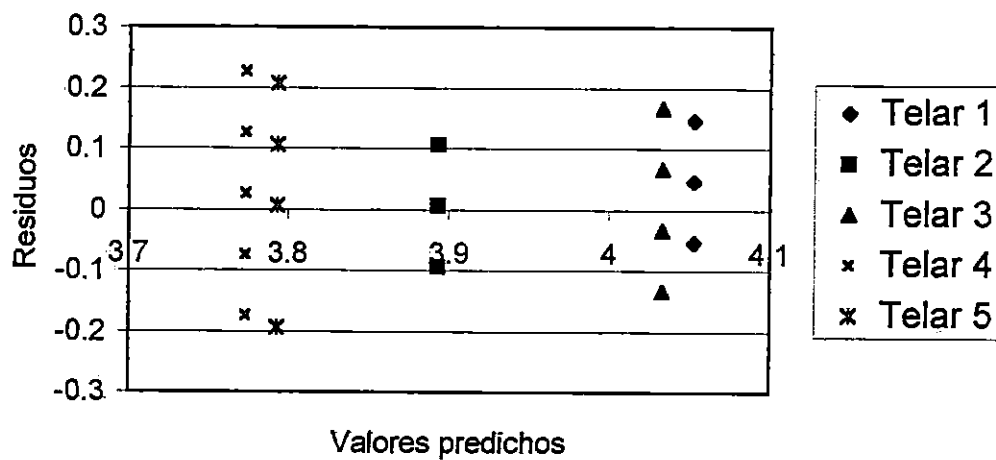
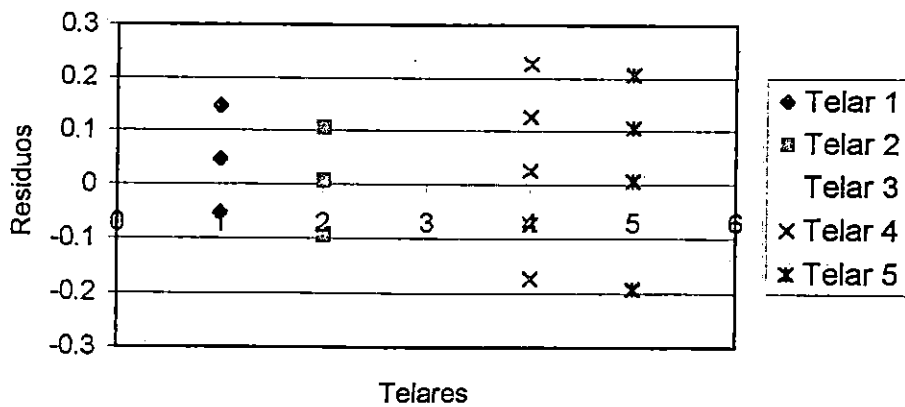


Figura 6.12. Grafica de residuos vs. Valores predichos



El examen del modelo muestra que los residuos se comportan con una distribución normal por lo tanto el modelo es adecuado para describir la variabilidad de los datos.

Ejemplo 6.3

Se diseñó un experimento para estudiar el desempeño de cuatro detergentes diferentes para limpiar inyectores de combustibles. Las siguientes lecturas de limpiezas se obtuvieron con equipos especialmente diseñados para 12 tanques de gas distribuidos en tres diferentes modelos de motores, presentados en la Tabla 6.26:

Determinar:

- Si existe diferencia en el desempeño de los cuatro detergentes y para los tres motores en forma significativa.
- Los estimadores μ , τ_i y β_j
- La idoneidad del modelo

Tabla 6.26. Lecturas obtenidas de tres tanques de autos diferentes.

Detergentes	Tanques		
	1	2	3
A	45	43	51
B	47	46	52
C	48	50	55
D	42	37	49

Solución: Los efectos que se desea demostrar se trata de un modelo unifactorial con bloques aleatorios (efectos fijos).

- Para determinar el desempeño de los detergentes y los motores se seguirá de la siguiente manera:
 - El modelo lineal es:

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

- Las hipótesis nulas son:

Para los detergentes: $H_0 : \tau_i = 0$

Para los motores: $H_0 : \beta_j = 0$

Las hipótesis alternativas son:

Para los detergentes: $H_1 : \tau_i \neq 0$ por lo menos para una i

Para los motores: $H_1 : \beta_j \neq 0$ por lo menos para una j

3. Nivel de significación: $\alpha=0.05$

4. Criterio de rechazo:

Para los tratamientos (detergentes)

Donde:

$$a=4$$

$$b=3$$

$$g.l \text{ de los tratamientos} = 3 \text{ (} v_1 \text{)}$$

$$g.l = (a-1)(b-1) = 6 \text{ (} v_2 \text{)}$$

Si $F_{\text{Tratamiento}} > F_{\alpha}$ (4.76)

Para los bloques (motores)

$$g.l = b-1 = 2 \text{ (} v_1 \text{)}$$

v_2 el mismo calculado anteriormente.

Si $F_{\text{Bloques}} > F_{\alpha}$ (5.14)

5. Cálculos previos: primeramente elaborar una tabla resumen de lecturas totales y promedio de las i y j , tal como se muestra en la Tabla 6.27.

Tabla 6.27. Resumen de lecturas subtotaes, totales y de promedios de i y j .

Detergentes	Tanques			Subtotales $\sum y_i$	Promedios \bar{y}_i
	1	2	3		
A	45	43	51	139	46.33
B	47	46	52	145	48.33
C	48	50	55	153	51
D	42	37	49	128	42.67
Subtotales $\sum y_i$	182	176	207	565	188.33
Promedios \bar{y}_j	45.5	44	51.75	141.25	47.083

6. Determinación de las sumas de cuadrados

Suma de cuadrado total, es:

$$SS_T = \sum_{i=1}^4 \sum_{j=1}^3 y_{ij}^2 - \frac{y_{..}^2}{ab} = 26867 - 26602.08333 = 264.91667$$

Para los tratamientos, es:

$$SS_D = \frac{1}{3} \sum_{i=1}^4 y_{i.}^2 - \frac{y_{..}^2}{ab} = 26713 - 26602.08333 = 110.91667$$

Para los bloques, es:

$$SS_B = \frac{\sum_{j=1}^3 y_{.j}^2}{a} - \frac{y_{..}^2}{ab} = 26737.25 - 26602.08333 = 135.1667$$

Para el error, es:

$$SS_E = SS_T - SS_D - SS_B = 264.91667 - 110.91667 - 135.1667 = 18.8333$$

7. El análisis de varianza, mostrado en la Tabla 6.28.

Tabla 6.28. Resumen del análisis de varianza del ejemplo 6.3

Fuente de variación	Suma de cuadrados	Grados de libertad	Media cuadrática	F _o	F _α
SS _D	110.91667	3	36.972	11.779	4.76
SS _B	135.1667	2	67.5834	21.531	5.14
SS _E	18.8333	6	3.1389		
SS _T	264.91667	11			

8. Decisiones: Como F_D (11.779) de los detergentes es mayor que el F de tablas (4.76) con .05 de nivel de significancia con 3 y 6 grados de libertad respectivamente; concluyendo que hay diferencia significativa en el desempeño de los cuatros detergentes. Además, para los bloques F calculado es mucho mayor que F de tablas, concluyendo que la diferencia entre los resultados obtenidos para los tres motores son significativos, esto es, que el bloque ha sido efectivo.

- b. Para la estimación de los parámetros del modelo lineal se tomará los cálculos de los promedios de los subtotales y totales presentados en la Tabla 6.27. Estos son:

Para la gran media total es:

$$\mu = \bar{y}_{..} = 47.083$$

Para los tratamientos es:

$$t_A = \mu_A - \mu = 46.33 - 47.083 = -0.753$$

$$t_B = \mu_B - \mu = 48.33 - 47.083 = 1.247$$

$$t_C = \mu_C - \mu = 51 - 47.083 = 3.927$$

$$t_D = \mu_D - \mu = 42.67 - 47.083 = -4.413$$

Para los bloques es:

$$\beta_1 = \mu_1 - \mu = 45.5 - 47.083 = -1.583$$

$$\beta_2 = \mu_2 - \mu = 44 - 47.083 = -3.083$$

$$\beta_3 = \mu_3 - \mu = 51.75 - 47.083 = 4.667$$

El modelo es:

$$\hat{y}_{ij} = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}$$

- c. Para las pruebas de idoneidad de modelo propuesto se determinará los residuos y luego se graficará con la frecuencia acumulada (probabilístico normal), con los valores predichos, con los detergentes y bloques respectivamente, a continuación se presentaran los valores predichos (haciendo uso de la ecuación que rige el modelo), residuos y las respectivas graficas:

Los valores predichos son:

Detergentes	Tanques		
	1	2	3
A	44.747	43.247	50.997
B	46.747	45.247	52.997
C	49.417	47.917	55.667
D	41.087	39.587	47.337

Y los residuos ($y_{ij} - \hat{y}_{ij}$), son:

Detergentes	Tanques		
	1	2	3
A	0.253	-0.247	0.003
B	0.253	0.753	-0.997
C	-1.417	2.083	-0.667
D	0.913	-2.587	1.663

La representación graficas se presentan en las Figuras 6.14, 6.15, 6.16 y 6.17. Como se muestra a continuación:

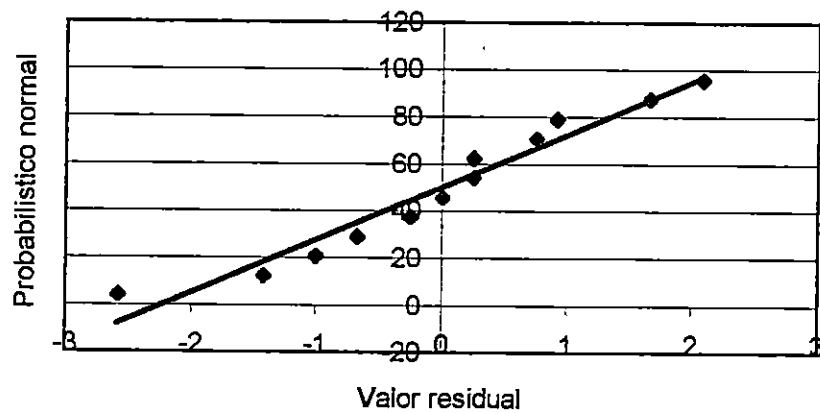


Figura 6.14. Grafica de probabilístico normal vs. los residuos

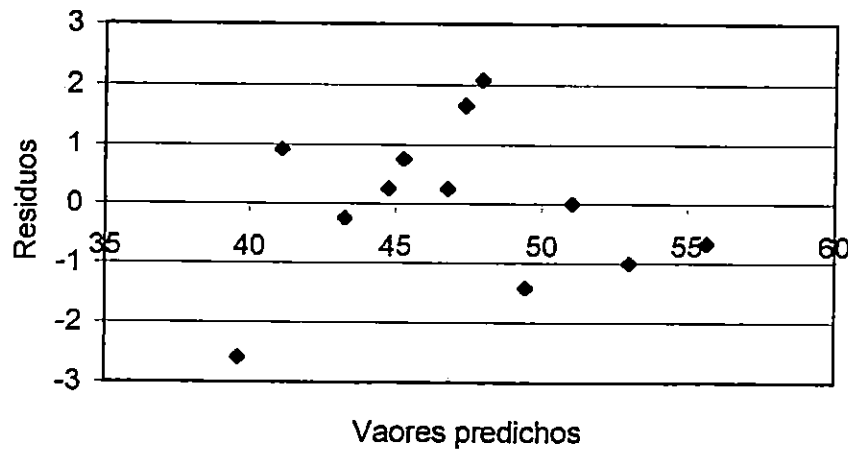


Figura 6.15. Grafica de residuos vs. valores predichos

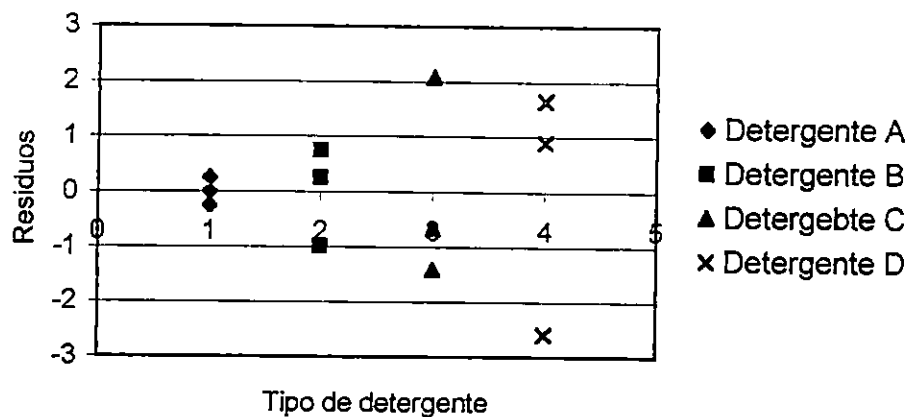


Figura 6.16. Grafica de residuos vrs. tipo de detergente .

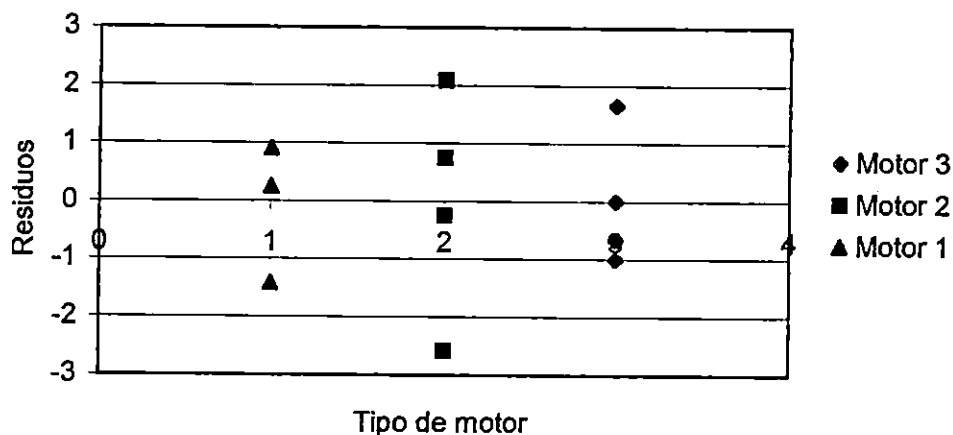


Figura 6.17. Grafica de residuos vrs. tipo de motor.

Ejemplo 6.4

Un investigador tiene tres diferentes productos de limpieza: A1, A2 y A3, y se desea seleccionar el más eficiente para limpiar una superficie metálica. La limpieza de una superficie se mide por su reflectibilidad, expresada en unidades arbitrarias como razón de reflectividad observada a la superficie de un espejo estándar. El análisis de covarianza debe usarse por que el efecto de un producto de limpieza sobre la reflectividad depende de la limpieza original. El investigador obtuvo los resultados mostrados en la tabla 6.29.

Tabla 6.29. Resultados obtenidos de reflectividad para el experimento.

A1		A2		A3	
Y	X	Y	X	Y	X
1.00	0.9	0.75	0.75	1.00	0.60
1.20	0.55	1.65	0.60	0.70	0.90
0.80	0.60	0.55	1.00	0.80	0.80
1.40	0.35	0.50	1.10	0.90	0.70

Donde:

Y es la reflectividad final

X es la reflectividad original

Determinar:

- Si el producto de limpieza tienen diferentes efectividades
- Los parámetros del modelo de regresión general
- La idoneidad del modelo.

Solución:

- Para determinar si los productos de limpiezas tienen diferentes efectividades, es necesario realizar un análisis de covarianza, de la siguientes forma:

- El modelo lineal, es:

$$Y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij}$$

- Hipótesis nula: $H_0: \tau_i = 0$

Hipótesis alternativa: $H_1: \tau_i \neq 0$ al menos una i

- Nivel de significancia: $\alpha = 0.05$
- Criterio: Rechazo de la hipótesis nula cuando $F_0 > F_{0.05}$ con 2 y 8 grados de libertad respectivamente.
- Cálculos previos:

$$S_{yy} = \sum_{i=1}^3 \sum_{j=1}^5 y_{ij}^2 - \frac{y_{..}^2}{ab} = 9.5 - \frac{10.20^2}{12} = 0.79$$

$$S_{xx} = \sum_{i=1}^3 \sum_{j=1}^5 x_{ij}^2 - \frac{X_{..}^2}{12} = 8.83 - \frac{9.5^2}{12} = 1.31$$

$$S_{xy} = \sum_{i=1}^3 \sum_{j=1}^5 x_{ij} y_{ij} - \frac{X_{..} Y_{..}}{3 \cdot 4} = 7.28 - 8.08 = -0.8$$

$$T_{yy} = \left(\sum_{i=1}^3 y_{i.}^2 \right) / 4 - \frac{Y_{..}^2}{4} = 9.17 - 8.67 = 0.50$$

$$T_{xx} = \frac{\sum_{i=1}^3 x_{i.}^2}{4} - \frac{X_{..}^2}{4} = 8.31 - 7.52 = 0.79$$

$$T_{xy} = \frac{\sum_{i=1}^3 x_{i.} y_{i.}}{4} - \frac{X_{..} Y_{..}}{12} = 7.45 - 8.08 = -0.63$$

$$E_{yy} = S_{yy} - T_{yy} = 0.79 - 0.50 = 0.29$$

$$E_{xx} = S_{xx} - T_{xx} = 1.31 - 0.79 = 0.52$$

$$E_{xy} = S_{xy} - T_{xy} = -0.80 - (-0.63) = -0.17$$

Finalmente para los ajustados se tiene:

$$SS'_E = S_{yy} - \frac{S_{xy}^2}{S_{xx}} = 0.79 - \frac{(-0.80)^2}{1.31} = 0.30$$

$$SS_E = E_{yy} - \frac{E_{xy}^2}{E_{xx}} = 0.29 - \frac{(-0.17)^2}{0.52} = 0.23$$

6. El análisis de covarianza presentado en la Tabla 6.30.

Tabla 6.30. Análisis de covarianza para los resultados del experimento de reflectividad por medio de los productos de limpieza.

Fuente de variación	Suma de cuadrados y productos				Ajustado para regresión		
	g.l	X	XY	Y	Y	g.l	Media de cuadrados
Limpiador	2	0.79	-0.63	0.50			
Error	9	0.52	-0.17	0.29	0.23	8	0.029
Total	11	1.31	-0.80	0.79	0.30	10	
Limpiador Ajustado					0.07	2	0.035

7. Decisión: Como $F=0.035/0.029=1.21$ y no excede el F de tabla (4.46), la hipótesis nula no puede rechazarse. En otras palabras, no puede concluirse que uno cualquiera de los productos de limpieza es más efectivo que los otros.

Ahora es necesario probar si: $\beta=0$ utilizando la estadística de prueba siguiente:

$$F_0 = \frac{E_{xy}^2 / E_{xx}}{MS_E} = \frac{(-0.17)^2 / 0.52}{0.029} = 1.916$$

y como $F_{0.05,1,8}=5.32$, no debe rechazarse la hipótesis. Por lo tanto no existe una relación lineal entre la limpieza final y la limpieza original o inicial. Lo cual no justifica el análisis de covarianza.

b. Los parámetros de regresión del modelo son:

$$\hat{Y}_j = \mu = \bar{y}_.. = 0.85$$

Dado que no existe efecto de los limpiadores y del coeficiente de regresión.

c. Para examinar la idoneidad del modelo se tiene:

$$\varepsilon = Y_{ij} - 0.85$$

Los valores ajustados se muestran en la tabla 6.31.

Tabla 6.31. Resumen de los valores predichos

A1		A2		A3	
Ajustado	X	Ajustado	X	Ajustado	X
0.85	0.9	0.85	0.75	0.85	0.60
0.85	0.55	0.85	0.60	0.85	0.90
0.85	0.60	0.85	1.00	0.85	0.80
0.85	0.35	0.85	1.10	0.85	0.70

Y los residuos se presentan en la tabla 6.32 a continuación:

Tabla 6.32. Resumen de los valores de los residuos.

A1		A2		A3	
Residuos	X	Residuos	X	Residuos	X
0.15	0.9	-0.10	0.75	0.15	0.60
0.35	0.55	-0.25	0.60	-0.15	0.90
-0.05	0.60	-0.30	1.00	-0.05	0.80
0.55	0.35	-0.35	1.10	0.05	0.70

Las graficas de residuos en probabilístico normal, valores predichos, valores iniciales de limpieza y tipo de limpiadores (tratamientos), se muestran en las Figuras 6.18, 6.19, 6.20 y 6.21.

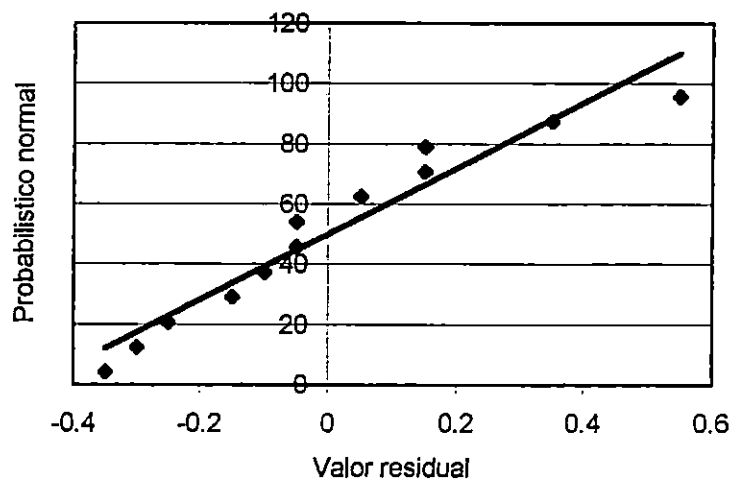


Figura 6.18. Grafica de probabilístico normal vs. residuos.

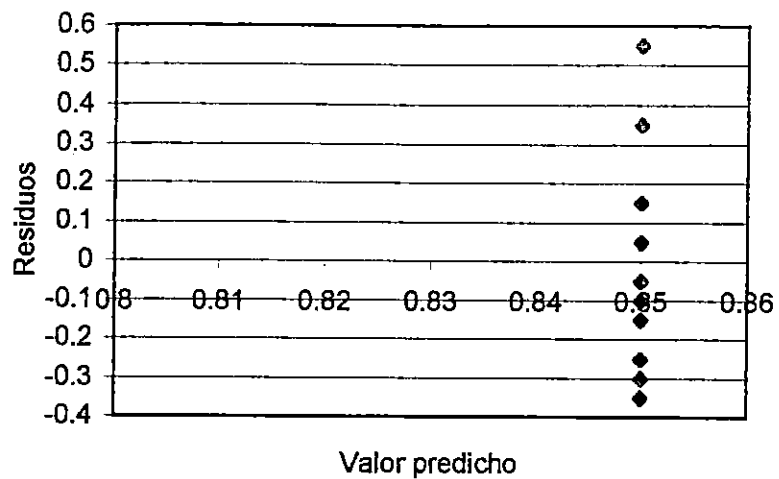


Figura 6.19. Grafica de residuos vrs. los valores predichos por el modelo

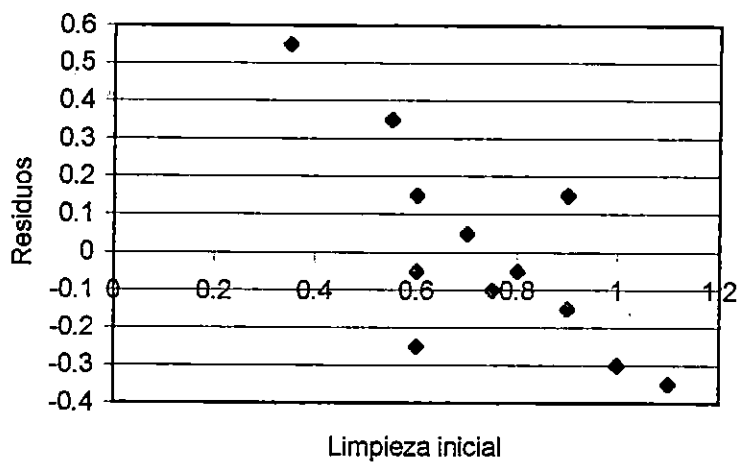


Figura 6.20. Grafica de residuos vrs. la limpieza inicial del material

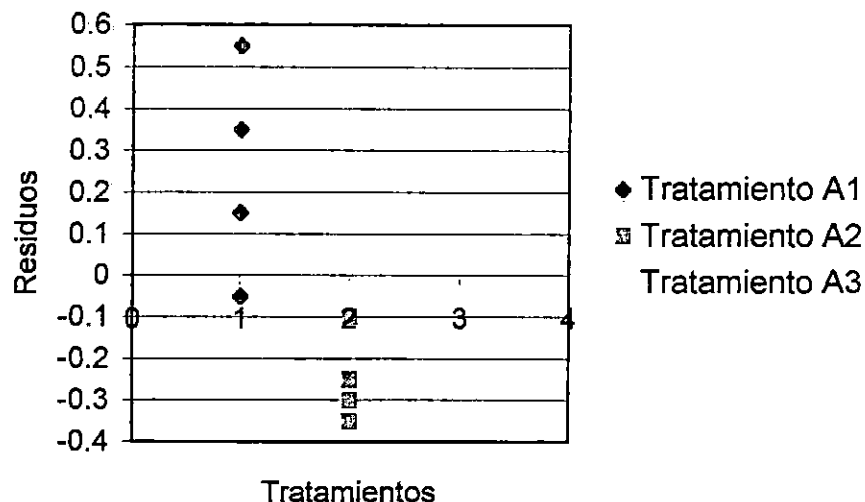


Figura 6.21. Gráfica de residuos vs. los tipos de limpiadores

Según el análisis de residuos, análisis de varianza el efecto de los tipos de limpiadores no es una variable que afecta la reflectividad.

Ejemplo 6.5

Un baño de ácido sulfúrico caliente se usa para remover óxido de la superficie de un metal antes revestido electrolíticamente y se desea determinar que factores en adición a la concentración del ácido sulfúrico podrían afectar la conductividad eléctrica del baño. Como se cree que la concentración de sal, así como la temperatura, podrían afectar la conductividad eléctrica. Para cubrir los intervalos de concentración y temperaturas normales encontrados, se decide utilizar los siguientes niveles de los tres factores:

Supóngase que dos replicas se realizan midiendo la conductividad eléctrica de los varios baños ácidos. Y los resultados se presentan en la tabla 6.33.

Factor	Nivel 1	Nivel 2	Nivel 3	Nivel 4
A concentración de ácido (%)	0	6	12	18
B concentración de sal (%)	0	10	20	
C temperatura del baño (°F)	80	100		

Tabla 6.33. Resultados del experimento de baños ácidos.

Niveles del factor	Conductividad (mhos/cm ³)		Total
	1	2	
A	B	C	
1	1	1	0.99
1	1	2	1.15
1	2	1	0.97
1	2	2	0.87
1	3	1	0.95
1	3	2	0.91
2	1	1	1.00
2	1	2	1.12
2	2	1	0.99
2	2	2	0.96
2	3	1	0.97
2	3	2	0.94
3	1	1	1.24
3	1	2	1.12
3	2	1	1.15
3	2	2	1.11
3	3	1	1.03
3	3	2	1.12
4	1	1	1.24
4	1	2	1.32
4	2	1	1.14
4	2	2	1.20
4	3	1	1.02
4	3	2	1.02

*Las condiciones experimentales son: 4x3x2=24

- a. Los efectos individuales y conjuntos de los tres factores
- b. El modelo matemático que rige el comportamiento de los datos
- c. La idoneidad del modelo propuesto

Solución:

- a. Para probar los efectos individuales ó la significancia del diseño, se hace de la siguiente manera:

1. Definir los niveles de los factores, replicas y el numero de lecturas, de la manera siguiente:

Factor A: $a=4$

Factor B: $b=3$

Factor C: $c=2$

Replicas: $n=2$

Numero de lecturas: $N=48$

2. Planteamiento de hipótesis nulas y alternativas, de la manera siguiente:

Hipótesis nula:

$$\tau_i = 0$$

$$\beta_j = 0$$

$$\gamma_k = 0$$

$$(\tau\beta)_{ij} = 0$$

$$(\tau\gamma)_{ik} = 0$$

$$(\beta\gamma)_{jk} = 0$$

$$(\tau\beta\gamma)_{ijk} = 0$$

Hipótesis alternativa:

$\tau_i \neq 0$ al menos una

$\beta_j \neq 0$ al menos una

$\gamma_k \neq 0$ al menos una



$$\begin{aligned}(\tau\beta)_{ij} &\neq 0 && \text{al menos una} \\(\tau\gamma)_{ik} &\neq 0 && \text{al menos una} \\(\beta\gamma)_{jk} &\neq 0 && \text{al menos una} \\(\tau\beta\gamma)_{ijk} &\neq 0 && \text{al menos una}\end{aligned}$$

3. Nivel de significancia: $\alpha=0.05$

4. Cálculos de suma de cuadrados:

Para los factores A, B y C. Además, totales es:

$$SS_T = 53.1005 - \frac{50.17^2}{48} = 0.6624$$

$$SS_A = \frac{11.24^2 + 12.24^2 + 13.01^2 + 13.68^2}{12} - \frac{50.17^2}{48} = 0.275$$

$$SS_B = \frac{18.21^2 + 16.37^2 + 15.59^2}{16} - \frac{50.17^2}{48} = 0.2262$$

$$SS_C = \frac{25.04^2 + 25.13^2}{24} - \frac{50.17^2}{48} = 0.0002$$

Para las interacción AB, AC y BC, se tiene:

$$SS_{AB} = \frac{4.06^2 + 3.61^2 + \dots + 4.05^2}{4} - \frac{50.17^2}{48} = 0.0289$$

$$SS_{AC} = 0.0085$$

$$SS_{BC} = 0.0042$$

Para la interacción ABC se tiene:

$$SS_{ABC} = 0.5712 - 0.275 - 0.2262 - 0.0289 - 0.0085 - 0.0042 - 0.0002 = 0.0282$$

Y para el error, se tiene:

$$SS_E = 0.5712 - 0.6624 = 0.0747$$

Para facilitar la suma de cuadrados se construye primero las siguientes tablas 6.34.

Tablas 6.34. Resumen para determinar sumas de cuadrados

		B			
		1	2	3	
A	1	4.06	3.61	3.57	11.24
	2	4.42	3.97	3.85	12.24
	3	4.73	4.16	4.12	13.01
	4	5	4.63	4.05	13.68
		18.21	4.63	15.59	50.17

		C		
		1	2	
A	1	5.61	5.63	11.24
	2	6.12	6.41	12.23
	3	6.60	6.41	13.01
	4	6.71	6.97	13.68
		25.0	25.13	50.17

		B			
		1	2	3	
C	1	8.99	8.25	7.80	25.04
	2	9.22	8.12	7.79	25.13
		18.21	16.37	15.59	50.17

5. El análisis de varianza se muestra en la tabla 6.35

Tabla 6.5. Análisis de varianza para el ejemplo 6.5

Fuente de variación	Suma de cuadrados	Grados de libertad	Media cuadrática	Fo	F α
A	0.275	3	0.0917	28.66	3.03
B	0.2262	2	0.1131	35.34	3.42
C	0.0002	1	0.0002	-----	4.28
AB	0.0289	6	0.0048	1.5	2.53
AC	0.0085	3	0.0028	-----	3.03
BC	0.0042	2	0.0021	-----	3.42
ABC	0.0282	6	0.0047	1.47	2.53
Error	0.0747	23	0.0032		
Total	0.6624	47			

Los efectos significativos de la variabilidad del proceso del baño electrolítico, es:

- a. La concentración de ácido sulfúrico
- b. La concentración de sal

En conclusión las hipótesis nulas que se rechazan son:

$$\tau_i \neq 0 \quad \text{al menos una, y}$$

$$\beta_j \neq 0 \quad \text{al menos una}$$

Con un coeficiente de determinación de:

$$R^2 = SS_{\text{MODELO}} / SS_T = (SS_A + SS_B) / SS_T = 0.7566$$

Lo cual el modelo propuesto describe el 75% de la variabilidad de los datos observados.

- b. Dado que los factores que afectan el proceso eléctrico son la concentración del ácido sulfúrico y de la sal del baño. Por lo tanto el modelo matemático lineal que regirá el comportamiento de los datos es:

$$y_{ijk} = \mu + \tau_i + \beta_j$$

Los estimadores son:

$$\mu = 1.045$$

$$\tau_1 = -0.108$$

$$\tau_2 = -0.025$$

$$\tau_3 = 0.035$$

$$\tau_4 = 0.095$$

$$\beta_1 = 0.093$$

$$\beta_2 = -0.022$$

$$\beta_3 = -0.075$$

Datos que se utilizaran para calcular los valores predichos. De la manera siguiente:

Cuando A=1 y B=1, es:

$$\hat{y}_{ijk} = 1.045 - 0.108 + 0.093 = 1.03$$

En forma análoga se calcula los valores predichos. En las Tablas 6.36 se muestra los resultados de los valores predichos:

Tabla 6.36. Resultados de los valores predichos

Niveles del factor			Conductividad (mhos/cm ³)	
A	B	C	1	2
1	1	1	1.03	1.03
1	1	2	1.03	1.03
1	2	1	0.915	0.915
1	2	2	0.915	0.915
1	3	1	0.865	0.865
1	3	2	0.865	0.865
2	1	1	1.113	1.113
2	1	2	1.113	1.113
2	2	1	0.998	0.998
2	2	2	0.998	0.998
2	3	1	0.945	0.945
2	3	2	0.945	0.945
3	1	1	1.173	1.173
3	1	2	1.173	1.173
3	2	1	1.058	1.058
3	2	2	1.058	1.058
3	3	1	1.005	1.005
3	3	2	1.005	1.005
4	1	1	1.233	1.233
4	1	2	1.233	1.233
4	2	1	1.118	1.118
4	2	2	1.118	1.118
4	3	1	1.065	1.065
4	3	2	1.065	1.065

c. Idoneidad del modelo:

Para ello, se calculará primero los residuos mediante la siguiente formula:

$$e_{ijk} = y_{ijk} - \hat{y}_{ij}$$

En las Tablas 6.37 se muestra los resultados de los valores de los residuos:

Tabla 6.37. Resumen de los residuos determinados

Niveles del factor			Conductividad (mhos/cm ³)	
A	B	C	1	2
1	1	1	-0.04	-0.1
1	1	2	0.12	-0.04
1	2	1	0.055	-0.005
1	2	2	-0.045	-0.055
1	3	1	-0.085	-0.005
1	3	2	0.045	-0.015
2	1	1	-0.113	0.057
2	1	2	0.007	0.017
2	2	1	-0.008	0.042
2	2	2	-0.038	-0.018
2	3	1	0.025	0.005
2	3	2	-0.005	0.045
3	1	1	0.067	0.047
3	1	2	-0.053	-0.023
3	2	1	0.092	-0.108
3	2	2	0.052	-0.108
3	3	1	0.025	0.005
3	3	2	0.115	-0.045
4	1	1	0.007	-0.033
4	1	2	0.087	0.007
4	2	1	0.022	-0.018
4	2	2	0.082	0.072
4	3	1	-0.045	-0.055
4	3	2	-0.045	-0.065

El análisis de gráficos de probabilístico normal, residuos vrs. valores predichos, residuos vrs. concentración de ácido sulfúrico y residuos vrs. concentración de sal (estos por ser los factores que presentan en el análisis de varianza efecto significativo). Se presentan sus gráficas en las Figuras 6.22, 6.23, 6.24 y 6.25 respectivamente a continuación:

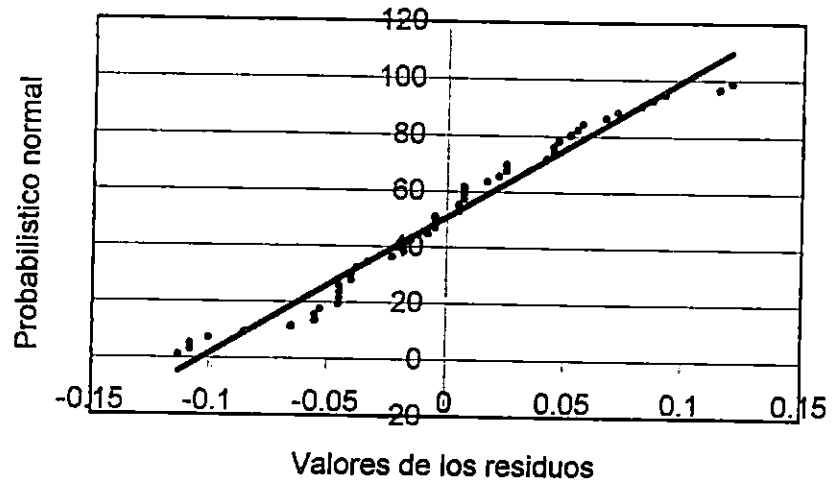


Figura 6.22. Grafico en probabilístico del ejemplo 6.5

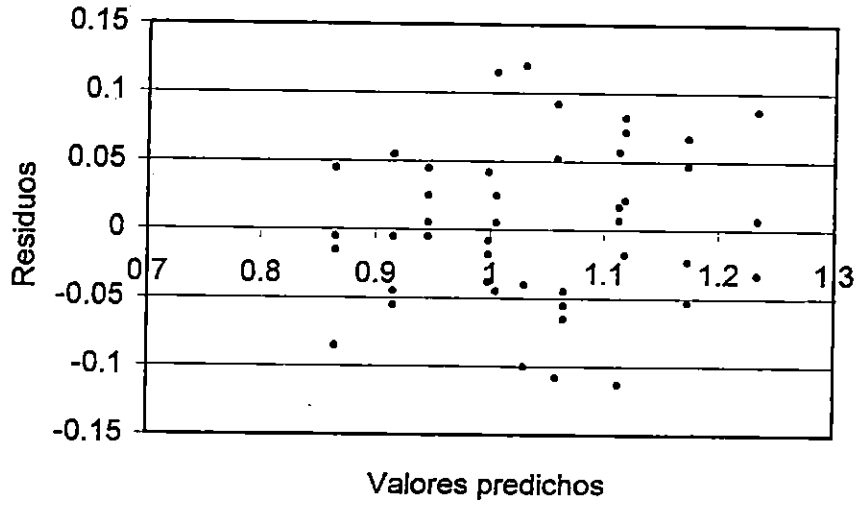


Figura 6.23. Gráfica de residuos vrs. valores predichos

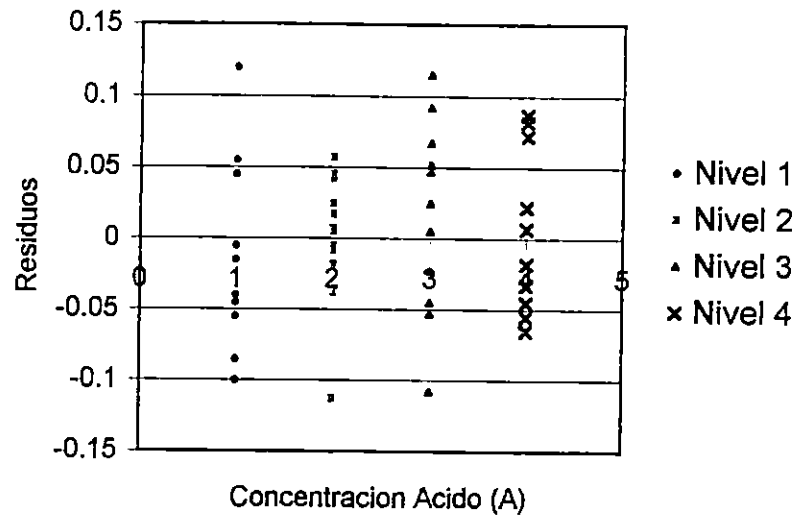


Figura 6.24. Grafica de residuos vrs. concentración de ácido sulfúrico.

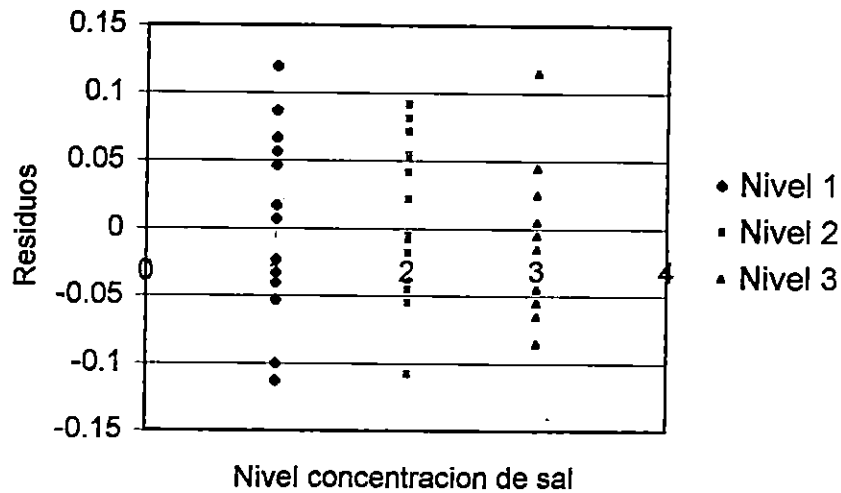


Figura 6.25. Grafica de residuos vrs. concentración de sal

Según el análisis de residuos muestra que el modelo encontrado es representativo para la variabilidad de los datos.

Ejemplo 6.6

Un químico realizó el siguiente experimento, por duplicado. Tal como se muestra en la tabla 6.36.

Tabla 6.36. Resultados obtenidos del experimento realizado.

Lectura	Temperatura (°C)	Catalizador (%)	pH	Rendimiento	
				1	2
1	-	-	-	60.4	62.1
2	+	-	-	75.4	73.1
3	-	+	-	61.2	59.6
4	+	+	-	67.3	66.7
5	-	-	+	66.0	63.3
6	+	-	+	82.9	82.4
7	-	+	+	68.1	71.3
8	+	+	+	75.3	77.1

Los valores altos y bajos de los tres factores son:

Nivel	Temperatura (°C)	Catalizador (%)	pH
-	130	1	6.8
+	150	2	6.9

Determinar:

- La significancia del diseño experimental
- La idoneidad del modelo

Solución:

- Para la significancia del diseño, se realiza de la siguiente manera:

- Planteamiento de hipótesis:

Hipótesis nula:

Ho: todos los factores=0 (A, B, C)

Ho: todas la interacciones dobles=0 (ab, ac, bc)

Ho: La interacción total=0 (abc)

Hipótesis alternativas:

H1: Los factores $\neq 0$, al menos uno

H1: Las interacciones dobles $\neq 0$, al menos una

H1: La interacción total $\neq 0$

Donde:

Factor A=la temperatura

Factor B=La concentración del catalizador

Factor C= El pH

2. Nivel de significancia: $\alpha=0.05$
3. Criterio de rechazo: $F_o > F_{0.05, 1, 8}$, es decir $F_o > 5.32$
4. Ordenamiento de datos, signos algebraicos, como se presentan en la tabla 6.37.

Tabla 6.37. Resumen de signos algebraicos y combinación de tratamientos para un diseño experimental 2^3 .

No.	Factor				Combinación De Tratamiento	Rendimiento total
	I	A	B	C		
1	+	-	-	-	1	122.5
2	+	+	-	-	a	148.5
3	+	-	+	-	b	120.8
4	+	+	+	-	ab	134.0
5	+	-	-	+	c	129.3
6	+	+	-	+	ac	165.3
7	+	-	+	+	bc	139.4
8	+	+	+	+	abc	152.4
					Total	1112.2

5. Determinación de los efectos, para ello se hará de la siguiente manera:

$$\text{Efecto} = \frac{\text{contraste}}{n2^{k-1}} = \frac{\text{contraste}}{2 * 2^2} = \frac{\text{contraste}}{8}$$

Para efectos principales:

$$\begin{aligned}
 A &= (1/8)(a+ab+ac+abc-b-c-bc-1) \\
 &= (1/8)(148.5+134+165+152.4-122.5-120.8-129.3-139.4) \\
 &= (1/8)(87.9)=10.9875
 \end{aligned}$$

De manera similar para los efectos B y C:

$$B=-2.125$$

$$C=7.575$$

Para las interacciones dobles:

$$\begin{aligned}
 AB &= (1/8)(1+ab+c+abc-a-b-ac-bc) \\
 &= (1/8)(122.5+134+129.3+152.4-148.5-120.8-165.3-139.4) \\
 &= (1/8)(-35.8)=-4.475
 \end{aligned}$$

De manera similar para las interacciones AC y BC

$$AC=1.225$$

$$BC=1.675$$

Para la interacción total:

$$\begin{aligned}
 ABC &= (1/8)(a+b+c+abc-1-ab-ac-bc) \\
 &= (1/8)(148.5+120.8+129.3+152.4-122.5-134-165.3-139.4) \\
 &= (1/8)(-10.2)=-1.275
 \end{aligned}$$

6. Graficar los efectos en pro balístico normal, tal como se muestra en la tabla 6.38.

Tabla 6.38. Resumen de efectos en orden ascendente que se graficará en pro balístico normal.

No. de efecto	$((Noi-0.5)/7)100$	Efecto	Tratamineto
1	7.14	-4.475	AB
2	21.43	-2.125	B
3	35.71	-1.275	ABC
4	50.00	1.225	AC
5	64.28	1.675	BC
6	78.57	7.575	C
7	92.86	10.9875	A

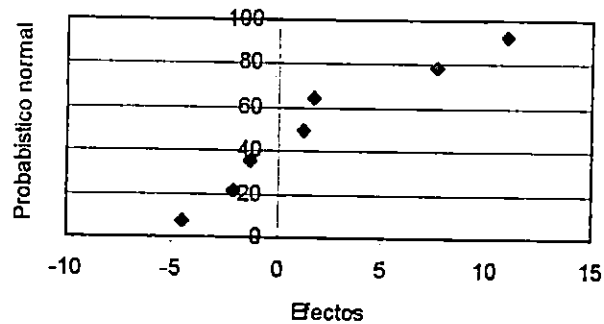


Figura 6.25. Grafica de efectos en probabilístico normal

7. Calcular la suma de cuadrados, de la forma siguiente:

$$SS = \frac{(\text{contraste})^2}{n \cdot 2^3} = \frac{(\text{contraste})^2}{16}$$

La suma de cuadrados son:

$$SS_A = (1/8)(87.9)^2 = 482.90$$

De manera similar para el resto de factores, interacciones doble y interacción triple, resultando:

$$SS_B = 18.0625$$

$$SS_C = 229.5225$$

$$SS_{AB} = 80.1025$$

$$SS_{AC} = 6.0025$$

$$SS_{BC} = 11.2225$$

$$SS_{ABC} = 6.5025$$

Para la suma de cuadrados totales se tiene:

$$SS_{TOTAL} = \sum y_{ijk}^2 - \frac{y^2}{16} = 78169.98 - 77311.8025 = 858.1775$$

Para el error, es:

$$\begin{aligned} SS_{ERROR} &= SS_{TOTAL} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} - SS_{ABC} \\ &= 858.1775 - 482.90 - 18.0625 - 229.5225 - 80.1025 - 6.0025 - 11.2225 - \\ &\quad 6.9249 \end{aligned}$$

=41.9249

8. Análisis de varianza, que se resume en la tabla 6.38, a continuación.

Tabla 6.38. Análisis de variancia para el ejemplo 6.6

Fuente de variación	Suma de cuadrados	G.L	Media de cuadrados	F _o	F _α
A	482.90	1	482.90	92.146*	5.32
B	18.0625	1	18.0625	3.447	5.32
C	229.5225	1	229.5225	43.797*	5.32
AB	80.1025	1	80.1025	15.285*	5.32
AC	6.0025	1	6.0025	1.145	5.32
BC	11.2225	1	11.2225	2.141	5.32
ABC	6.5025	1	6.5025	1.241	5.32
Error	41.9249	8	5.2406		
Total	858.1775	15			

9. Conclusión: Los factores que tienen un efecto significativo en el rendimiento de la reacción son: A, C y la interacción AB (ligeramente). Con un coeficiente de determinación (R^2)=0.9235, es decir que el modelo da un 92% aproximadamente la variancia de los datos.

- b. Idoneidad del modelo: para ello es necesario determinar los valores ajustados (Factores que tienen efecto significativo en las observaciones), residuos y graficarlos, estas graficas, son:
1. Probabilístico normal (Figura 6.26)
 2. Residuos vrs. Valores predichos (Figura 6.27)
 3. Residuos vrs. niveles de temperatura (Factor A) (Figura 6.28)
 4. Residuos vrs. niveles de pH (Factor C) (Figura 6.29)

Para ello se realizara de la siguiente manera:

1. El modelo lineal es:

$$\hat{y} = \bar{y}_{\dots} + \frac{A}{2}x_1 + \frac{C}{2}x_3 + \frac{AB}{2}x_1x_2$$

es decir:

$$\hat{y} = 69.5125 + 5.4937x_1 + 3.7875x_3 - 2.2375x_1x_2$$

2. Determinar los valores predichos por el modelo lineal. De la forma siguiente:

Cuando: $x_1=-1$, $x_2=-1$ y $x_3=-1$

$$y=57.9938$$

De forma similar para el resto de las combinaciones, en la tabla 6.39 se muestra el resumen de los valores predicho por el modelo.

Tabla 6.39. Resumen de los valores predichos por el modelo para los datos del ejemplo 6.6

Lectura	Temperatura (°C)	Catalizador (%)	pH	Rendimiento predicho
1	-	-	-	57.9938
2	+	-	-	73.4562
3	-	+	-	62.4688
4	+	+	-	68.9812
5	-	-	+	65.5688
6	+	-	+	81.0312
7	-	+	+	70.0438
8	+	+	+	76.5562

3. Determinar los residuos ($y_{ijk}-\hat{y}$), presentados en la tabla 6.40. a continuación:

Tabla 6.40. Resumen de residuos

Lectura	Temperatura (°C)	Catalizador (%)	pH	Residuos	
				1	2
1	-	-	-	2.4062	4.1062
2	+	-	-	1.9438	-0.3562
3	-	+	-	-1.2688	-2.8688
4	+	+	-	-1.6812	-2.2812
5	-	-	+	0.4312	-2.2688
6	+	-	+	1.8688	1.3688
7	-	+	+	-1.9438	1.2562
8	+	+	+	-1.2562	0.5438

4. Graficar los residuos, tales graficas se muestran en las figuras 6.26, 6.27, 6.28 y 6.29, a continuación:

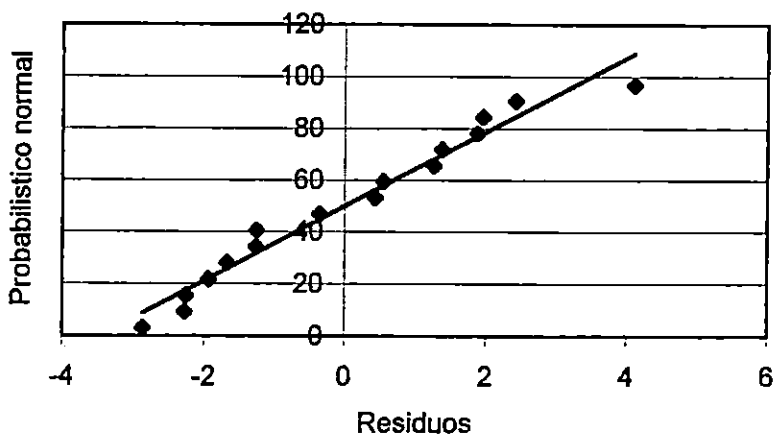


Figura 6.26. Graficas de residuos en probabilístico normal

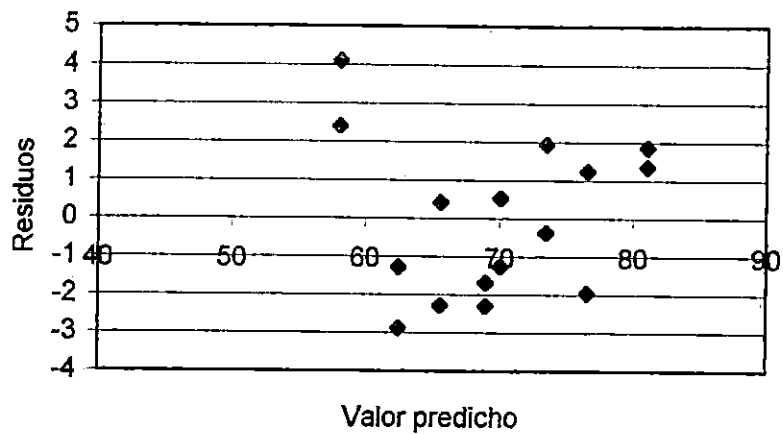


Figura 6.27. Grafico de residuos vrs. los valores predichos

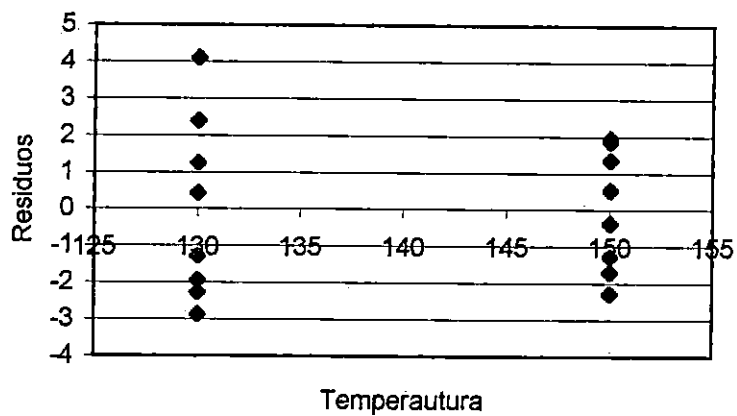


Figura 6.28. Grafico de residuos vrs. niveles de temperatura

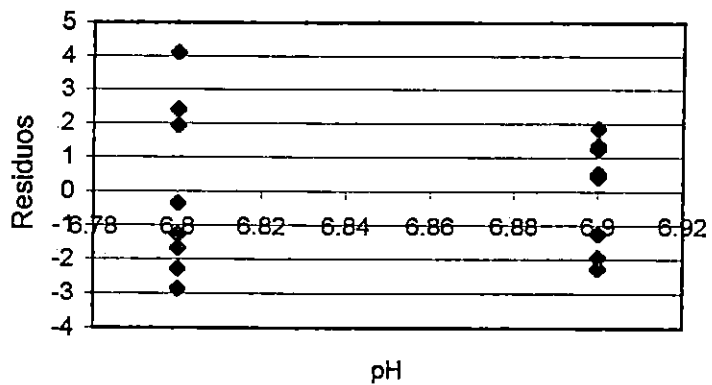


Figura 6.29. Grafico de residuos vrs. niveles de pH

Ejemplo 6.7

Con las observaciones del ejemplo 6.6, se desea realizar un diseño experimental con dos bloques y comprobar el efecto de replicas y de los bloques.

Solucion:

Para realizar la significancia del diseño, se hará de la siguiente manera:

1. Como son tres factores, dos replicas; los tratamientos y los bloques quedan de la siguiente forma:

Replica 1		Replica 2	
Bloque 1	Bloque 2	Bloque 1	Bloque 2
(1)=60.4	a=75.4	(1)=62.1	a=73.1
ab=67.3	b=61.2	ab=66.7	b=59.6
ac=82.9	c=66.0	ac=82.4	c=63.3
bc=68.1	abc=75.3	bc=71.3	abc=77.1
278.7	277.9	282.5	273.1

ABC confundido.

2. Determinar la suma de cuadrados

Para los factores, interacciones dobles y total es de la misma manera que en el ejemplo 7.9.

Para replicas es de la siguiente manera:

Replica 1=556.6

Replica 2=556.6

$$SS_{Repl\acute{c}as} = \frac{556.6^2 + 556.6^2}{8} - \frac{1112.2^2}{16} = 0.0625$$

Para los bloques, es:

$$SS_{BLOQUE} = \frac{278.7^2 + \dots + 273.1^2}{4} - \frac{1112.2^2}{16} = 11.1875$$

3. El análisis de varianza se muestra en la tabla 6.40 , a continuación

Tabla 6.40. Análisis de varianza para un diseño experimental 2^3 confundido en dos bloques

Fuente de variación	Suma de cuadrados	G.L	Media de cuadrados	Fo	F _α
Replicas	0.0625	1	0.0625	-----	5.59
Bloques	11.1875	1	11.1875	2.553	5.59
A	482.90	1	482.90	110.19 ^b	5.59
B	18.0625	1	18.0625	4.123	5.59
C	229.5225	1	229.5225	52.377	5.59
AB	80.1025	1	80.1025	18.279	5.59
AC	6.0025	1	6.0025	1.37	5.59
BC	11.2225	1	11.2225	2.561	5.59
ABC	6.5025	1	6.5025	1.484	5.59
Error	30.6749	7	4.3821		
Total	858.1775	16			

Conclusión: No existe efecto significativo debido a los bloques y replicas. Además, los efectos significativos en el rendimientos son los mismo determinados en el ejemplo 6.6 y por tanto el examen de idoneidad del modelo dará como resultados los mismo realizados en el ejemplo antes mencionado.

Ejemplo 6.8

Un artículo en "Solid State Technology" sobre ("Orthogonal Design for Process Optimization and its Application in Plasma Etching", mayo 1987, p 127-132). Describe la aplicación de los diseños factoriales en el desarrollo de un proceso de grabado en nitruro en un grabador de plasma en una sola oblea. El proceso emplea C_2F_6 como gas reactante. Es posible variar el flujo de gas, la potencia aplicada en el cátodo, la presión en la cámara del reactor, y el espaciado entre el ánodo y el cátodo (entrehierro). Los resultados se presentan en la tabla 6.41. con dos niveles para cada factor, tal como se muestra en la tabla 6.42.

Tabla 6.41. Niveles de los factores empleados en el experimento

Nivel	Factores			
	Entrehierro A (cm)	Presión B (mtorr)	Flujo de C_2F_6 C (SCCM)	Potencia D (W)
Bajo (-)	0.80	450	125	275
Alto (+)	1.20	550	200	325

Tabla 6.42. Resultados obtenidos expresadas en concentraciones en el grabado para el nitruro de silicio (A°/min).

A	B	C	D	Tasa de grabado (A°/min)
-	-	-	-	550
+	-	-	-	669
-	+	-	-	604
+	+	-	-	650
-	-	+	-	633
+	-	+	-	642
-	+	+	-	601
+	+	+	-	635
-	-	-	+	1037
+	-	-	+	749
-	+	-	+	1052
+	+	-	+	868
-	-	+	+	1075
+	-	+	+	860
-	+	+	+	1063
+	+	+	+	729

Determinar:

- Si los efectos principales afectan significativamente en el proceso grabado
- Un análisis de residuos para validar el modelo

Solución:

- Como el diseño se ejecuto con una replica y se tiene un diseño 2^4 . Para el caso se aplicará la dispersión del principio de los efectos; esto es, los efectos principales y las interacciones de orden menor suelen dominar el sistema.

Aplicando el algoritmo de Yates, así como se muestra en la tabla 6.43.

Tabla 6.43. Algoritmo de Yates para el ejemplo 6.8

Tratamientos	Lectura	1	2	3	4	Suma de cuadrados	Estimación de efectos
1	550	1219	2473	4984	12417	-----	-----
a	669	1254	2511	7433	-814	41412.25	-101.75
b	604	1275	3706	207	-13	10.5625	-1.625
ab	650	1236	3727	-1021	-60	225.00	-7.5
c	633	1786	164	-4	59	217.5625	7.375
ac	642	1920	43	-9	-198	2450.25	-24.75
bc	601	1935	-472	-49	-351	7700.625	-43.875
abc	635	1792	-549	-11	-128	1024	-16
d	1037	119	35	38	-2449	374850.0625	306.125
ad	749	45	-39	21	-1228	92249	-153.5
bd	1052	9	134	-121	-5	1.5625	-0.625
abd	868	34	-143	-77	38	90.25	8.5
cd	1075	-288	-74	-74	-17	18.0625	-2.125
acd	860	-184	25	-277	44	121	5.5
bcd	1063	-215	108	99	-203	2575.5625	-25.375
abcd	729	-334	-119	-227	-326	6642.25	-40.75

En la figura 6.30, se muestra los resultados de los efectos en probabilístico normal.

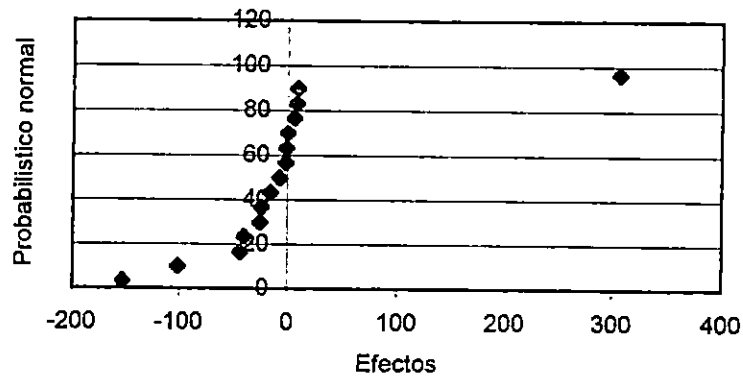


Figura 6.30. Grafica de los efectos en probabilístico normal

Según el grafico de probabilístico normal los efectos significativos son:

A (Enterhierro)

D (Potencia)

AD (Interacción Entrehierro-Potencia)

La asunción asumida al principio se despreciará la interacción efectos totales y triples, entonces el análisis de varianza queda como se muestra en la tabla 6.44.

Tabla 6.44. Análisis de varianza para el ejemplo 6.8

Fuente de variación	Suma de cuadrados	G.L	Media de cuadrados	Fo	F _α
A	41412.25	1	41412.25	20.28	6.61
B	10.5625	1	10.5625	<1	6.61
C	217.5625	1	217.5625	<1	6.61
D	374850.0625	1	374850.0625	183.99	6.61
AB	225	1	225	<1	6.61
AC	2450.25	1	2450.25	1.21	6.61
AD	94249	1	94249	48.79	6.61
BC	7700.625	1	7700.625	3.78	6.61
BD	1.5625	1	1.5625	<1	6.61
CD	18.0625	1	18.0625	<1	6.61
Error	10286.0005	5	2037.363		
Total	531420.938	15			

Conclusión: Se rechaza la hipótesis nula para los factores A, D y la interacción AD, ya que los efectos son significativos en el grabado por plasma. Puesto que, $A=-101.75$, el efecto del aumento del entrehierro entre el cátodo y el ánodo es disminuir la tasa de grabado. Sin embargo cuando se tiene que $D=306.125$, significa que a potenciales más altos incrementará la tasa de grabado.

b. Para realizar el examen de idoneidad del modelo, se realiza de la siguiente manera:

1. Encontrar el modelo lineal, en este caso es:

$$\hat{y} = 776.0625 - (101.725 / 2)x_1 + (306.125 / 2)x_4 - (153.5 / 2)x_1x_4$$

Es decir:

$$\hat{y} = 776.0625 - 50.8625x_1 + 153.0625x_4 - 76.75x_1x_4$$

2. Determinar por medio de la ecuación lineal los valores predichos, como se muestra en la tabla 6.45.

Tabla 6.45. Valores predichos para el ejemplo 6.8

A	B	C	D	Valor predicho
-	-	-	-	597.1875
+	-	-	-	648.875
-	+	-	-	597.1875
+	+	-	-	648.875
-	-	+	-	597.1875
+	-	+	-	648.875
-	+	+	-	597.1875
+	+	+	-	648.875
-	-	-	+	1056.3125
+	-	-	+	801.5
-	+	-	+	1056.3125
+	+	-	+	801.5
-	-	+	+	1056.3125
+	-	+	+	801.5
-	+	+	+	1056.3125
+	+	+	+	801.5

3. Determinar el valor de los residuos, como se muestra en la tabla 6.46.

Tabla 6.46. Resumen de los residuos para el ejemplo 6.8

A	B	C	D	Residuos
-	-	-	-	-4.7
+	-	-	-	20.125
-	+	-	-	7
+	+	-	-	1.125
-	-	+	-	41
+	-	+	-	-6.875
-	+	+	-	4
+	+	+	-	-13.875
-	-	-	+	-19.3125
+	-	-	+	-52.5
-	+	-	+	-4.3125
+	+	-	+	66.5
-	-	+	+	18.6875
+	-	+	+	58.5
-	+	+	+	6.6875
+	+	+	+	-72.5

4. Graficas de residuos, como se muestran en las figura 6.31, 6.32, 6.33 y 6.34.

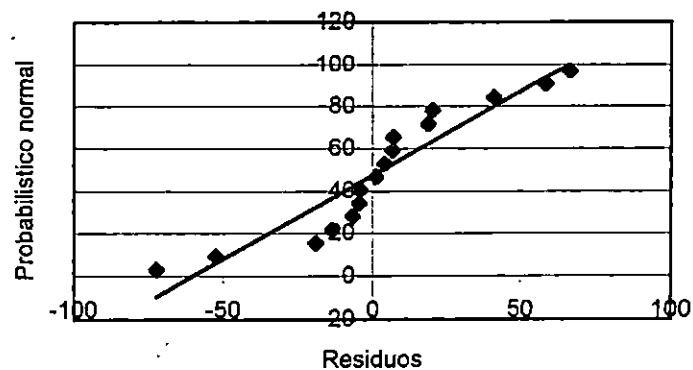


Figura 6.31. Grafica de residuo en probabilístico normal

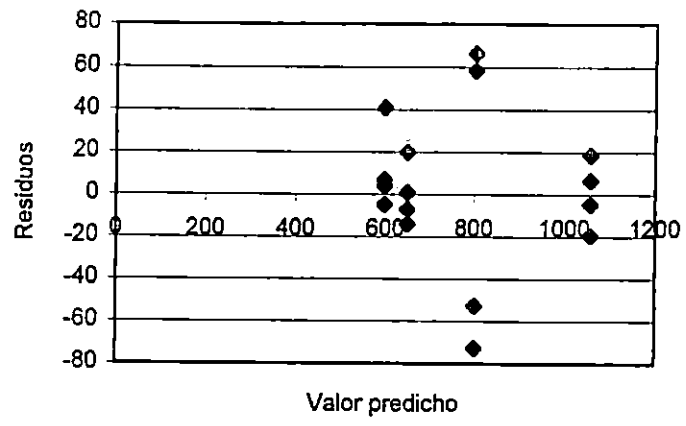


Figura 6.32. Grafica de residuos vrs. valor predicho

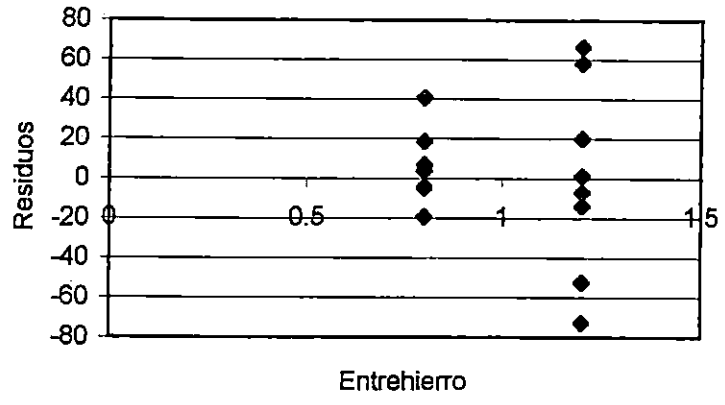


Figura 6.33. Grafica de residuos vrs. Entrehierro

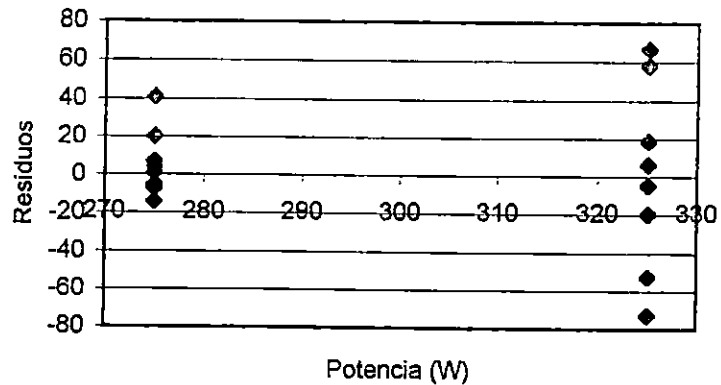


Figura 6.34. Grafica de residuos vrs. Potencia

Ejemplo 6.9

Considérese el empleo de un diseño experimental fraccional medio para el experimento descrito en el ejemplo 6.8. Sobre el grabado con plasma. Utilizando un 2^{4-1} con I=ABCD, para estudiar los cuatro factores: Entrehierro (A), Presión (B), Tasa de flujo de C_2F_6 (C) y potencia (D). Construyendo un diseño 2^3 con los factores A, B y C, fijando $D=ABC$.

Determinar:

- La significancia del modelo
- La idoneidad del modelo

Solución:

- Para determinar la significancia del modelo se hará de la siguiente manera:
 - Determinar los pseudónimos con las interacciones de tres factores, de la siguiente manera:

$$A.I=A.ABCD$$

$$A=A^2BCD$$

De manera similar:

$$B=ACD$$

$$C=ABD$$

$$D=ABC$$

Los otros seudónimos son:

$$AC=BD$$

$$AD=BC$$

2. Estimar los efectos principales y sus seudónimos, usando uso de la tabla 6.47.

Los efectos son:

$$l_A = A + BCD = \frac{1}{4}(-550 + 749 - 1052 + 650 - 1075 + 642 - 601 + 729) = -127.0$$

Tabla 6.47. Diseño experimental 2^{4-1} con definición I=ABCD

A	B	C	D=ABC	Fuente de variacion	Tasa de grabado
-	-	-	-	(1)	550
+	-	-	+	ad	749
-	+	-	+	bd	1052
+	+	-	-	ab	650
-	-	+	+	cd	1075
+	-	+	-	ac	642
-	+	+	-	bc	601
+	+	+	+	abcd	729
					6048

De manera similar para las otros seudónimos o alias:

$$l_B = B + ACD = 4.0$$

$$l_C = C + ABD = 11.50$$

$$l_D = D + ABC = 290.51$$

Y para las interacciones, es:

$$l_{AB} = AB + CD = -10.00$$

$$l_{AC} = AC + BD = -25.50$$

$$l_{AD} = AD + BC = -197.50$$

O por el algoritmo de Yates, como se muestra en la tabla 6.48.

Tabla 6.48. Algoritmo de Yates para el diseño factorial 2_{IV}^{4-1} del ejemplo 6.9

Combinación de tratamientos	Respuesta	(1)	(2)	(3)	Efecto	Estimación de efecto $2*(3)/2^3$
(1)	550	1299	3001	6048	-----	-----
a(d)	749	1702	3047	-764	A+ABC	-191
b(d)	1052	1717	-203	16	B+ACD	4
ab	650	1330	-561	-40	AB+CD	-10
c(d)	1075	199	403	46	C+ABC	11.5
ac	642	-402	-387	-358	AC+BD	89.5
bc	601	-433	-601	-790	BC+AD	197.5
abc(d)	729	128	561	1162	ABC+D	290.5

Los efectos principales son:

$$A = -127.00$$

$$D = 290.51$$

$$AD = -197.50$$

3. Graficar los efectos en probabilístico normal, para examinar los factores que tienen mayor efecto en el grabado en plasma. Ordenándolo de la forma mostrada en la tabla 6.49 y graficando la frecuencia acumulada vrs. valor de los efectos, como se muestra en la Figura 6.35.

Tabla 6.49. Tabla resumen de los efectos principales ordenados en forma ascendente

No. de efectos	Frecuencia acumulada	Efecto	Tratamiento
1	7.14	-197.5	AD
2	21.43	-127.00	A
3	35.71	-25.50	AC
4	50.0	-10.00	AB
5	64.28	4.00	B
6	78.57	11.50	C
7	92.86	290.51	D

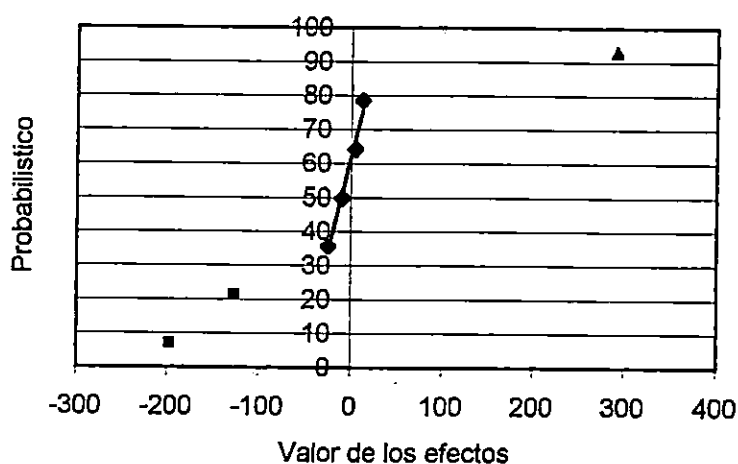


Figura 6.35. Grafica de efectos en probabilidad normal

4. Análisis de varianza, se basará en los efectos principales, como se muestra en la Tabla 6.50.

Tabla 6.50. Análisis de varianza para el ejemplo 6.9.

Fuente de variación	Suma de cuadrados	G.L	Media de cuadrados	F _o	F _α
A	32258	1	32258	73.99	6.61
D	168792.12	1	168792.12	387.16	6.61
AD	77618	1	77618	178.03	6.61
Error	2179.88	5	435.976		
Total	280848	8			

Conclusión: Al igual que en el ejemplo 7.11, los factores significativo son los mismos, es decir: Entrehierro (A), Potencia (D) y la interacción entre estos dos factores.

- b. La idoneidad del modelo, se procede de la siguiente manera:
1. Determinar el modelo lineal, para el caso es:

$$\hat{y} = 756 + \left(\frac{-127}{2}\right)x_1 + \left(\frac{190.51}{2}\right)x_4 + \left(\frac{-197.5}{2}\right)x_1x_4$$

Resolviendo se tiene:

$$\hat{y} = 756 - 63.5x_1 + 145.255x_4 - 98.75x_1x_4$$

2. Determinar los valores predichos por el modelo, los resultados obtenidos se muestran en la tabla 6.51.

Tabla 6.51. Resumen de los valores predichos por el modelo lineal

A	B	C	D=ABC	Valores predichos
-	-	-	-	575.495
+	-	-	+	739.255
-	+	-	+	1063.505
+	+	-	-	645.995
-	-	+	+	1063.505
+	-	+	-	645.995
-	+	+	-	575.495
+	+	+	+	739.255

3. Determinar los valores de los residuos, cuyos valores se muestran en la tabla 6.52.

Tabla 6.52. Tabla resumen de los residuos para el ejemplo 6.9

A	B	C	D=ABC	Residuos
-	-	-	-	-25.495
+	-	-	+	9.745
-	+	-	+	-11.505
+	+	-	-	1.005
-	-	+	+	11.495
+	-	+	-	-3.995
-	+	+	-	25.505
+	+	+	+	-10.255

4. Realizar el examen de idoneidad, graficando:
- Los residuos en probabilístico normal (Figura 6.36)
 - Los residuos contra los valores predichos (Figura 6.37)
 - Los residuos contra los niveles de entrehierro (Figura 6.38)
 - Los residuos contra los niveles de potencia (Figura 6.39)

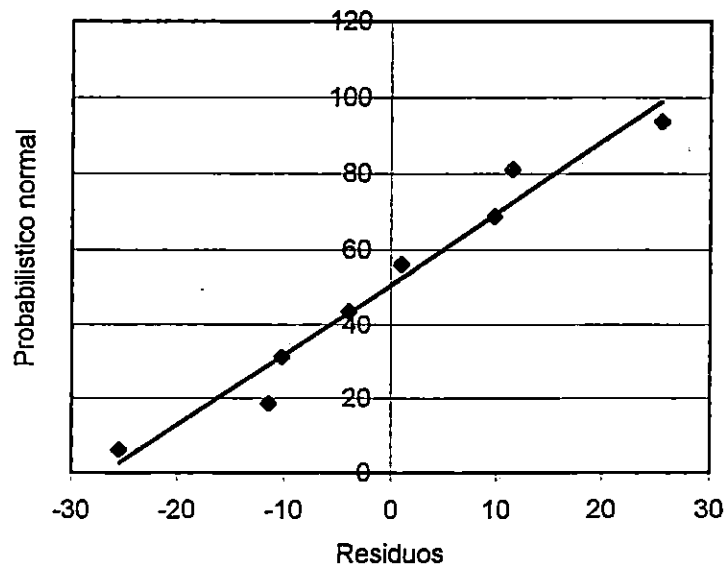


Figura 6.36. Gráfica de residuos en probabilístico normal

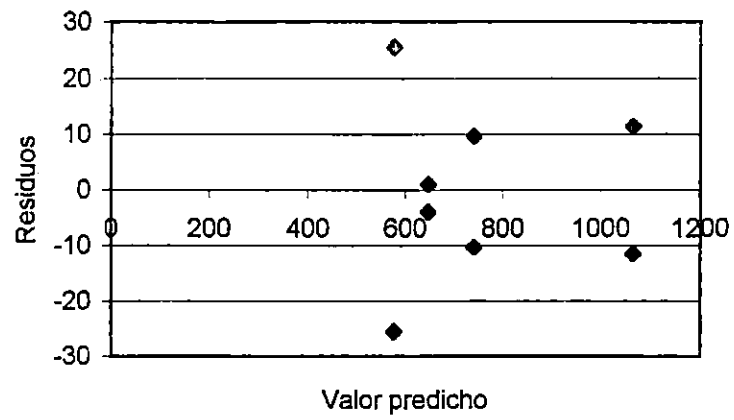


Figura 6.37. Grafica de los residuos vrs. los valores predichos

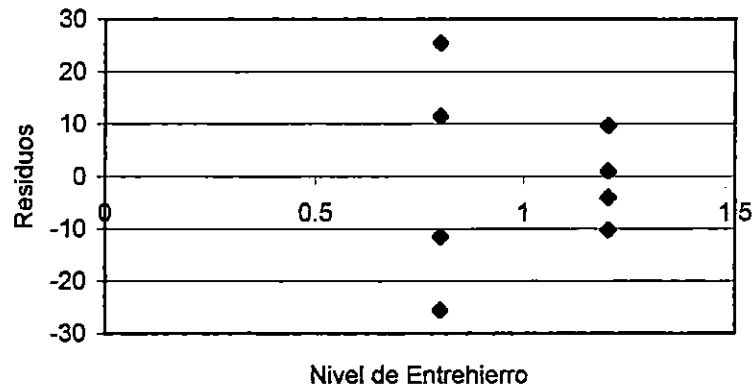


Figura 6.38. Grafica de residuos vrs. niveles de entrehierro

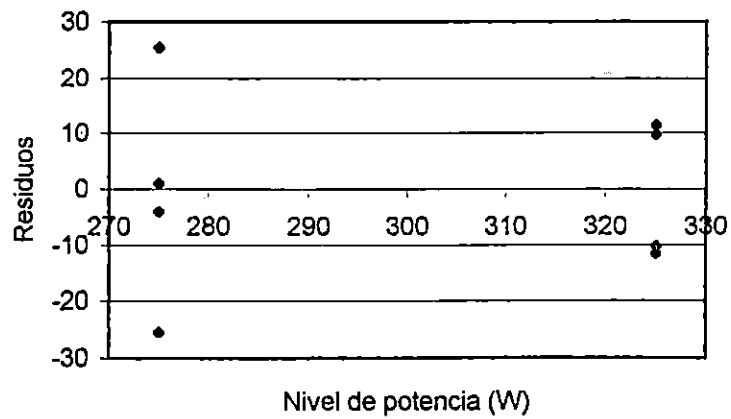


Figura 6.39. Grafica de residuos vrs. niveles de potencia

6.6 PROBLEMAS PROPUESTOS

- 6.1 Una fabrica de papel que emplea bolsa para comestibles, se interesa en mejorar la resistencia a la tensión del producto. Y un ingeniero de producción considera que la resistencia a la tensión es una función de la concentración de madera dura en la pulpa, y que el intervalo de concentración de madera dura de interés practico esta entre 5 y 20 por ciento. Los resultados obtenidos se muestran a continuación:

Resistencia a la tensión del papel (Psi)

Concentración de madera (%)	Observación					
	1	2	3	4	5	6
5	7	8	15	11	9	10
10	12	17	13	18	19	15
15	14	18	19	17	16	18
20	19	25	22	23	18	20

Asumiendo un diseño unifactorial completamente aleatorio; determinar:

- a. Si los porcentajes de concentración de la madera dura afecta la resistencia a la tensión de las cajas de papel.
 - b. Los parámetros del modelo
 - c. La idoneidad del modelo
- 6.2 En "Orthogonal Disign for Process Optimization and its Aplications to Plasma Etching" (Solid State Technology, mayo de 1987), G.S Yin y D. W. Jillie. Describen un experimento para determinar el efecto de la tasa de flujo de C_2F_6 sobre la uniformidad del grabado en una obleta de silicio empleado en una manufactura de un circuito intregado. Se utilizan tres tasas de flujo en el experimento, y la uniformidad resultante (en porcentaje) para 6 replicas se muestran enseguida.

Flujo de C_2F_6	Observaciones					
	1	2	3	4	5	6
125	2.7	4.6	2.6	3.0	3.2	3.8
160	4.9	4.6	5.0	4.2	3.6	4.2
200	4.6	3.4	2.9	3.5	4.1	5.1

Determinar:

- Si la tasa de flujo de C_2F_6 afecta la uniformidad del grabado
- Los parámetros del modelo
- La idoneidad del modelo

- 6.3 Una compañía de manufactura textil produce una fibra en un gran número de telares. La compañía está interesada en la variabilidad de la resistencia a la tensión entre los telares. Para investigar esto, un ingeniero de manufactura selecciona cuatro telares al azar y realiza cuatro determinaciones de resistencia en muestras de fibras elegidas al azar en cada telar. Los datos se muestran a continuación:

Telar	Observaciones			
	1	2	3	4
1	98	97	99	96
2	91	90	93	92
3	96	95	97	95
4	95	96	99	98

Determinar:

- Si no difieren significativamente en su capacidad para producir fibra resistente uniforme.
- La variabilidad entre telares y sus estimaciones
- Idoneidad del modelo

- 6.4 Un fabricante supone que existe diferencia en el contenido de calcio en lotes de materia prima que le son suministrado por un proveedor. Actualmente hay una gran cantidad de lotes en la bodega. Cinco de estos son elegidos aleatoriamente. Un químico realiza cinco prueba sobre cada lote y obtiene los siguientes resultados:

Lote 1	Lote 2	Lote 3	Lote 4	Lote 5
23.46	23.59	23.51	23.28	23.29
23.48	23.46	23.64	23.40	23.46
23.56	23.42	23.46	23.37	23.37
23.39	23.49	23.52	23.46	23.32
23.40	23.50	23.49	23.39	23.38

Determinar:

- Si hay variación significativa en el contenido de calcio de un lote a otro
 - Estime los componentes de variancia del modelo
 - La idoneidad del modelo
- 6.5 Se efectuó un experimento para determinar el efecto de cuatro diferentes compuestos químicos en la resistencia de una fibra. Estos compuestos se emplearon como parte del proceso de acabado de planchado permanente. Se seleccionaron cinco muestras de fibra, y un diseño de bloque aleatorio se efectuó probando cada tipo de compuesto químico en orden aleatorio en cada muestra de fibra. Obteniendo los siguientes resultados:

Tipo de compuesto químico	Muestra de fibra				
	1	2	3	4	5
1	1.3	1.6	0.5	1.2	1.1
2	2.2	2.4	0.4	2.0	1.8
3	1.8	1.7	0.6	1.5	1.3
4	3.9	4.4	2.0	4.1	3.4

Determinar:

- a. Si el efecto de los compuestos químicos es significativo.
- b. Estimar los parámetros
- c. Análisis de residuos y la idoneidad del modelo

6.6 Un técnico de laboratorio mide la resistencia a la ruptura de cada una de cinco clase de hilos por medio de cuatro instrumentos diferentes y obtiene los siguientes resultados (en onzas).

Hilo	Instrumento de medicion			
	1	2	3	4
1	20.6	20.7	20.0	21.4
2	24.7	26.5	27.1	24.3
3	25.2	23.4	21.6	23.9
4	24.5	21.5	23.6	25.2
5	19.3	21.5	22.2	20.6

Determinar:

- a. Si tiene una variabilidad significativa el tipo de hilo. Además si el efecto de las maquinas de prueba es significativa.
 - b. Estimar los parámetros
 - c. La idoneidad del modelo
- 6.7 Se están probando cuatro diferentes formulas de un pegamento industrial. La resistencia a la tensión del pegamento también está relacionada con el espesor de la capa de adherente. Cinco observaciones de la resistencia (X) en libras y el espesor (Y) en centésimas de pulgadas, se obtuvieron de cada formula. Los resultados se muestran en la siguiente tabla. Analícelos y exprese las conclusiones apropiadas.

Formula de pegamento							
1		2		3		4	
Y	X	Y	X	Y	X	Y	X
46.5	13	48.7	12	46.3	15	44.7	16
45.9	14	49.0	10	47.1	14	43.0	15
49.8	12	50.1	11	48.9	11	51.0	10
46.1	12	48.5	12	48.2	11	48.1	12
44.3	14	45.2	14	50.3	10	48.6	11

6.8 Se usan tres maquinas distintas para producir fibras monofilamentosa para una compañía textil. El ingeniero de proceso está interesado en determinar si existe diferencia en la resistencia a la ruptura de la fibra producida por las tres maquinas. Sin embargo, la resistencia de una fibra depende del grosor de la misma, siendo generalmente más resistentes las fibras de mayor grosor. Se selecciona una muestra aleatoria de cinco fragmentos por cada maquina. La resistencia de cada fibra (Y) y el grosor (X) correspondientes, se presentan a continuación:

Maquina 1		Maquina 2		Maquina 3	
Y	X	Y	X	Y	X
36	20	40	22	35	21
41	25	48	28	37	23
39	24	39	22	42	26
42	25	45	30	34	21
49	32	44	28	32	15

6.9 Se están investigando los efectos sobre la resistencia de papel del porcentaje de concentración de madera dura en la pulpa cruda, la limpieza, y el tiempo de cocido de la pulpa. Analizar los datos haciendo uso de la significancia del modelo y de la idoneidad del modelo. Los resultados obtenidos se muestran en la siguiente tabla, suponiendo que los tres factores son fijos.

% de concentración de la madera	Tiempo de cocido de 1.5 horas			Tiempo de cocido de 2.0 horas		
	Limpieza			Limpieza		
	400	500	650	400	500	650
10	96.6	97.7	99.4	98.4	99.6	100.6
	96.0	96.0	99.8	98.6	100.4	100.9
15	98.5	96.0	98.4	97.5	98.7	99.6
	97.2	96.9	97.6	98.1	98.0	99.0
20	97.5	95.6	97.4	97.6	97.0	98.5
	96.6	96.2	98.1	98.4	97.8	99.5

6.10 Un ingeniero sospecha que el acabado superficial de una pieza metálica es afectado por el tipo de pintura utilizado y el tiempo de secado. Selecciona tres tipos de secado: 20, 25 y 30 minutos. Y elige al azar dos tipos de pintura de varios de los que dispone. El ingeniero realiza un experimento y obtiene los datos que se muestran adelante. Analice los datos y extraiga conclusiones. Estime las componentes de la varianza.

Pintura	Tiempo de secado (min)		
	20	25	30
1	74	73	78
	64	61	85
	50	44	92
	92	98	66
2	86	73	45
	68	88	85

6.11 Se desea determinar los efectos de la temperatura del gas y el ancho del horno sobre el tiempo requerido para fabricar coque. Las condiciones experimentales son:

Ancho del horno (pulgadas)	Temperatura del gas (°F)
4	1600
4	1900
8	1600
8	1900
12	1600
12	1900

Los resultados se presentan en la siguientes tabla por triplicado

Factor A Ancho del horno	Factor B Temperatura del gas	Replicas		
		1	2	3
4	1600	3.5	3.0	2.7
4	1900	2.2	2.3	2.4
8	1600	7.1	6.9	7.5
8	1900	5.2	4.6	6.8
12	1600	10.8	10.6	11.0
12	1900	7.6	7.1	7.3

Determinar:

- El análisis de varianza tomando los datos como efecto fijo, aleatorias y fijando el factor a y el factor b aleatorio
- La significación de los efectos para un nivel de significación del 0.05
- Idoneidad del modelo

6.12 son los Para determinar las condiciones optimas para un baño galvanoplástico, los efectos de la concentración de sulfuro y de la temperatura del baño sobre la reflectividad del metal revestido se estudian en un experimento factorial 2x5 los resultados de tres replicas siguientes:

Concentración (gramos/lts)	Temperatura (°F)	Reflectividad		
		1	2	3
5	75	35	39	36
5	100	31	37	36
5	125	30	31	33
5	150	28	20	23
5	175	19	18	22
10	75	38	46	41
10	10	36	44	39
10	125	39	32	38
10	150	35	47	40

10	175	30	38	31
----	-----	----	----	----

Analice y determine la condición o condiciones del baño que produce la máxima capacidad de reflexión. También, construya un intervalo de confianza de 95 por ciento para la capacidad de reflexión del baño galvanoplástico correspondiente de las condiciones óptimas.

6.13 Durante la elaboración de la cerveza, se añade a este un ingrediente retardador del deterioro. Para determinar en que medida se prolonga la vida útil de la cerveza es afectado por la cantidad de este ingrediente agregado a cada lote, y como tales cambios de la vida útil podrían depender de la temperatura de pasteurización de la cerveza, se diseñó un experimento factorial 3x4 con dos replicas. La vida útil de la cerveza (semanas) se presenta a continuación:

Cantidad de ingrediente (gr/lote)	Temperatura de pasteurización (°C)	Vida útil (semanas)	
		Replica 1	Replica 2
2	70	2.1	1.6
2	75	2.6	1.9
2	80	2.9	2.4
3	70	1.4	1.7
3	75	1.9	2.2
3	80	2.3	2.7
4	70	0.5	0.9
4	75	1.2	0.8
4	80	1.7	1.4
5	70	1.0	1.6
5	75	2.2	1.3
5	80	2.3	2.1

- a. Calcule las sumas de cuadrados correspondientes a los efectos principales y la interacción, y presente los resultados en una tabla de análisis de varianza.
- b. Interprete los resultados del experimento

c. Compruebe la idoneidad del modelo.

6.14 En el departamento de control de calidad de una planta de acabado de tela se está estudiando el efecto de diversos factores sobre el teñido de la tela de algodón y fibra sintética, la cual se utiliza en la manufactura de camisas para caballero. Se seleccionaron tres operadores, tres niveles de tiempo y dos niveles de temperatura, y se tiñeron tres pequeñas muestras de tela dentro de cada conjunto de condiciones. La tela teñida fue comparada con un estándar y se le asignó una calificación numérica. Los resultados se muestran a continuación. Analice los resultados obtenidos y realice las recomendaciones adecuadas. Comente sobre la idoneidad del modelo.

Tiempo de ciclo (minutos)	Temperatura					
	300 °C			350 °C		
	Operario			Operario		
	1	2	3	1	2	3
40	23	27	31	24	38	34
	24	28	32	23	36	36
	25	26	29	28	35	39
50	36	34	33	37	34	34
	35	38	34	39	38	36
	36	39	35	35	36	31
60	28	35	26	26	36	28
	24	35	27	29	37	26
	27	34	25	25	34	24

6.15 Se piensa que cuatro factores posiblemente afecten el sabor de un refresco: el tipo de saborizante (A), proporción entre jarabe y agua (B), el nivel de carbonatación (C) y la temperatura (D). Cada factor puede ejecutarse en dos niveles produciendo un diseño 2^4 . En cada ejecución en el diseño se dan muestras de la bebida a un grupo de prueba compuesto de 20 personas. Cada una de ellos asigna una calificación de 1 a 10 al refresco.

La calificación total es la variable de respuesta, y el objetivo es encontrar una formulación que maximice la calificación total. Se ejecutan dos replicas de este diseño, y los resultados se presentan aquí. Analice los datos y extraiga conclusiones. Y compruebe la idoneidad del modelo.

Combinación de tratamientos	Replicas		Combinación de tratamientos	Replicas	
	1	2		1	2
(1)	190	193	d	198	195
a	174	178	ad	172	176
b	181	185	bd	187	183
ab	183	180	abd	185	186
c	177	178	cd	199	190
ac	181	180	acd	179	175
bc	188	182	bcd	187	184
abc	173	170	abcd	180	180

- 6.16 Para determinar el efecto sobre el gusto de tres factores diferentes en la fabricación de latas para gaseosas se efectuó un experimento en el que el sabor de una gaseosa fue calificada por un juez con base en una escala del 1 al 10. los resultados fueron los siguientes:

A Lubricante	B Calor	C Resina	Calificación	
			1	2
Fresco	Sin calentar	A	6	8
Fresco	Sin calentar	B	8	7
Fresco	Calentado	A	9	9
Fresco	Calentado	B	1	2
Viejo	Sin calentar	A	6	7
Viejo	Sin calentar	B	6	8
Viejo	Calentado	A	9	8
Viejo	Calentado	B	2	3

Desarrollar:

- El análisis de varianza y concluya sobre los resultados obtenidos
- Concluir sobre la idoneidad del modelo

- 6.17 Un ingeniero químico, para estudiar una reacción realizó un factorial 2^3 y obtuvo los siguientes resultados, mostrados a continuación:

Experimento elemental	Temperatura (°C)	Concentración (%)	Velocidad de agitación (rpm)	Producción (%)
1	50	6	60	54
2	60	6	60	57
3	50	10	60	69
4	60	10	60	70
5	50	6	100	55
6	60	6	100	54
7	50	10	100	80
8	60	10	100	81

Determinar:

- a. La significancia del modelo y concluya sobre ello
 - b. La idoneidad del modelo y concluya sobre ello
- 6.18 De acuerdo a los resultados presentados en los problemas 6.15, 6.16 y 6.17, desarrolle un arreglo para un modelo experimental 2^k confundido para dos y cuatro bloques, y determinar:

Determinar:

- a. La significancia del modelo y comparar y concluir sobre los resultados obtenidos.
 - b. La idoneidad del modelo y hacer conclusiones de ello
- 6.19 De acuerdo a los resultados presentados en los problemas 6.15, 6.16 y 6.17, desarrolle un arreglo para un modelo experimental 2^{k-1} y determinar:

- a. La significancia del modelo y comparar y concluir sobre los resultados obtenidos.
- b. La idoneidad del modelo y hacer conclusiones de ello

CAPÍTULO 7

MÉTODOS Y DISEÑOS DE SUPERFICIES DE RESPUESTAS

La metodología de superficie de respuestas, (MSR ó RSM, por sus siglas en inglés) es un conjunto de técnicas estadísticas-matemáticas útiles para modelar y analizar problemas en las cuales una respuesta de interés e influenciada por varias variables, y el objetivo es optimizar esta respuesta.

Por ejemplo: un ingeniero químico desea determinar los niveles de temperatura (x_1) y presión (x_2) que maximizan el rendimiento (y) de una reacción química que ocurre en un proceso a gran escala, laboratorio ó planta piloto. Ó, un ingeniero en Alimentos desea determinar los niveles de temperatura (x_1) y contenido de humedad (x_2) que minimice la oxidación (y) de un producto alimenticio ó su vida útil, para ello se dice que:

$$y = f(x_1, x_2) + \varepsilon \quad 7.1$$

Donde ε representa el ruido ó error observado en la respuesta (y). Si la respuesta se denota por $E(y)=f(x_1, x_2)=\eta$, entonces la superficie de respuesta se representa por:

$$\eta = f(x_1, x_2) \quad 7.2$$

La cual se denomina superficie de respuesta.

Estas técnicas se utilizan en las áreas de ingeniería química y áreas afines, con el objeto de responder a interrogantes de varios tipos, tales:

1. ¿Cómo, en una región un conjunto de variables afectan a la variable respuesta?

2. ¿Qué niveles, si que existen algunos, de las variables de control, permitirán obtener una respuesta deseada?
3. ¿Qué valores de las variables de control producirán un máximo ó minimice en una respuesta y cómo es la superficie en las inmediaciones de este máximo ó mínimo?

En la mayoría de los problemas de RSM, la forma de la relación entre la respuesta y variables independientes se desconoce. Por ello, el primer paso en la RSM consiste en determinar una aproximación apropiada a la relación funcional real entre (y) y el conjunto de variables independientes. Por lo general se emplea un polinomio de orden bajo sobre alguna región de las variables independientes. Si la respuesta es descrita adecuadamente por una función lineal de las variables independientes, la función aproximada es el modelo de primer orden, se tiene:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad 7.3$$

Si existe una región de la RSM que satisfaga este modelo, se debe usar un polinomio de orden mayor, por ejemplo uno de segundo orden, como el siguiente:

$$\hat{y} = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ij} x_i^2 + \sum_{i=1}^k \sum_{j=1}^n \beta_{ij} x_i x_j + \varepsilon \quad 7.4$$

Esto, cuando existe curvatura en el sistema. Casi todos los problemas de RSM utilizan uno ó ambos polinomios de aproximación. Por supuesto, es improbable que un modelo polinomial sea una aproximación razonable de la relación entre las variables reales. Mas sin embargo, se aproximan muy bien para intervalos o niveles de la variables pequeñas.

En el capítulo 5, se describe la determinación de los parámetros de los coeficientes de estos polinomios, mediante el método de los mínimos cuadrados.

Tal análisis de regresión será aproximadamente reproducido el comportamiento real del sistema. También, la estimación de los parámetros del modelo polinomial se hace efectivo si se utilizan los diseños de experimentos apropiados para la recopilación de datos. Estos modelos determinados a partir del diseño de experimento se ilustró en el Capítulo 6.

La RSM es una técnica secuencial. Y a menudo, cuando se considera un punto ó región sobre la superficie de respuesta alejado del óptimo, el polinomio de primer orden es apropiado, porque existe poca curvatura en el sistema. Lo cual consiste en guiar al experimentador rápida y eficazmente a los alrededores del punto óptimo. Una vez que se encuentra próximo a la región del punto óptimo, puede aplicarse un modelo de orden mayor (más elaborado) ó de segundo orden, y determinar el punto óptimo. La cual puede ser:

- a. Respuesta máxima: ascenso de una loma
- b. Respuesta mínima: descenso al valle

Las representaciones gráficas que genera este tipo de análisis son: grafica de superficie de respuestas y sus contornos de esta superficie, como se muestran las Figuras 7.1 y 7.2.

7.1 MÉTODOS PARA DETERMINAR LA COMBINACIÓN OPTIMA DEL SISTEMA

El objetivo es establecer un programa experimental para determinar los niveles de los factores que afectan a la variable respuesta (y) la hacen máxima ó mínima (óptima). Por ejemplo: los niveles que hacen máxima la conversión ó rendimiento de una reacción química ó los niveles de los factores que hacen mínima el deterioro de un alimento. La cual la respuesta puede ser una cantidad real de la respuesta ó una medida de calidad ó evolución de la misma.

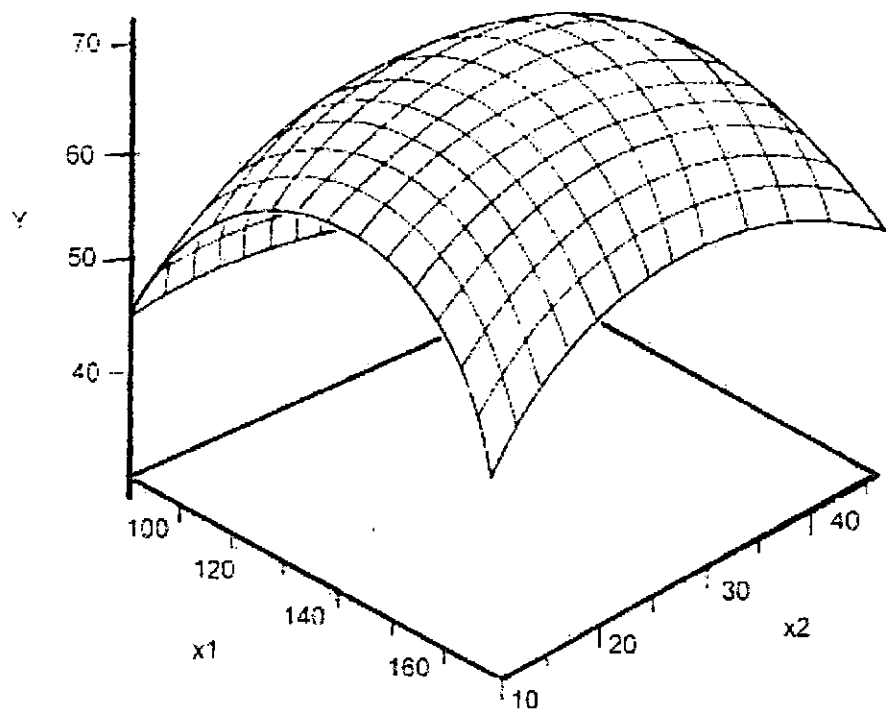


Figura 7.1. Superficie de respuesta tridimensional en la que se observa la variable respuesta(η) en función de X_1 y X_2 .

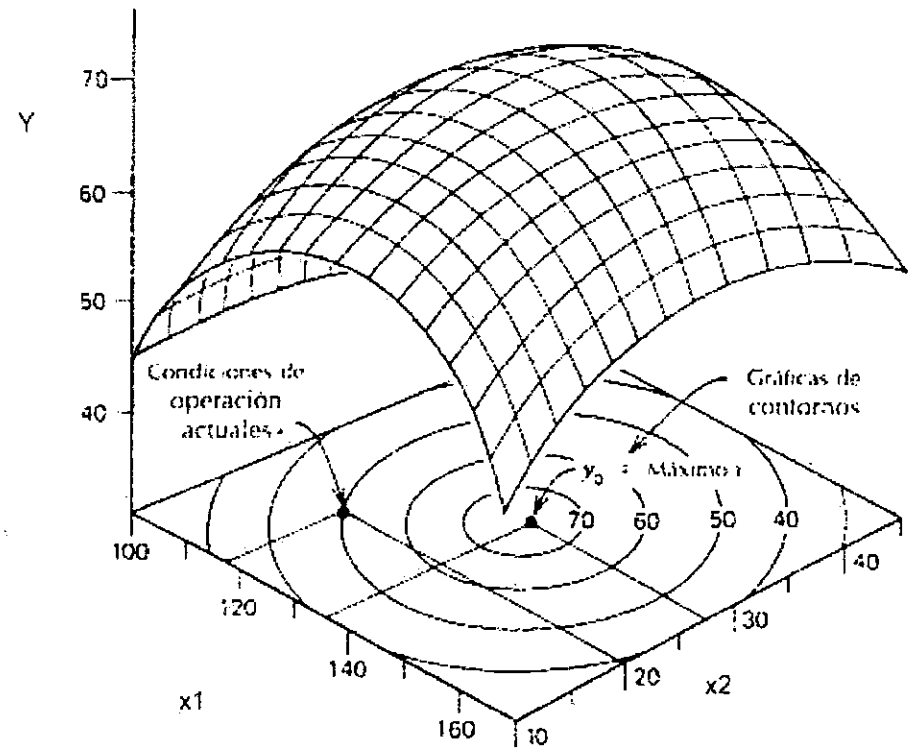


Figura 7.2. Gráfica de contornos de una superficie de respuesta

Además, de localizar los niveles de los factores que maximizan ó minimizan (y), es valioso visualizar como varía (y) en los alrededores del óptimo, cuando estos niveles de los factores se cambian de su valor óptimo. Y esto es valioso, por varias razones:

- a. La aplicación a gran escala de los resultados puede ser no factible establecer las condiciones exactamente a su nivel óptimo. Y alguna combinación de los niveles diferentes del óptimo puede ser más económica de mantener.
- b. A menudo, más de una característica de la variable respuesta es importante. Un cambio de un factor, alejándose del óptimo, puede ser preferible a causa de su efecto sobre una de estas otras características.
- c. La forma de la superficie de respuesta cerca del óptimo puede dar indicación de la naturaleza del fenómeno estudiado.
- d. La superficie de respuesta puede carecer de un máximo ó de un mínimo en la región de experimentación (o en la región en la cual el proceso se lleve a cabo en la practica).

Para determinar un punto óptimo se requiere comúnmente de una serie de técnicas que secuencialmente se puede lograr. Entre los métodos más usados se pueden mencionar:

1. Método del factor común
2. Método de la máxima pendiente
3. Ajuste de diseños factoriales
4. Selección al azar de puntos de prueba.

Para objeto del desarrollo del presente trabajo, se aplicará los métodos 2 y 3.

7.2 MÉTODO DE LA MÁXIMA PENDIENTE

Se asume que las estimaciones iniciales de operación óptima para un sistema estará alejada del óptimo real. El objetivo es, moverse en forma analítica a las cercanías ó entorno del óptimo. Para ello es necesario suponer un modelo de primer orden, la cual es adecuado a la superficie real en regiones pequeñas de las X (alejado del punto estacionario es decir del óptimo). Para ello, es necesario recurrir a los métodos planteados e ilustrados en el Capítulo 5.

El método de máxima pendiente con ascenso o descenso es un procedimiento para recorrer secuencialmente a través de la trayectoria de la máxima pendiente, es decir, la dirección donde muestra el modelo máxima pendiente o incremento de la variable respuesta. Las curvas a nivel de \hat{y} consta de una serie de rectas paralelas, siendo la dirección paralela a la normal de la superficie de respuesta ajustada (Figura 7.3). En forma general, la trayectoria de máxima pendiente en ascenso se toma como la recta que atraviesa el centro de la región de interés y es normal a la superficie ajustada.

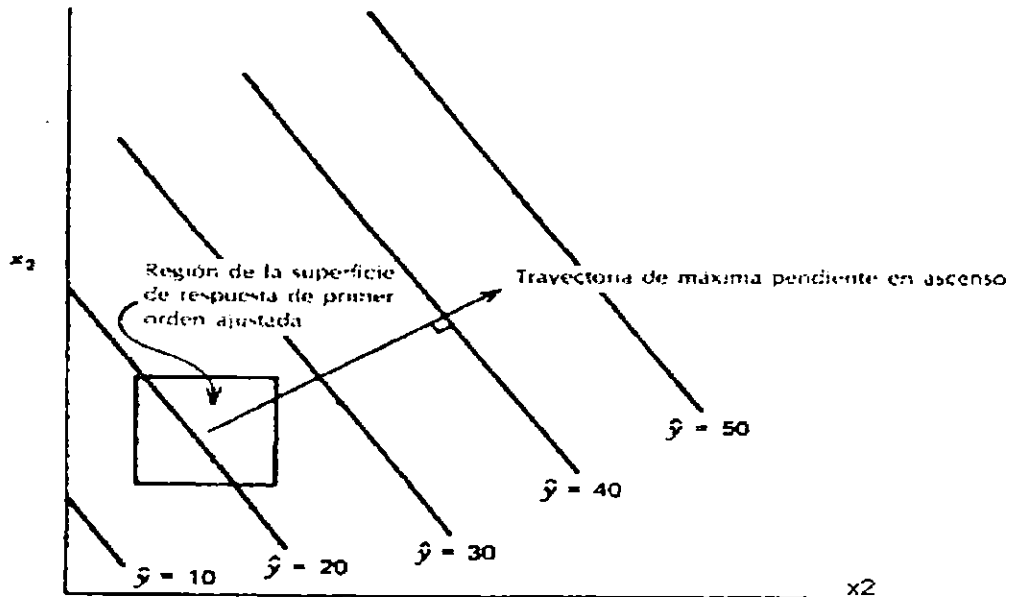


Figura 7.3. Superficie de respuesta de primer orden y trayectoria de máxima pendiente en ascenso

Para desarrollar este método, este consta de dos partes:

1. Diseños con modelos lineales ó de primer orden
2. Diseños no lineales, generalmente de segundo orden (cuadráticos)

7.2.1 DISEÑOS DE PRIMER ORDEN

El objeto de proponer un diseño lineal, se debe que los niveles lejanos del punto óptimo, es descrito adecuadamente por un modelo lineal. Ya que por lo general, tiene un comportamiento lineal. Los parámetros de este modelo lineal propuesto, puede determinarse por:

1. Regresión lineal múltiple (método de los mínimos cuadrados)
2. El modelo lineal de los diseños experimentales factoriales (2^k y fraccionarios), generalmente para investigaciones exploratorias

Los diseños proporcionados en la Tabla 8.1, el experimentador puede juzgar sobre si la ecuación que rige el modelo lineal se ajusta adecuadamente, ya que tiene el valor de la variancia del error experimental a su disposición.

Tabla 7.1. Algunos diseños útiles factoriales 2^k para ajustar ecuaciones lineales

Número de factores	Tamaño de experimento (unidades)	Fracción de una repetición	Número de G.L para la "falta de ajuste"
3	8	1	4
4	8	$\frac{1}{2}$	3
5	8	$\frac{1}{4}$	2
5	16	$\frac{1}{2}$	10
6	8	$\frac{1}{8}$	1
6	16	$\frac{1}{4}$	9

7.2.2 DISEÑOS DE SEGUNDO ORDEN

Generalmente debido a la curvatura de la superficie real, el investigador requiere de un modelo de grado mayor o igual a 2, para aproximar la respuesta cuando se encuentra relativamente cercana al óptimo. En la mayoría de los casos, el modelo de segundo orden es la ecuación 8.4

$$\hat{y} = \beta_0 + \sum_{i=1}^k \beta_i X_i + \sum_{i=1}^k \beta_{ii} X_i^2 + \sum_{i=1}^k \sum_{j=1}^n \beta_{ij} X_i X_j + \varepsilon$$

Con este modelo, el fin es encontrar el conjunto óptimo de trabajo para las X y caracterizar la naturaleza de la superficie de respuesta. Para estimar los coeficientes de regresión para este modelo, cada variable X debe tomarse cuando menos tres diferentes niveles.

7.2.2.1 Localización del Punto Estacionario

La respuesta se maximiza cuando las derivadas parciales de la respuesta predicha con respecto a X_i es igual a cero, es decir:

$$\frac{\partial \hat{y}}{\partial X_1} = \frac{\partial \hat{y}}{\partial X_2} = \dots = \frac{\partial \hat{y}}{\partial X_k} \quad 7.5$$

Es decir, $X_{1,0}$, $X_{2,0}$, ..., $X_{k,0}$ la cual se denota:

1. Respuesta máxima
2. Respuesta mínima
3. Punto de silla

Estas tres posibilidades se muestran en la Figura 7.4

Puede obtenerse usando la notación matricial para el modelo de segundo orden, como sigue:

$$\hat{y} = \beta_0 + x'b + x'\beta x$$

7.6

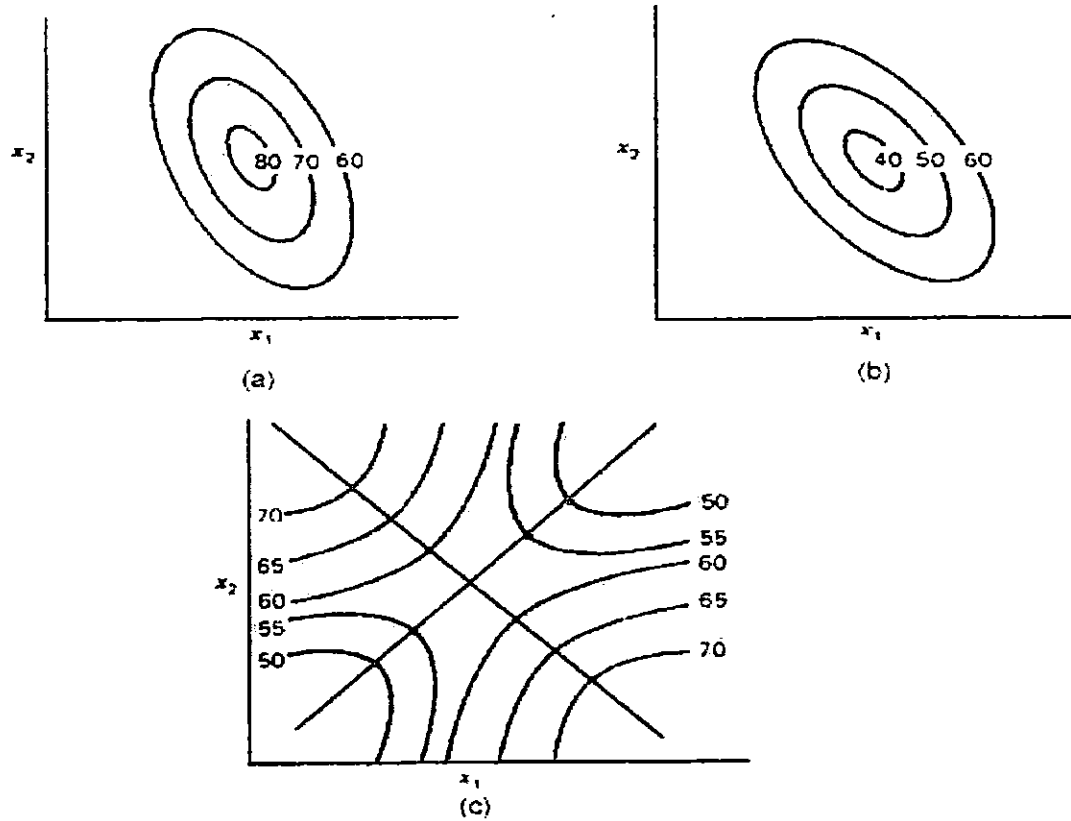


Figura 7.4. Puntos estacionarios en una superficie de respuesta de segundo orden ajustada.
 (a) respuesta máxima, (b) respuesta mínima, (c) punto de silla

donde:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \quad b = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \quad y \quad \beta = \begin{bmatrix} \beta_{11} & \beta_{12}/2 & \dots & \beta_{1k}/2 \\ \cdot & \beta_{22} & \dots & \beta_{2k}/2 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \beta_{kk} \end{bmatrix} \quad 8.7$$

Para el valor predicho en el punto estacionario, es:

$$\hat{y} = \beta_0 + \frac{1}{2} x_0^T b \quad 7.8$$

donde: $x_0 = -\frac{1}{2} \beta^{-1} b$

7.2.2.2 Caracterización de la Superficie de Respuesta.

Por caracterizar se entiende, determinar:

1. Punto de respuesta máxima
2. Punto de respuesta mínima
3. Punto de silla
4. La sensibilidad relativa de la respuesta a las variables x_i .

La forma conveniente consiste en examinar la gráfica de los contornos del modelo ajustado. Para ello, se debe transformar el modelo a un sistema de coordenadas con origen en el punto estacionario x_0 y rotar (girar) los ejes del sistema hasta que sean paralelas a los ejes principales de la superficie de respuesta ajustada. Esto, es posible demostrar el modelo ajustado por:

$$\hat{y} = \hat{y}_0 + \lambda_1 W_1^2 + \lambda_2 W_2^2 + \dots + \lambda_k W_k^2 \quad 7.9$$

donde: W_i : Son variables independientes transformadas
 λ Son constantes (valores propios) llamadas raíces características autovalores ó eigen valores.

Esta forma del modelo ajustado, es llamada forma canónica del modelo. A menudo puede ser necesario determinar la relación entre las variables canónicas (W_i) y las del diseño (x_i). Esto es, para cuando es imposible operar el proceso a gran escala en el punto estacionario, debido que la operación de los niveles óptimo de los factores cuesta demasiado. Entonces se desea alejarse del óptimo a

un menor costo y sin sacrificar demasiado el rendimiento. En general X se relaciona con las variables canónicas (W) mediante la siguiente relación:

$$W = M'(x - x_0) \quad 7.10$$

donde: M : Es la matriz ($k \times k$) ortogonal, entonces m_i es la i -ésima columna de M , es decir:

$$(\beta - \lambda_i I)m_i = 0 \quad 7.11$$

donde: $\sum_{j=1}^k m_{ji}^2 = 1$

En general la naturaleza de la superficie de respuesta puede determinarse a partir del punto estacionario a partir de los signos y magnitudes de las λ_i , nos dice para un modelo de segundo orden, lo siguiente:

1. Si todas las λ_i son positivas, x_0 es un punto de respuesta mínima.
2. Si todas las λ_i son negativas, x_0 es un punto de respuesta máxima
3. Si las λ_i son de distintos signos, x_0 es un punto de silla.

7.3 SISTEMAS DE LOMAS

A menudo se encuentran variaciones de las superficies de respuestas máxima, mínima y punto de silla puras, que se consideran en la sección anterior. Considérese el modelo de segundo grado presentado en la Ecuación 8.9:

$$\hat{y} = \hat{y}_0 + \lambda_1 W_1^2 + \lambda_2 W_2^2 + \dots + \lambda_k W_k^2$$

Se debe considerar:

1. Una de las λ_i es muy pequeña, por lo tanto es insignificante el efecto que ocasiona a la variable respuesta.
2. El modelo de segundo grado queda representado por las λ_i que resulten con magnitud significativa.

Entonces, si $\lambda_1 \approx 0$, su valor es cercano a cero, el modelo de segundo grado se reduce a la siguiente expresión:

$$\hat{y} = \hat{y}_0 + \lambda_2 W_2^2 + \dots + \lambda_k W_k^2 \quad 7.12$$

Las superficies de respuesta que resultan por este modelo ajustado, pueden ser:

1. Sistema de loma estacionario: Es cuando el verdadero valor del punto óptimo está en cualquier lugar de la región de la superficie de respuesta.
2. Sistema de lomas ascendente: Este tipo de sistema de lomas no es posible hacer inferencia acerca de la superficie verdadera ó el punto estacionario. Dado que x_0 está fuera de la región donde se ha ajustado el modelo

La distancia al punto estacionario desde el centro del diseño, es:

$$d = \left(\sum_{i=1}^k x_{i0}^2 \right)^{1/2} \quad 7.13$$

donde: x_{i0} , $i=1, 2, \dots, k$

Para lomas ascendentes (d) suele ser mayor que la unidad. El mejor método es, continuar la exploración a lo largo de la dirección óptima. En estos casos, suele ser útil otra forma canónica, como se muestra a continuación:

$$\begin{aligned} \hat{y} &= \hat{\beta}_0 + \theta_1 W_1 + \theta_2 W_2 + \dots + \theta_k W_k + \lambda_1 W_1^2 + \lambda_2 W_2^2 + \dots + \lambda_k W_k^2 \\ &= \hat{\beta}_0 + W'\theta + W'\Lambda W \end{aligned} \quad 7.14$$

donde: $\theta = M'b$

$\Lambda =$ Diagonal $(\lambda_1, \lambda_2, \dots, \lambda_k)$ que determina el tipo de superficie ajustada

$\theta =$ Miden las pendientes de la superficie en el origen inicial $x=0$, en las direcciones de los ejes rotados (W_1, W_2, \dots, W_k) .

7.4 DISEÑOS EXPERIMENTALES PARA AJUSTAR SUPERFICIE DE RESPUESTAS.

El ajuste y análisis de superficie de respuesta se facilitan en gran medida con la elección apropiada de un diseño experimental. Cuando se seleccionan una RSM, algunas de las características de un diseño deseable, son:

1. Proporciona una distribución razonable de puntos de datos (y por tanto información) en toda la región de interés.
2. Permite investigar la idoneidad del modelo, incluyendo la falta de ajuste.
3. Permite realizar experimentos de bloques.
4. Permite la construcción secuencial de diseños de orden superior.
5. Proporciona una estimación de error interna.
6. No requiere un número grande de corridas.
7. No requiere demasiados niveles de las variables independientes.
8. Asegura simplicidad de los cálculos de los parámetros del modelo.

7.4.1 DISEÑOS PARA AJUSTAR EL MODELO DE PRIMER ORDEN.

El modelo de primer orden, es:

$$\hat{y} = \beta_0 + \sum_{i=1}^k \beta_i X_i + \varepsilon \quad 7.15$$

La cual solo se considera los efectos puros de las variables que intervienen en el experimento y la suma de productos cruzados o interacciones entre los factores se consideran igual a cero. Esta clase de diseños ortogonales de primer orden incluyen diseños, como:

1. Factoriales 2^k .
2. Fraccionales de la serie 2^k , donde los efectos principales no son alias entre si.
3. Factoriales 2^k con adición de puntos centrales.
4. Diseños simples.

7.4.2 DISEÑOS PARA AJUSTAR EL MODELO DE SEGUNDO ORDEN

un diseño experimental para ajustar un modelo de segundo orden debe tener por lo menos tres niveles cada factor. Existen muchos modelos que podrían emplearse para ajustar un modelo de segundo orden. Para un modelo de primer orden, la ortogonalidad es el principal propiedad del diseño óptimo ya que minimiza la variancia de la respuesta predicha.

Los modelos de primer orden con puntos centrales son diseños experimentales que son rotables ó girables si la variancia de la respuesta predicha \hat{y} en algún punto de (x) es función solo de la distancia al punto desde el centro del diseño y no en función de la dirección. Además, un diseño con esta propiedad dejará sin cambio la variancia de \hat{y} cuando el diseño se haga alrededor del centro; e aquí el nombre de diseño rotable.

La rotabilidad es una propiedad muy importante en la elección de un diseño de superficie de respuesta. Dado que la finalidad de la RSM es la optimización y se desconoce la localización del óptimo antes de realizar el experimento. Obsérvese que todos los diseños ortogonales de primer orden es rotable. En las Figura 7.5 se muestran las graficas para diseños de compuestos centrales para $k=2$ y $k=3$.

Además, en la tabla 7.2 se presentan un resumen de diseños de primer orden 2^k y fraccionarios de compuestos centrales girables ortogonales y de precisión uniforme.

Además, existen otros diseños, entre ellos se tiene:

1. diseño compuesto central en las caras: los puntos axiales se encuentran en el centro de la caras del cubo ($k=3$) y no son rotables, este diseño se muestra en la Figura 7.6.
2. Diseños de tres variables de Box-Behnken: Así, como se muestra en la Figura 7.7. Su diseño se muestra en la Tabla 7.3 y son casi rotables.

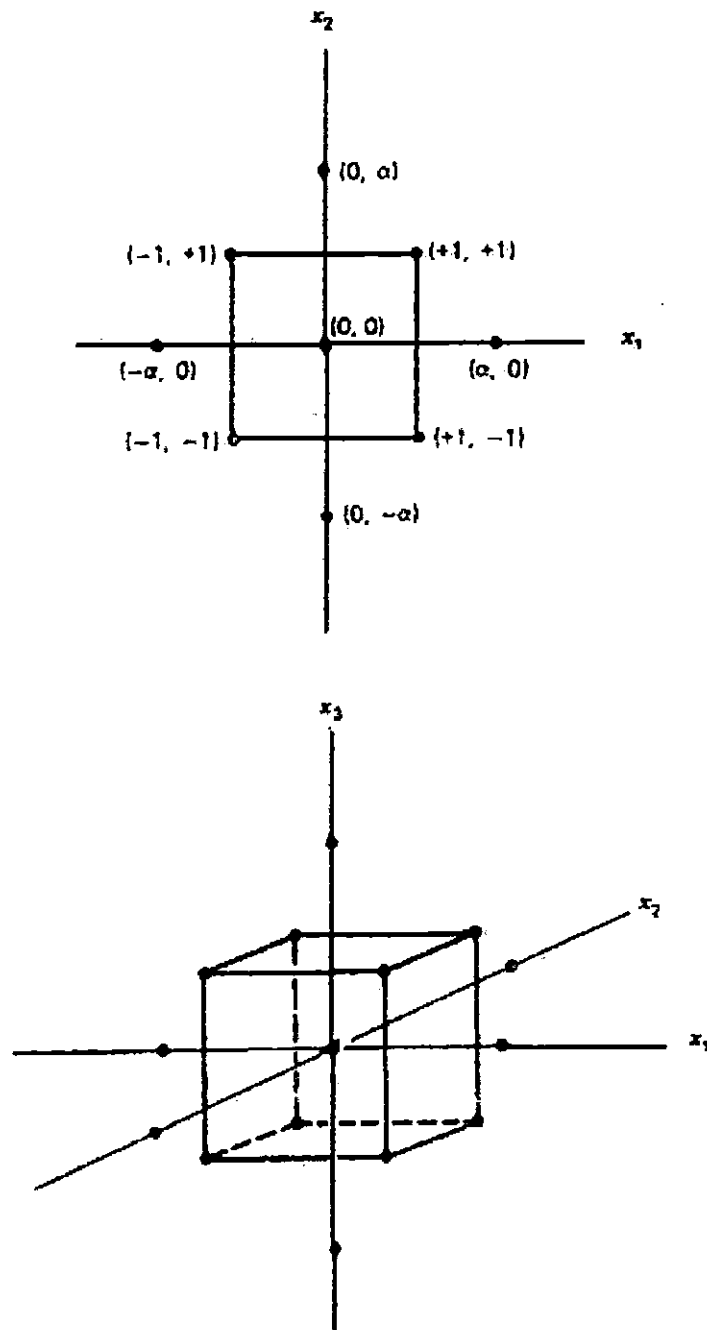


Figura 7.5. Diseños compuestos centrales para $k=2$ y $k=3$

7.5 FORMACION DE BLOQUES EN DISEÑOS DE SUPERFICIE DE RESPUESTA

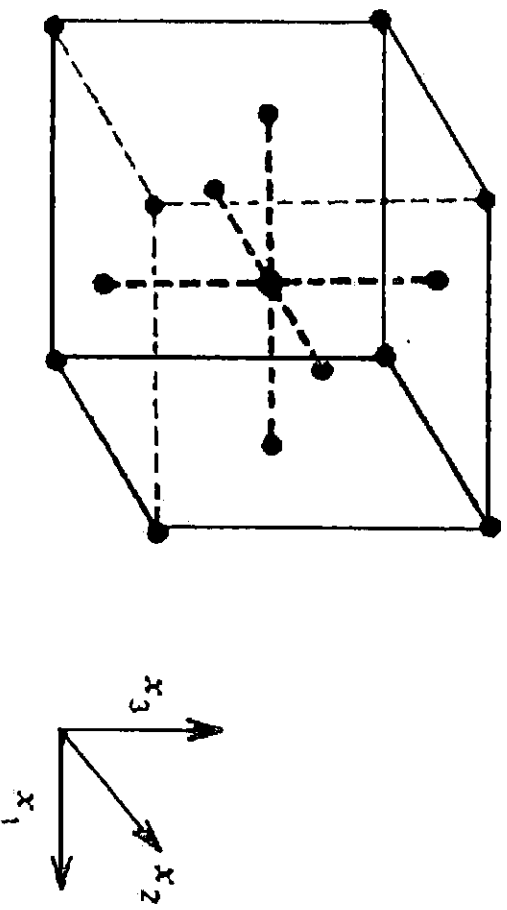
Los diseños de superficie de respuesta, con frecuencia se hace necesario considerar las formaciones de bloques a fin de eliminar efectos inconvenientes ó no manejables. Por ejemplo: no se puede reproducir un experimento en el cual secuencialmente se diseña un experimento de segundo orden después de realizarse el ensayo de un modelo de primer orden. Esto es, puede transcurrir un tiempo considerable entre la corrida de los experimentos de segundo orden, puede ser, que durante ese tiempo las condiciones de pruebas cambien. Con esto, se justifica la formación de bloques para corregir este posible efecto.

Tabla 7.2. Resumen de diseños compuestos centrales girables y de precisión uniforme.

k	2	3	4	5	5 ½ rep	6	6 ½ rep	7 ½ rep	8 ½ rep
n_i	4	8	16	32	16	64	32	64	128
Puntos axiales n_a	4	6	8	10	10	12	12	14	16
n_0 (P.U)	5	6	7	10	6	15	9	14	20
n_0 (ortg)	8	9	12	17	10	24	15	22	33
N (P.U)	13	20	31	52	32	91	53	92	164
N (ortg)	16	23	36	59	36	100	59	100	177
α	1.414	1.682	2.000	2.378	2.000	2.828	2.378	2.828	2.364

Tabla 7.3. Diseño de Box-Behnken de tres variables.

Corrida	X_1	X_2	X_3
1	-1	-1	0
2	-1	1	0
3	1	-1	0
4	1	1	0
5	-1	0	-1
6	-1	0	1
7	1	0	-1
8	1	0	1
9	0	-1	-1
10	0	-1	1
11	0	1	-1
12	0	1	1
13	0	0	0
14	0	0	0
15	0	0	0

Figura 7.6. Diseño compuesto centrado en las caras para $k=3$

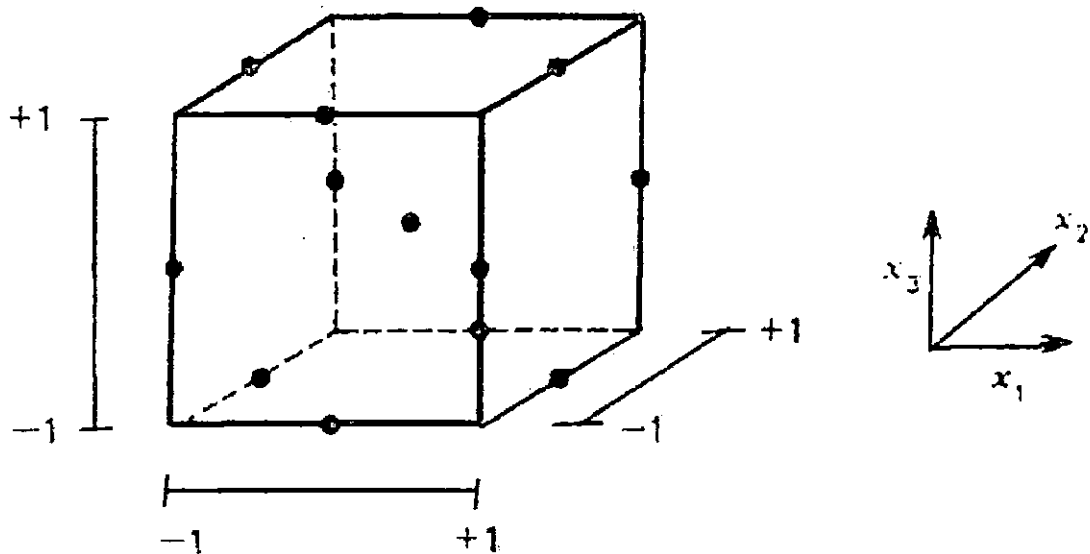


Figura 7.7. Diseño de Box-Behnken para tres factores

Para dividir en bloques ortogonales un diseño de segundo es necesario satisfacer dos condiciones. Cuando n_b observaciones en el b-ésimo bloque, las dos condiciones son:

1. Cada bloque debe ser un diseño ortogonal de primer orden, es decir:

$$\sum_{u=0}^{n_b} x_{iu} x_{ju} = 0 \quad i \neq j = 1, 2, \dots, k \quad \text{para toda } b \quad 7.16$$

donde: x_{iu} y x_{ju} son los niveles para la variable i -ésima y j -ésima corrida del experimento con $x_{0u}=1$ para toda u

2. De la suma total de cuadrados para cada variable, la fracción por cada bloque debe ser igual a la fracción de las observaciones totales que ocurre en el bloque; esto es:

$$\frac{\sum_{u=1}^{n_b} x_{iu}^2}{\sum_{u=1}^N x_{iu}^2} = \frac{n_b}{N} \quad i = 1, 2, \dots, k \quad \text{para toda } b \quad 7.17$$

donde: N es el número de corridas en el diseño

Para un diseño compuesto central siempre se puede construir por un modelo de primer orden ortogonal, quedando por primera condición: para el bloque 1: n_f puntos factoriales más n_{cf} puntos centrales, y el segundo consistirá: en $n_a=2*k$ puntos axiales más n_{ca} puntos centrales. Para la segunda condición, da como resultado:

$$\frac{\sum_u^{n_2} X_{iu}^2}{\sum_u^{n_1} X_{iu}^2} = \frac{n_a + n_{ca}}{n_f + n_{cf}} \quad 7.18$$

la división ortogonal se determina por la relación siguiente:

$$\alpha = \left[\frac{n_f(n_a + n_{ca})}{2(n_f + n_{cf})} \right]^{1/2} \quad 7.19$$

7.6 EJEMPLOS DEL CAPÍTULO

EJEMPLO 7.1

Una ingeniera química está interesada en determinar las condiciones de operación que maximizan el rendimiento de una reacción. Dos variables controlables influyen en este rendimiento: el tiempo y la temperatura de reacción. Actualmente ella opera el proceso con un tiempo cerca de 35 min y una temperatura de 155 °F. Esto produce un rendimiento de cerca de 40%. Ya que es poco probable que esta región contenga el óptimo decide ajustar un modelo de primer orden, aplicando el método de la máxima pendiente. Además elige la región de exploración para ajustar el modelo de primer orden de (30, 40) minutos de reacción y (150, 160 °F). Para simplificar los cálculos, las variables independientes se codifican al intervalo (-1, 1); quedando las variables codificadas de la siguiente manera (Montgomery, 1991):

Para la variable natural tiempo es:

$$X_1 = \frac{\varepsilon_1 - 35}{5}$$

y, para la variable natural temperatura es:

$$X_2 = \frac{\varepsilon_2 - 155}{5}$$

Los datos obtenidos del proceso por la ingeniera se muestran en la Tabla 7.4 a continuación:

Solución:

Aplicando el método de los mínimos cuadrados (capítulo 5), se obtiene un modelo matemático en función de las variables codificadas, este modelo es el siguiente:

$$\bar{y} = 40.44 + 0.775X_1 + 0.325X_2$$

Tabla 7.4. Datos del proceso para ajustar un modelo de primer orden

Variables naturales		Variables codificadas		Respuesta
ε_1	ε_2	X_1	X_2	y
30	150	-1	-1	39.3
30	160	-1	1	40.0
40	150	1	-1	40.9
40	160	1	1	41.5
35	155	0	0	40.3
35	155	0	0	40.5
35	155	0	0	40.7
35	155	0	0	40.2
35	155	0	0	40.6

Ahora es necesario explorar la idoneidad del modelo, este se resume en la Tabla 7.5. La cual consiste en un análisis de varianza.

Tabla 7.5. Resumen del análisis de varianza

Fuente de variación	Suma de cuadrados	G.L	Media de cuadrados	F_0
Regresión	2.8250	2	1.4125	47.83
Residuo	0.1772	6		
Interacción	0.0025	1	0.0025	0.058
Cuadrático puro	0.0027	1	0.0027	0.063
Error puro	0.1720	4	0.043	
Total	3.0022	8		

Donde:

La estimación del error es:

$$\sigma^2 = \frac{(40.3)^2 + (40.5)^2 + (40.7)^2 + (40.2)^2 + (40.6)^2 - (202.3)^2 / 5}{4} = 0.043$$

La suma de cuadrados para la interacción es:

$$SS_{INTERACCIÓN} = \frac{[(1 * 39.3) + (1 * 41.5) + (-1 * 40.0) + (-1 * 40.9)]^2}{4}$$

La suma de cuadrados de los cuadráticos puros con $\bar{y}_i = 40.425$ y $\bar{y}_c = 40.46$ es:

$$SS_{Cudraticapura} = \frac{n_f n_c (40.425 - 40.46)^2}{n_f + n_c} = 0.0027$$

Calcular el incremento a la trayectoria de máxima pendiente, de la siguiente manera:

Se decide incrementar el tiempo en 5 minutos, la cual equivale a un incremento de la variable codificada $\Delta X_1 = 1$ y $\Delta X_2 = (0.0325 / 0.775) \Delta X_1 = 0.42$ tal plan de trayectoria aparece en la Tabla 7.6 los incrementos se muestran tanto para variables codificadas (manejo matemático) como variables naturales (proceso).

Tabla 7.6. Experimento de máximo ascenso para el ejemplo 7.1

Incrementos	Variables codificadas		Variables naturales		Respuesta y
	X_1	X_2	ε_1	ε_2	
Origen 1	0	0	35	155	
Δ	1	0.42	5	2	
Origen + Δ	1	0.42	40	157	41.0
Origen + 2 Δ	2	0.84	45	159	42.9
Origen + 3 Δ	3	1.26	50	161	47.1
Origen + 4 Δ	4	1.68	55	163	49.7
Origen + 5 Δ	5	2.10	60	165	53.8
Origen + 6 Δ	6	2.52	65	167	59.9
Origen + 7 Δ	7	2.94	70	169	65.0
Origen + 8 Δ	8	3.36	75	171	70.4
Origen + 9 Δ	9	3.78	80	173	77.6
Origen + 10 Δ	10	4.20	85	175	80.3
Origen + 11 Δ	11	4.62	90	177	76.2
Origen + 12 Δ	12	5.04	95	181	75.1

Lo que significa que los alrededores del óptimo está cerca en el punto ($\varepsilon_1=85$, $\varepsilon_2=175$). Tal como se muestra en la figura 7.8.

Ahora se ajusta un nuevo modelo de primer orden en la región de exploración para ε_1 entre 80 y 90 minutos, y ε_2 entre 170 y 180; por lo tanto, las variables codificadas son:

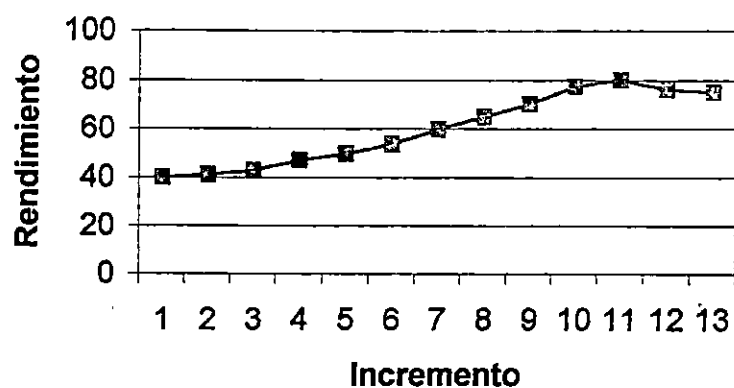


Figura 7.8. Rendimiento vrs. incremento de la trayectoria de la pendiente máxima

$$X_1 = \frac{\varepsilon_1 - 85}{5}$$

y

$$X_2 = \frac{\varepsilon_2 - 175}{5}$$

Los resultados obtenidos se presentan en la tabla 7.7.

Tabla 7.7. Datos para el segundo modelo de primer orden.

Variables naturales		Variables codificadas		Respuesta
ε_1	ε_2	X_1	X_2	y
80	170	-1	-1	76.5
80	180	-1	1	77.0
90	170	1	-1	78.0
90	180	1	1	79.5
85	175	0	0	79.9
85	175	0	0	80.3
85	175	0	0	80.0
85	175	0	0	79.7
85	175	0	0	79.8

El modelo de primer orden ajustado se:

$$\hat{y} = 78.97 + 1 * X_1 + 0.50 * X_2$$

Y su prueba de significancia se presenta en la tabla 7.8.

Tabla 7.8. Análisis de varianza para el segundo modelo de primer orden

Fuente de variación	Suma de cuadrados	G.L	Media de cuadrados	F ₀
Regresión	5.00	2		
Residuo	11.12	6		
Interacción	0.25	1	0.25	4.72
Cuadrático puro	10.658	1	10.658	201.09
Error puro	0.212	4	0.053	
Total	16.12	8		

El análisis de varianza para el modelo de primer orden presentado en la Tabla 7.8. las prueba de interacción y cuadrático puro son grandes, por lo cual las hipótesis nulas formuladas se rechazan y sus efectos son significativos en la variabilidad de los datos. Por tanto, el modelo de primer orden no es una aproximación adecuada. La curvatura en la superficie real puede indicar la proximidad al punto óptimo. En este caso debe hacerse un análisis adicional para localizar el óptimo con mayor precisión.

Ejemplo 7.2

Se continua el análisis de la reacción química del Ejemplo 7.1. no se puede ajustar un modelo de segundo orden para las variables X_1 y X_2 usando los datos de la tabla 7.7. La experimentadora decide aumentar estos datos con suficientes puntos para ajustar un modelo de segundo grado. Obtiene cuatro observaciones en $(X_1=0, X_2 \pm 1.414)$ y en $(X_1 = \pm 1.414, X_2)$. El conjunto completo de datos se muestra en la tabla 7.9 y el diseño se muestra en la Figura 7.9.

Un modelo de segundo orden se ajusta utilizando los mínimos cuadrados usando los datos codificados. Resultando:

$$\bar{y} = 79.9408 + 0.9949X_1 + 0.5151X_2 - 1.377X_1^2 - 1.0018X_2^2 + 0.25X_1X_2$$

Tabla 7.9. Diseño compuesto central para el ejemplo 7.2

Variables naturales		Variables codificadas		Respuesta
ε_1	ε_2	X_1	X_2	y
80	170	-1	-1	76.5
80	180	-1	1	77.0
90	170	1	-1	78.0
90	180	1	1	79.5
85	175	0	0	79.9
85	175	0	0	80.3
85	175	0	0	80.0
85	175	0	0	79.7
85	175	0	0	79.8
92.07	175	1.414	0	78.4
77.93	175	-1.414	0	75.6
85	182.07	0	1.414	78.5
85	167.93	0	-1.414	77.0

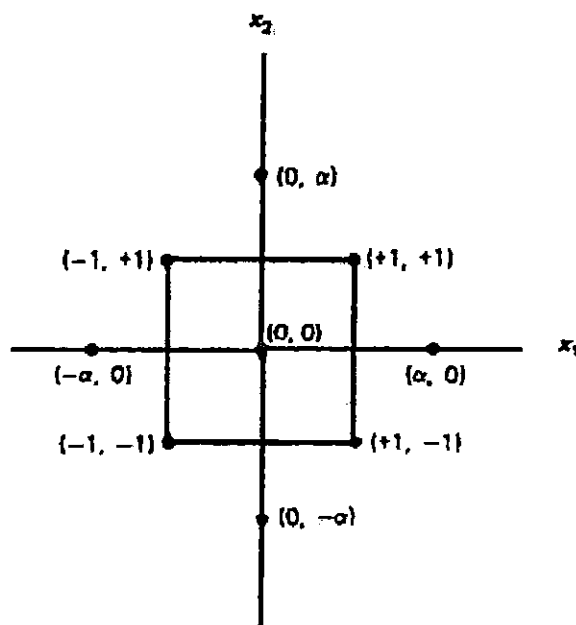


Figura 7.9. Diseño compuesto central para el ejemplo 7.2

La significancia del modelo se muestra en la tabla 7.10

Tabla 7.10. Análisis de varianza para el modelo de segundo orden

Fuente de variación	Suma de cuadrados	G.L.	Media de cuadrados	F ₀
Regresión	28.2560	5	5.6512	81.20 ^a
Error	0.4871	7	0.0696	---
Falta de ajuste	0.2751	3	0.0917	1.73
Error puro	0.212	4	0.0530	
Total	28.7431	12		

El análisis de varianza para este modelo presenta una falta de ajuste no significativa y la regresión sí lo es, por lo tanto el modelo de segundo orden se aproxima adecuadamente a la superficie real. Determinándose el punto estacionario de la siguiente forma:

$$b = \begin{vmatrix} 0.9949 \\ 0.5151 \end{vmatrix} \quad \beta = \begin{vmatrix} -1.377 & 0.125 \\ 0.125 & -1.0018 \end{vmatrix}$$

Aplicando a β determinante inversa se tiene:

$$X_0 = -\frac{1}{2} \begin{vmatrix} -0.7345 & -0.0917 \\ -0.0917 & -1.0096 \end{vmatrix} \begin{vmatrix} 0.9949 \\ 0.5151 \end{vmatrix} = \begin{vmatrix} 0.389 \\ 0.3056 \end{vmatrix}$$

Entonces:

$$X_{1,0} = 0.389$$

Y

$$X_{2,0} = 0.3056$$

Si:

$$0.389 = \frac{\varepsilon_1 - 85}{5}$$

y

$$0.3056 = \frac{\varepsilon_2 - 175}{5}$$

El punto estacionario es, cuando el tiempo toma el valor de $86.945 \approx 87$ minutos, y cuando la temperatura vale $176.528 \approx 176.5$ °F.

7.7 EJERCICIOS PROPUESTOS

7.1 Una planta química produce oxígeno mediante la licuefacción de aire y separándolo en sus gases componentes mediante destilación fraccionaria. La pureza del oxígeno es función de la temperatura del condensador principal y de la razón de la presión entre las columnas superior e inferior. Las condiciones actuales de operación son temperatura (ε_1) = -220 °C y la razón de presión (ε_2) = 1.2. Usando los siguientes datos determine la trayectoria de ascenso máximo.

ε_1	ε_2	Pureza
-225	1.1	82.8
-225	1.3	83.5
-215	1.1	84.7
-215	1.3	85.0
-220	1.2	84.1
-220	1.2	84.5
-220	1.2	83.9
-220	1.2	84.3

7.2 Los siguientes datos fueron recopilados por un ingeniero químico y corresponde al tiempo de filtración, X_1 es la temperatura y X_2 la presión. Ajuste un modelo de segundo orden u analice la superficie ajustada.

X_1	X_2	y
-1	-1	54
-1	1	45
1	-1	32
1	1	47
-1.414	0	50
1.414	0	53
0	-1.414	47
0	1.414	51
0	0	41
0	0	39
0	0	44
0	0	42
0	0	40

CONCLUSIONES

- 1 Las técnicas estadísticas aplicadas a las áreas de la Ingeniería Química e Ingeniería de Alimentos, que tienen mayor importancia y utilidad son: las técnicas de regresión lineal, los diseños de experimentos y el análisis de superficies de respuesta. 2 El uso de las herramientas estadísticas aplicadas a las investigaciones en las áreas de interés conlleva a la toma de decisiones más objetivas, confiables y eficientes.
- 3 Por medio de las técnicas estadísticas se realiza un análisis e interpretación de los datos que permite visualizar cuáles son las variables que tienen efecto significativo en un fenómeno investigado.
- 4 El análisis de Covarianza es una técnica estadística de gran importancia en las áreas de Ingeniería Química e Ingeniería de Alimentos para evaluar los efectos representados por las variables ruido y ayuda a tomar decisiones con mayor objetividad y criterio a considerar a esta ó estas en la investigación.
- 5 Los diseños experimentales se pueden realizar investigaciones preliminares con el fin de evaluar los factores que más afectan significativamente la variable respuesta. Y con ello eliminar las variables que tienen un efecto no significativo en ella, para luego realizar un diseño de la investigación más depurado y menos complejo.
- 6 El análisis de superficie de respuesta permite optimizar los niveles de los factores que intervienen en la variabilidad de la variable respuesta. Además, permite conocer cómo se comportan estas en los alrededores del punto óptimo.
- 7 Las técnicas estadísticas permiten comprobar la relación existente entre las variables independientes y dependiente. Permite con que el

investigador tenga un dominio adecuado del experimento, y por que no decir de un proceso en particular.

RECOMENDACIONES

- 1 Incluir dentro de la curricula de las áreas de Ingeniería Química y de Alimentos asignaturas de Estadística para la investigación y Control Estadístico de Calidad, de esta forma se volverá más efectivo el desarrollo de trabajos de graduación y otros trabajos de investigación en el que se vea involucrado el profesional dentro de su ámbito de trabajo.
- 2 Que los aspirantes a optar al título de Ingeniero Químico e Ingeniero de Alimentos, que realicen algún tipo de investigación hagan uso de las herramientas estadísticas para diseñar el experimento de su investigaciones y validar sus resultados.
- 3 Que se continúe con los estudios estadísticos especializados en posteriores trabajos de graduación en las áreas de interés, como por ejemplo: Control de calidad, monitoreo de procesos, etc.

BIBLIOGRAFÍA

1. Box, George E. P, Hunter, William G., Hunter, J. Estuart, **"Estadística para Investigadores. Introducción al Diseño de Experimentos, Análisis de Datos y Construcción de Modelos"**, Tercera Edición, Editorial Reverte, S. A., 1989, Madrid, España.
2. Cochran, Williams G, Gertrude M. Cox, **"Diseños de Experimentos"**, Segunda Edición, Trillas S.A de C.V, México, 1991.
3. García, Eduardo Enrique, Juan Rodolfo Ramírez Guzmán, **"Recuperación de Grasas Residuales de Tenerías y su uso Industrial"**, Trabajo de Graduación para optar al título de Ingeniero Químico, UES, El Salvador, 1994.
4. Hines, Williams W, Douglas C Montgomery, **"Probabilidad y Estadística para Ingeniería y Administración"**, Tercera Edición, Cuarta Reimpresión, CECSA, México, 1997.
5. Johnson, Richard A, **"Probabilidad y Estadística para Ingenieros de Miller y Freund"**, Quinta Edición, Printice-Hall Hispanoamericana S.A, México, 1997
6. Montgomery, Douglas C, **"Diseños y Análisis de Experimentos"**, Primera Edición, Grupo Editorial Iberoamericana, S.A de C.V, México, 1991.
7. Nuila de Mejía, Julia Amalia; Marcos Antonio Mejía Mejía, **"Manual de Diseños Experimentales, con aplicación a la Agricultura y Ganadería"**, Trabajo de investigación, San Salvador, mayo de 1990.

- 8 Romero Villafranca, Rafael, Luisa Rosa Zúñiga de Ramejo, **"Estadística (Proyecto de Innovación Educativa) Conceptos Básicos, Análisis de Regresión, Análisis de Varianza, Diseños de Experimentos y Procesos Estocásticos"**, Departamento de Estadística e Investigación operativa, Universidad Politécnica de Valencia, septiembre de 1993.
- 9 Tamayo y Tamayo, Mario, **"El Proceso de la Investigación Científica"**, 3ª Edición, Editorial Limusa S.A. de C.V, México D.F., 1999.
- 10 Sampieri Hernández, Roberto; Fernández Collado, Carlos y Lucio, Baptista Pilar, **" Metodología de la Investigación"**, 2ª Edición, McGraw-Hill, México, 1998.

ANEXOS

ANEXO A GLOSARIO DE TÉRMINOS

Análisis de dependencia Grupo de técnicas estadísticas caracterizadas por considerar una variable o un conjunto de ellas como *dependientes* y el resto como *independientes*. El objetivo es predecir la(s) variable(s) dependiente(s) a partir de la(s) independiente(s). Un ejemplo lo constituye el análisis de regresión.

Análisis de interdependencia Grupo de técnicas estadísticas en que las variables no se dividen en *dependientes* e *independientes* (por ejemplo, análisis factorial), sino que se consideran como un único conjunto.

Análisis de la varianza (ANOVA) Técnica estadística empleada para determinar si las muestras provienen de poblaciones con medias iguales. El análisis univariante de la varianza utiliza una variable dependiente, mientras que el análisis multivariante de la varianza compara muestras basadas en dos o más variables dependientes.

Análisis multivariante Análisis de varias variables en una única relación o un conjunto de relaciones.

Análisis univariante de la varianza (ANOVA) Técnica estadística que sirve para determinar, sobre la base de una medida dependiente, si las muestras provienen de poblaciones con igual media.

Coefficiente de correlación (r) Indica la fuerza de la asociación entre las variables independientes y la variable criterio. El signo (+ o -) indica la dirección de la relación. Puede tomar valores entre -1 y +1, con +1 indicando una relación positiva perfecta, 0 indicando la ausencia de relación y -1 indicando una relación inversa o negativa perfecta (a medida que una crece, otra disminuye).

Coefficiente de correlación parcial Medidas de la fuerza de la relación entre el criterio o variable criterio y una variable predictor simple donde los efectos de la otra variable predictor del modelo se mantienen constantes. Por ejemplo, r_{Y, X_2, X_1} , mide la variación en Y asociada con X_2 cuando el efecto de X_1 sobre X_2 e Y es constante. Usado en la estimación del modelo en el método de regresión para la selección secuencial de una variable, con el fin de identificar la variable independiente con el mayor poder predictivo *incrementado* por encima de las variables predictor ya presentes en el modelo.

Coefficiente de determinación (R^2) Medida de la proporción de la varianza de la variable criterio sobre su media, que es explicada por las variables independientes o predictor. El coeficiente puede variar entre 0 y 1. Si el modelo de regresión es estimado y aplicado apropiadamente, el investigador puede asumir que cuanto mayor

Coefficiente de determinación corregido (R^2 corregido)

Medida modificada del *coeficiente de determinación* que tiene en cuenta el número de variables predictor incluidas en la ecuación de regresión. Aunque la adición de variables a la regresión siempre provocará el aumento del coeficiente de determinación, el coeficiente de determinación corregido puede caer si las variables independientes añadidas tienen escaso poder explicativo y/o si los *grados de libertad* se vuelven demasiado pequeños. Este estadístico es muy útil para comparar entre ecuaciones con un número diferente de variables independientes, o tamaños muestrales diferentes, o ambos.

Coefficiente de regresión (b_n) Valor numérico del *parámetro* estimado directamente asociado con las variables independientes; por ejemplo, en el modelo $Y = b_0 + b_1X_1$, el valor b_1 es el coeficiente de regresión de la variable X_1 . En el modelo de regresión múltiple (es decir, $Y = b_0 + b_1X_1 + b_2X_2$), los coeficientes de regresión son parciales porque cada uno tiene en cuenta no sólo las relaciones entre Y y X_1 y entre Y y X_2 , sino también entre X_1 y X_2 . El coeficiente no está limitado en su rango, ya que está basado tanto en el grado de asociación como en las unidades de escala de la variable predictor. Por ejemplo, dos variables con la misma asociación a Y tendrían diferentes coeficientes si una variable predictor se midió en una escala sobre 7 puntos y otra basada en una escala de 100 puntos.

Constante (b_0) Valor del eje Y (eje de la variable criterio) donde la línea definida por la ecuación de regresión $Y = b_0 + b_1X_1$ corta el eje. Se describe por el término constante b_0 en la ecuación de regresión. Además de su papel en la predicción, la constante puede o no tener una interpretación práctica. La constante representa el cambio en la variable criterio en ausencia de la variable predictor. Por ejemplo, cuando estimamos las ventas respecto de los gastos previos de publicidad, la constante representa el nivel de ventas esperado si se elimina el efecto de la publicidad. Pero en muchos casos, la constante sólo tiene valor predictivo dado que no existe situación en la cual están ausentes todas las variables predictor. Un ejemplo es la predicción de la preferencia por el producto basándose en las actitudes del consumidor. Todos los individuos tienen una actitud, de tal forma que la constante no tiene un uso para la toma de decisiones pero ayuda a realizar una predicción más precisa.

Covarianzas o análisis de covarianzas Empleo de procedimientos similares a la regresión para eliminar la variación extraña (molesta) de las variables dependientes debida a una o más variables independientes métricas incontroladas (covariaciones). Se supone que las covariaciones están linealmente relacionadas con las variables dependientes. Después del ajuste debido a la influencia de las covariaciones, se realiza un ANOVA (o MANOVA) estándar. Este proceso de ajuste (conocido como ANCOVA y MANCOVA) generalmente permite el uso de contrastes más sensibles de los efectos de tratamientos.

- Curtosis** Medida del apuntamiento o llanura de una distribución cuando se compara con una distribución normal. Un valor positivo indica una distribución relativamente puntiaguda y un valor negativo indica una distribución relativamente plana.
- Chi-cuadrado** Método de estandarización de datos en una *tabla de contingencia* comparando la frecuencia de la celda efectiva con la frecuencia de la celda esperada. La frecuencia de la celda esperada se basa en las probabilidades marginales de su fila y columna (probabilidad de una fila y columna entre todas las filas y columnas).
- Datos métricos** También denominados *datos cuantitativos*, datos de *intervalos* o de *ratios*, estas medidas identifican o describen sujetos (u objetos) no sólo por la posesión de un atributo sino también por la cantidad o grado en que el sujeto está caracterizado por el atributo. Por ejemplo, la edad de una persona o su peso son medidas métricas.
- Datos no métricos** También denominados *datos cualitativos*, son atributos, características o propiedades categóricas que identifican o describen a un sujeto o a un objeto.
- Diseño** Conjunto específico de estímulos conjuntos creados para mostrar las propiedades estadísticas específicas de *ortogonalidad* y *equilibrio*.
- Diseño de experimentos** Plan de investigación en el cual el investigador manipula o controla directamente una o más variables predictoras (véase *tratamiento*) y valora sus efectos sobre las variables dependientes. Ya habitual en la ciencia física, está ganando popularidad en las ciencias económicas y sociales. Por ejemplo, a los encuestados se les muestra *anuncios* publicitarios distintos que varían sistemáticamente en una característica, tales como diferentes pretensiones (emocional frente a racional) o tipos.
- Diseño de puente** *Diseño* de estímulos para un gran número de *factores* (atributos) en los que los atributos se parten en varios grupos reducidos. Cada grupo de atributos tiene ciertos atributos que están contenidos en otros grupos, de tal forma que los resultados de cada grupo se pueden combinar, o construir puentes.
- Diseño equilibrado** *Diseño* de estímulos en los que en cada nivel dentro de un *factor* aparece un número igual de veces.
- Diseño factorial** Método de diseño de *estímulos* para evaluación mediante la generación de *todas* las posibles combinaciones de *niveles*. Por ejemplo, un análisis conjunto de tres factores con tres niveles por factor resultaría en 27 combinaciones que podrían actuar como estímulos.
- Diseño factorial fraccionado** Aproximación, como alternativa al *diseño factorial*, que utiliza sólo un subconjunto de posibles estímulos necesarios para estimar los resultados basándose en una regla de composición supuesta. Su tarea principal es reducir el número de las evaluaciones recogidas mientras se mantiene la *ortogonalidad* entre los *niveles* y subsecuentes estimaciones de los *componentes parciales de la utilidad total*. El diseño más simple es un *modelo aditivo*, en el cual sólo se estiman los *efectos principales*. Si los *términos de interacción* están incluidos, entonces se crean estímulos adicionales. El diseño puede ser creado tanto por la relación con las fuentes publicadas como utilizando los programas informáticos que acompañan a la mayoría de los paquetes de análisis conjunto.
- Diseño óptimo ortogonal** Restricción matemática que se exige para que las estimaciones de los *componentes parciales de la utilidad total* sean independientes unas de las otras. En el análisis conjunto, la ortogonalidad se refiere a la capacidad para medir el efecto de cambiar cada nivel de atributo y separarlo de los efectos del cambio sobre otros niveles de atributos y de un error experimental.
- Distribución normal** Una distribución de probabilidad continua puramente teórica en la que el eje horizontal representa todos los posibles valores de una variable y el eje vertical representa la probabilidad de que ocurran esos valores. Los valores de la variable están agrupados en torno a la media de forma simétrica y unimodal, conocida como la curva en forma de campana de Gauss o normal.
- Error de predicción** Diferencia entre los valores reales y de predicción de la variable criterio para cada observación en la muestra (véase *residuo*).
- Error de Tipo I** Probabilidad de rechazar la hipótesis nula siendo cierta, es decir, concluir que dos medias son significativamente diferentes cuando de hecho son iguales. Valores pequeños de alfa (por ejemplo, 0,05 ó 0,01), (véase *alfa*), llevan a que el rechazo de la hipótesis nula sea insostenible y a la aceptación de la hipótesis alternativa de que las medias poblacionales son distintas.
- Error de Tipo II** Probabilidad de no rechazar la hipótesis nula cuando es falsa rechazada, es decir, concluir que las dos medias no son significativamente diferentes cuando de hecho lo son. También conocido como error *beta* (β). En términos más sencillos, la posibilidad de no encontrar correlación o diferencia de medias cuando existen. Está inversamente relacionado con el *error de Tipo I*. El valor 1 menos el error de Tipo II se define como la *potencia*.
- Error estándar** Medida de la dispersión de las medias o de las diferencias de las medias esperada debido a la variación muestral. Denota la gama esperada del coeficiente a través de las muestras múltiples de los datos. Resulta de utilidad en pruebas estadísticas de significación para ver si el coeficiente es sustancialmente diferente de cero (es decir, si la gama esperada del coeficiente contiene el valor cero a un nivel de confianza específico). El valor *t* de un *coeficiente de regresión* es el coeficiente dividido por su error estándar.
- Estadístico t** Contraste estadístico que valora la significación estadística entre dos grupos en una única variable dependiente (véase *contraste t*).
- Estadístico U** Véase *lambda de Wilks*.
- Estimación de máxima verosimilitud (MLE)** Método de estimación habitualmente empleado en modelos de ecuaciones estructurales, incluyendo LISREL y EQS. Método alternativo a los mínimos cuadrados ordinarios utilizados en la regresión múltiple, MLE es un procedimiento que mejora iterativamente los parámetros estimados para minimizar una función de ajuste especificada.

- Factor (Análisis conjunto/MANOVA)** Variable independiente no-métrica, también denominada variable de experimento o de *tratamiento*.
- Factor (Análisis factorial)** Combinación lineal (valor teórico) de las variables originales. Los factores también representan las dimensiones subyacentes (construcciones) que resumen o justifican la serie original de variables en observación.
- Gráfico de cajas y bigotes** Método de representación de la distribución de la variable. Una caja representa la mayor parte de la distribución y las extensiones, llamadas bigotes, llegan al final de la distribución. Muy útil para realizar comparaciones de una sola variable a lo largo de grupos o entre varias variables.
- Gráfico de distribución normal** Comparación gráfica de la forma de la distribución respecto a la *distribución normal*. En el gráfico, la distribución normal se representa por una línea recta con un ángulo de 45 grados. La distribución efectiva se dibuja contra esta línea, de tal forma que cualquier diferencia se muestra como desviación de la línea recta, haciendo muy aparente e interpretable la identificación de estas diferencias.
- Gráficos de dispersión** Representación de las relaciones entre dos variables métricas que muestren los valores conjuntos de cada observación en un gráfico de dos dimensiones.
- Hipótesis nula** Hipótesis de que las muestras provienen de poblaciones con medias iguales para una variable dependiente (contraste univariante) o para un conjunto de variables dependientes (contraste multivariante). La hipótesis nula puede ser aceptada o rechazada dependiendo de los resultados del contraste de significación estadística.
- Histograma** Representación gráfica de la distribución de una única variable. Al formar el recuento de frecuencias en categorías, puede verse el perfil de la distribución de la variable. Utilizado para realizar una comparación visual con la *distribución normal*.
- Medida multivariante** Uso de dos o más variables como *indicadores* de una *medida compuesta* única. Por ejemplo, un contraste de personalidad puede suministrar las respuestas a una serie de preguntas individuales (*indicadores*), que se combinan para formar un único registro (*escala aditiva*) que representa un rasgo de la personalidad.
- Mínimos cuadrados** Procedimiento de estimación utilizado en la regresión simple y múltiple por lo que se estiman los coeficientes de regresión para aminorar la suma total de los *residuos* cuadrados.
- Modelo** Conjunto especificado de relaciones de dependencia que puede ser contrastado empíricamente —una operacionalización de una teoría. El propósito de un modelo es proporcionar concisamente una representación amplia de las relaciones a examinar. El modelo puede ser formalizado en un diagrama de secuencias o en un conjunto de relaciones estructurales.
- Ortogonalidad** Independencia estadística o una ausencia de relación. Los *valores teóricos* ortogonales explican la varianza única, sin ofrecer ninguna explicación de la varianza que es compartida por ellos. Los *contrastos* de ortogonalidad son *comparaciones planificadas* que son estadísticamente independientes y representan comparaciones únicas de las medias de los grupos.
- Parámetro** Cantidad (medida) característica de la población. Por ejemplo, (k) y (p) son los símbolos usados para los parámetros de la población media (k) y varianza (p). Habitualmente se estiman a partir de los datos de la muestra donde la media aritmética muestral se utiliza como medida de la media poblacional y la varianza muestral se usa para estimar la varianza de la población.
- Regresión múltiple** Modelo de regresión con dos o más variables independientes.
- Regresión simple** Modelo de regresión con una única variable independiente.
- Residuo** Medida del ajuste predictivo para una única observación. Se calcula como la diferencia entre los valores reales y predichos de la variable dependiente. Se supone que los residuos tienen una media de cero y una varianza constante. Juegan un papel importante no sólo en la determinación del cumplimiento de los supuestos subyacentes de regresión, sino también sirven como una herramienta de diagnóstico en la identificación de *casos atípicos* y *observaciones influyentes*.
- Simetría** Medida de la simetría de una distribución; en muchos casos la comparación se hace para una *distribución normal*. Una distribución simétrica positiva contiene pocos valores elevados y se inclina hacia la derecha, y una distribución simétrica negativa tiene pocos valores bajos y se inclina hacia la izquierda. Cuando los valores de simetría caen fuera del rango de -1 a $+1$ indican una distribución sustancialmente simétrica.
- Suma del cuadrado de la regresión (SSR)** Suma de las diferencias al cuadrado entre la media y los valores de la predicción de la variable independiente para todas las observaciones. Representa la cantidad de mejora en la explicación de la variable independiente atribuible a las variables independientes.
- Suma del cuadrado de los errores (SSE)** Suma de los errores de *predicción al cuadrado* (*residuos*) a lo largo de todas las observaciones. Usada para denominar a la varianza en las variables dependientes no tenidas todavía en cuenta en el modelo de regresión. Si no se utilizan variables independientes para la predicción, se convierte en los errores cuadrados usando la media como el valor predictor y por tanto iguala a la *suma total de los cuadrados*.
- Suma total de los cuadrados (TSS)** Cantidad total de variación que existe para ser explicada por las variables independientes. Esta «línea de fondo» se calcula sumando las diferencias al cuadrado entre la media y los valores efectivos de la variable criterio para todas las observaciones.

APÉNDICE B

DIAGNOSTICO DEL GRADO DE USO DE LA ESTADISTICA APLICADA A LA INVESTIGACIÓN.

El diagnóstico del grado de uso de la estadística aplicada a la investigación se realiza para obtener resultados que dará valiosa información, como:

- Las metodologías estadísticas mas conocidas y aplicadas en las investigaciones.
- La aplicación de metodologías estadísticas para la investigación se hace por norma.
- Los software empleados para procesar los datos obtenidos en la investigación y la importancia de esta herramienta.
- Si han sido capacitados en el área de estadística aplicada a la investigación y su importancia.

A través de los resultados del diagnóstico se podrá asegurar que las investigaciones se realizan en forma correcta, objetivas y confiables. Cumpliendo con ello la validez interna y externa de la investigación; justificando con ello la importancia de la realización del presente trabajo investigativo sobre la estadística aplicada a la investigación en el área de Ingeniería Química e Ingeniería de Alimentos.

2.1 LUGARES DONDE SE REALIZO EL DIAGNOSTICO

Los lugares seleccionados para realizar el diagnóstico del grado de uso de la estadística aplicada a la investigación, se realizó que se cumpliera con los siguientes requisitos:

1. Accesibilidad para obtener la información, según la política interna y externa del lugar (con la disponibilidad de colaborar con la investigación).
2. Accesibilidad geográfica del lugar (área de San Salvador).
3. Lugares que se realicen investigaciones (científicas, tecnológicas y económicas) afines al área de Ingeniería Química e Ingeniería de Alimentos.

Los lugares seleccionados fueron las Escuelas de diferentes Facultades dentro de la Universidad Nacional de El Salvador , Escuela De Ingeniería Química de la Universidad José Simeón Cañas (UCA), Unidad Investigación y Desarrollo de la Universidad Don Bosco y PROCAFE. Las Escuelas y Facultades seleccionadas dentro de la Universidad Nacional de El Salvador fueron:

1. Escuela de Ingeniería Química e Alimentos.
2. Facultad de Ingeniería Agronómica.
3. Facultad de Química y Farmacia.
4. Escuela de Biología.
5. Escuela de Química.
6. Escuela de Física.
7. Facultad de Medicina (Unidad de Patología).
8. Escuela de Economía.

2.2 COMO SE REALIZO EL DIAGNOSTICO

El diagnóstico se realizó en los lugares seleccionados por medio de un cuestionario (Encuesta) diseñado con el objetivo de extraer la información (antes descrita).

2.3 DISEÑO DE LA ENCUESTA

Para el diseño de la encuesta se tiene como objetivo obtener la mayor información posible de interés y que sea valiosa para el fin que se percibe. Para su diseño se consideraran cuatro factores de interés, como:

1. Las metodologías más aplicadas en la investigación.
2. Si se aplican por norma o no.
3. Los software utilizados para procesar los datos de las investigaciones y su importancia.
4. Si los investigadores están capacitados en el área de estadística o no.

En base a estos cuatro factores críticos de interés se realiza el diseño de la encuesta mostrada en la sección 2.4.

2.4 ELABORACION DE ENCUESTA

Tomando en cuentas los cuatro factores críticos determinados con anterioridad pasa a elaborar la encuesta en la cual se diagnosticará el grado de uso de estadística aplicada a la investigación en los lugares antes seleccionados. La encuesta obtenida se mostrará a continuación, Anexo A.

Ejemplar de la encuesta diseñada para diagnosticar el grado de uso de la estadística aplicada a la investigación. Ejemplo, Se muestra en el Apéndice J.

2.5 RESULTADOS DEL DIAGNOSTICO

Los resultados obtenidos en las encuestas según las fuentes de investigación y desarrollo seleccionados, fueron:

Nota: Las respuestas se clasificarán por pregunta de la encuesta, según su importancia.

1. Que tipos de investigaciones realizan en el departamento.

Económica Social Tecnológica Científica

Tipo de investigación	Frecuencia
Económica	1
Social	1
Tecnológica	4
Científica	7

3. Cuales de las siguientes metodologías estadísticas conoce y cuales ha aplicado en sus investigaciones.

	<i>Metodología</i>					
Respuesta	Prueba de hipótesis	Diseño de experimento	Análisis de regresión	Series temporales	Técnica de muestreo	Análisis de multivariabl e
Conoce	9	9	5	1	6	3
Aplicado	9	7	5	1	4	2

Otros :

- Comparación del fenómeno teórico y experimental (simulación matemática).

5. El uso de las metodologías estadísticas que usted utiliza, lo hace por norma:

SI NO

6. Han recibido capacitación en metodologías estadísticas para la investigación

SI NO

Nota: Como las respuestas son estrictamente SI o NO se presentará en la misma tabla de respuestas.

Respuesta	Pregunta No. 5	Pregunta No. 6
SI	8	7
NO	1	2

7. En que grado es necesario la capacitación en metodologías estadísticas

Ninguno Bajo Medio Alto

8. Cual es grado de importancia que usted considera que tiene la aplicación de metodologías estadísticas a la investigación.

Ninguno Bajo Medio Alto

Nota: Las respuestas de la preguntas 7 y 8 tienen la misma clasificación, los resultados se presentará en la misma tabla de respuesta.

Clasificación	Pregunta No. 7	Pregunta No. 8
Ninguno	-	-
Bajo	-	-
Medio	-	-
Alto	9	9

9. Aplican algún software para la realizar el análisis estadístico de los datos.

SI

NO

Clasificación	Respuesta
SI	8
NO	1

Cuales:

- Statistical Analysis System
- Mstat (2)
- Risk Analysis
- Excel (2)
- Visual Fox
- SX
- Systat (2)
- EPI-INFO (OMS/OPS) (2)
- ULC
- Origin
- SAS
- SPS
- Stat Graphic
- Harvard Graphic
- Math Cad

Nota: Los software que aparece entre paréntesis el numero dos, se debe que el software es utilizado en dos centros de investigación y desarrollo.

10. En que grado cree que es necesario utilizar software para el análisis estadístico de datos:

Ninguno Bajo Medio Alto

Clasificación	Respuesta
Ninguno	-
Bajo	-
Medio	5
Alto	4

COMENTARIOS:

- Lo importante es que el investigador sepa la importancia y el valor que tiene el análisis estadístico en las investigaciones para garantizar la calidad, validez interna y externa de la investigación y buscar apoyo en otros profesionales.

2.6 IMPORTANCIA DEL DIAGNOSTICO

La importancia de realizar el diagnóstico a los lugares de investigación y desarrollo seleccionados, fue para obtener valiosa información referente el uso de las metodologías estadísticas aplicadas y sus herramientas e importancia de su uso en las investigaciones, como:

- Metodologías estadísticas más aplicadas
- Si las aplican por norma
- La aplicación de software como herramienta en procesamiento de datos.
- Si los investigadores son diestro en metodologías aplicadas a la investigación.

Con estos parámetros se puede tener una idea de cuales son las condiciones reales en el campo de investigación y por tal razón de realizar el diagnóstico. Partiendo de los resultados del diagnóstico se puede dar énfasis aun mas las metodologías estadísticas que se aplique en las áreas investigación de los lugares seleccionados, lo cual realizan sus investigaciones objetivas, confiables, con resultados apegados a la realidad y garantizar la validez externa como interna de las investigaciones.

2.7 OBSERVACIONES

1. De los centros de investigación y desarrollo seleccionados para la realizar el diagnóstico del grado de uso de estadística aplicada a la investigación algunos se negaron a colaborar, siendo estos: la Escuela de Ingeniería Química de la Universidad Centroamericana José Simeón Cañas (UCA) y la Universidad Don Bosco, por la cual de los once lugares de investigación y desarrollo solo se obtuvo respuestas de las restantes nueve lugares seleccionados.
2. Los software especializados en el procesamientos de datos tienen gran uso e importancia en las investigaciones realizadas. Y existen una gran gama de software en el mercado aplicados a áreas especializadas específicamente a los campos de investigación seleccionados. Aunque algunos son de uso generalizados.

2.8 RECOMENDACIONES

1. Se debería realizar un estudio a fondo para diagnosticar las deficiencias en el campo de la investigación en los lugares donde se realicen investigación. Con estos lograr visualizar sus debilidades existentes para luego preparar cursos para su adiestramiento en estadística aplicada a la investigación y los

software necesario para obtener investigaciones mas precisas, funcionales y menos tediosas. Apegándose a las necesidades reales existentes y disponer de las herramientas disponibles para superar la tediosa tarea de investigación y que se lleven a cabo con el mayor éxito posible.

2. La Universidad de El Salvador en estadística haga mayor énfasis en el área de investigación, siguiendo la metodología científica y sustentar los resultados obtenidos en las investigaciones sean sustentados por métodos estadísticos aplicados a ella. Con esto capacitando a los estudiantes a nivel de pregrado en las técnicas estadísticas aplicadas a la investigación, para garantizar que las investigaciones tengan validez interna como externa, realizándola en forma transparentes, objetiva, confiables, seguras y apegadas a la realidad del fenómeno en estudio.
3. Las metodologías estadísticas más empleadas y conocidas en el campo de la investigación seleccionados, son: planteamiento de hipótesis, diseño de investigación, regresión lineal y técnicas de muestreo, y las de menor uso es le análisis de series temporales.

**Ejemplar de la encuesta utilizada para realizar el diagnostico y uso de
técnica estadísticas aplicadas a la investigación.**

Prueba para Diagnosticar el uso de Técnicas Estadística a la Investigación

Gracias!. La siguiente es una encuesta para diagnosticar el grado de uso de la estadística aplicada en sus investigaciones.

NOTA: La prueba se aplicará en departamentos donde se realicen ó hallan realizado investigaciones.

1. Que tipos de investigaciones realizan en el departamento.

Económica Social Tecnológica Científica

2. Mencione las investigaciones de mayor relevancia que halla realizado con su equipo de trabajo.

1. _____
2. _____
3. _____

3. Cuales de las siguientes metodologías estadísticas conoce y cuales ha aplicado en sus investigaciones.

Metodologías	Conoce	Ha aplicado
Prueba de Hipótesis		
Diseño de Experimentos		
Análisis de Regresión		
Análisis de Series Temporales		
Técnicas de muestreo		
Análisis de Multivariable		

Otros _____

4. Si ha utilizado alguna, favor contestar, si no continuar con el numeral 6.
5. El uso de las metodologías estadísticas que usted utiliza, lo hace por norma:
SI NO
6. Han recibido capacitación en metodologías estadísticas para la investigación
SI NO
7. En que grado considera necesaria la capacitación en metodologías estadísticas
Ninguno Bajo Medio Alto
8. Cual es grado de importancia que usted considera que tiene la aplicación de metodologías estadísticas a la investigación.
Ninguno Bajo Medio Alto
9. Aplican algún software para la realizar el análisis estadístico de los datos.
SI Cuales: _____

NO
10. En que grado cree que es necesario utilizar software para el análisis estadístico de datos:
Ninguno Bajo Medio Alto

ANEXO C. Distribución normal acumulativa estándar

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

z	0.00	0.01	0.02	0.03	0.04	Z
.0	.500 00	.503 99	.507 98	.511 97	.515 95	.0
.1	.539 83	.543 79	.547 76	.551 72	.555 67	.1
.2	.579 26	.583 17	.587 06	.590 95	.594 83	.2
.3	.617 91	.621 72	.625 51	.629 30	.633 07	.3
.4	.655 42	.659 10	.662 76	.666 40	.670 03	.4
.5	.691 46	.694 97	.698 47	.701 94	.705 40	.5
.6	.725 75	.729 07	.732 37	.735 65	.738 91	.6
.7	.758 03	.761 15	.764 24	.767 30	.770 35	.7
.8	.788 14	.791 03	.793 89	.796 73	.799 54	.8
.9	.815 94	.818 59	.821 21	.823 81	.826 39	.9
1.0	.841 34	.843 75	.846 13	.848 49	.850 83	1.0
1.1	.864 33	.866 50	.868 64	.870 76	.872 85	1.1
1.2	.884 93	.886 86	.888 77	.890 65	.892 51	1.2
1.3	.903 20	.904 90	.906 58	.908 24	.909 88	1.3
1.4	.919 24	.920 73	.922 19	.923 64	.925 06	1.4
1.5	.933 19	.934 48	.935 74	.936 99	.938 22	1.5
1.6	.945 20	.946 30	.947 38	.948 45	.949 50	1.6
1.7	.955 43	.956 37	.957 28	.958 18	.959 07	1.7
1.8	.964 07	.964 85	.965 62	.966 37	.967 11	1.8
1.9	.971 28	.971 93	.972 57	.973 20	.973 81	1.9
2.0	.977 25	.977 78	.978 31	.978 82	.979 32	2.0
2.1	.982 14	.982 57	.983 00	.983 41	.983 82	2.1
2.2	.986 10	.986 45	.986 79	.987 13	.987 45	2.2
2.3	.989 28	.989 56	.989 83	.990 10	.990 36	2.3
2.4	.991 80	.992 02	.992 24	.992 45	.992 66	2.4
2.5	.993 79	.993 96	.994 13	.994 30	.994 46	2.5
2.6	.995 34	.995 47	.995 60	.995 73	.995 85	2.6
2.7	.996 53	.996 64	.996 74	.996 83	.996 93	2.7
2.8	.997 44	.997 52	.997 60	.997 67	.997 74	2.8
2.9	.998 13	.998 19	.998 25	.998 31	.998 36	2.9
3.0	.998 65	.998 69	.998 74	.998 78	.998 82	3.0
3.1	.999 03	.999 06	.999 10	.999 13	.999 16	3.1
3.2	.999 31	.999 34	.999 36	.999 38	.999 40	3.2
3.3	.999 52	.999 53	.999 55	.999 57	.999 58	3.3
3.4	.999 66	.999 68	.999 69	.999 70	.999 71	3.4
3.5	.999 77	.999 78	.999 78	.999 79	.999 80	3.5
3.6	.999 84	.999 85	.999 85	.999 86	.999 86	3.6
3.7	.999 89	.999 90	.999 90	.999 90	.999 91	3.7
3.8	.999 93	.999 93	.999 93	.999 94	.999 94	3.8
3.9	.999 95	.999 95	.999 96	.999 96	.999 96	3.9

Distribución normal acumulativa estándar (continuación)

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sigma\sqrt{2\pi}} e^{-u^2/2} du$$

z	0.05	0.06	0.07	0.08	0.09	Z
0.0	.519 94	.523 92	.527 90	.531 88	.535 86	0
0.1	.559 62	.563 56	.567 49	.571 42	.575 34	0.1
0.2	.598 71	.602 57	.606 42	.610 26	.614 09	0.2
0.3	.636 83	.640 58	.644 31	.648 03	.651 73	0.3
0.4	.673 64	.677 24	.680 82	.684 38	.687 93	0.4
0.5	.708 84	.712 26	.715 66	.719 04	.722 40	0.5
0.6	.742 15	.745 37	.748 57	.751 75	.754 90	0.6
0.7	.773 37	.776 37	.779 35	.782 30	.785 23	0.7
0.8	.802 34	.805 10	.807 85	.810 57	.813 27	0.8
0.9	.828 94	.831 47	.833 97	.836 46	.838 91	0.9
1.0	.853 14	.855 43	.857 69	.859 93	.862 14	1.0
1.1	.874 93	.876 97	.879 00	.881 00	.882 97	1.1
1.2	.894 35	.896 16	.897 96	.899 73	.901 47	1.2
1.3	.911 49	.913 08	.914 65	.916 21	.917 73	1.3
1.4	.926 47	.927 85	.929 22	.930 56	.931 89	1.4
1.5	.939 43	.940 62	.941 79	.942 95	.944 08	1.5
1.6	.950 53	.951 54	.952 54	.953 52	.954 48	1.6
1.7	.959 94	.960 80	.961 64	.962 46	.963 27	1.7
1.8	.967 84	.968 56	.969 26	.969 95	.970 62	1.8
1.9	.974 41	.975 00	.975 58	.976 15	.976 70	1.9
2.0	.979 82	.98 03	.980 77	.981 24	.981 69	2.0
2.1	.984 22	.984 61	.985 00	.985 37	.985 74	2.1
2.2	.987 78	.988 09	.988 40	.988 70	.988 99	2.2
2.3	.990 61	.990 86	.991 11	.991 34	.991 58	2.3
2.4	.992 86	.993 05	.993 24	.993 43	.993 61	2.4
2.5	.994 61	.994 77	.994 92	.995 06	.995 20	2.5
2.6	.995 98	.996 09	.996 21	.996 32	.996 43	2.6
2.7	.997 02	.997 11	.997 20	.997 28	.997 36	2.7
2.8	.997 81	.997 88	.997 95	.998 01	.998 07	2.8
2.9	.998 41	.998 46	.998 51	.998 56	.998 61	2.9
3.0	.998 86	.998 89	.998 93	.998 97	.999 00	3.0
3.1	.999 18	.999 21	.999 24	.999 26	.999 29	3.1
3.2	.999 42	.999 44	.999 46	.999 48	.999 50	3.2
3.3	.999 60	.999 61	.999 62	.999 64	.999 65	3.3
3.4	.999 72	.999 73	.999 74	.999 75	.999 76	3.4
3.5	.999 81	.999 81	.999 82	.999 83	.999 83	3.5
3.6	.999 87	.999 87	.999 88	.999 88	.999 89	3.6
3.7	.999 91	.999 92	.999 92	.999 92	.999 92	3.7
3.8	.999 94	.999 94	.999 95	.999 95	.999 95	3.8
3.9	.999 96	.999 96	.999 96	.999 97	.999 97	3.9

ANEXO D. Puntos porcentuales de la distribución χ^2

V	0.995	0.990	0.975	0.950	0.900	0.500	0.100	0.050	0.025	0.010	0.005
1	0.00+	0.00+	0.00+	0.00+	0.02	0.45	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	1.39	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.56	2.37	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	6.36	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.38	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.98	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72

Puntos porcentuales de la distribución^a χ^2 (continuación)

v	0,995	0,990	0,975	0,950	0,900	0,500	0,100	0,050	0,025	0,010	0,005
18	6,26	7,01	8,23	9,39	10,87	17,34	25,99	28,87	31,53	34,81	37,16
19	6,84	7,63	8,91	10,12	11,65	18,34	27,20	30,14	32,85	36,19	38,58
20	7,43	8,26	9,59	10,85	12,44	19,34	28,41	31,41	34,17	37,57	40,00
21	8,03	8,90	10,28	11,59	13,24	20,34	29,62	32,67	35,48	38,93	41,40
22	8,64	9,54	10,98	12,24	14,04	21,34	30,81	33,92	36,78	40,29	42,80
23	9,26	10,20	11,69	13,09	14,85	22,34	32,01	35,17	38,08	41,64	44,18
24	9,89	10,88	12,40	13,85	15,66	23,34	33,20	36,42	39,36	42,98	45,56
25	10,52	11,52	13,12	14,61	16,47	24,34	34,28	37,65	40,65	44,31	46,93
26	11,16	12,20	13,84	15,38	17,29	25,34	35,56	38,89	41,92	45,64	48,29
27	11,81	12,88	14,57	16,15	18,11	26,34	36,74	40,11	43,19	46,96	49,65
28	12,46	13,57	15,31	16,93	18,94	27,34	37,92	41,34	44,46	48,28	50,99
29	13,12	14,26	16,05	17,71	19,77	28,34	39,09	42,56	45,72	49,59	52,34
30	13,79	14,95	16,79	18,49	20,60	29,34	40,26	43,77	46,98	50,89	53,67
40	20,71	22,16	24,43	26,51	29,05	39,34	51,81	55,76	59,34	63,69	66,77
50	27,99	29,71	32,38	34,76	37,69	49,33	63,17	67,50	71,42	76,15	79,49
60	35,53	37,48	40,48	43,19	46,48	59,33	74,40	79,40	83,30	88,38	91,95
70	43,28	45,44	48,78	51,74	55,33	69,33	85,53	90,53	95,02	100,42	104,22
80	51,17	53,54	57,15	60,39	64,28	79,33	96,58	101,88	106,63	112,33	116,32
90	59,20	61,75	65,65	69,13	73,29	89,33	107,57	113,57	118,14	124,12	128,90
100	67,33	70,06	74,22	77,93	82,83	99,33	118,50	124,34	129,56	135,81	140,17

^av=grados de libertad

ANEXO E. Puntos porcentuales de la distribución t

	0.4000	0.2500	0.1000	0.0500	0.0250	0.0100	0.0050	0.0025	0.0010	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.320	318.310	636.620
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
x	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

ANEXO F Puntos porcentuales de la distribución F

F_{25, v1, v2}

v ₁	v ₂	Grados de libertad para el numerador (v ₁)																							
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞					
1	1	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.28	9.32	9.41	9.49	9.56	9.63	9.67	9.71	9.76	9.80	9.85					
2	1	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.41	3.43	3.43	3.44	3.45	3.46	3.47	3.48					
3	1	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.46	2.46	2.47	2.47	2.47	2.47	2.47	2.48					
4	1	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08					
5	1	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.88	1.88	1.88	1.88	1.87	1.87	1.87					
6	1	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.77	1.78	1.76	1.75	1.75	1.74	1.74	1.74	1.74					
7	1	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.70	1.69	1.68	1.68	1.67	1.67	1.66	1.66	1.65	1.65	1.65					
8	1	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.63	1.62	1.62	1.61	1.60	1.60	1.59	1.59	1.59	1.59					
9	1	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.54	1.53	1.53					
10	1	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.51	1.50	1.49	1.48					
11	1	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.48	1.47	1.46	1.45	1.45					
12	1	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.45	1.44	1.43	1.42					
13	1	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.40	1.40					
14	1	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.41	1.40	1.39	1.38					
15	1	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.35					
16	1	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34					
17	1	1.42	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34					
18	1	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32					
19	1	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.40	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32					
20	1	1.40	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31	1.29					
21	1	1.40	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32	1.31	1.29					
22	1	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32	1.31	1.30	1.28					
23	1	1.39	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32	1.31	1.30	1.28					
24	1	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.36	1.35	1.33	1.32	1.31	1.30	1.29	1.28	1.26					
25	1	1.39	1.47	1.46	1.44	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32	1.31	1.29	1.28	1.25					
26	1	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.35	1.34	1.32	1.31	1.30	1.29	1.28	1.26	1.25					
27	1	1.38	1.46	1.45	1.43	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.33	1.32	1.31	1.30	1.29	1.28	1.26	1.24					
28	1	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.31	1.30	1.29	1.28	1.25	1.24					
29	1	1.38	1.45	1.45	1.43	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.32	1.31	1.30	1.29	1.27	1.26	1.25	1.23					
30	1	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.34	1.32	1.30	1.29	1.28	1.27	1.26	1.24	1.23					
40	1	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.31	1.30	1.28	1.26	1.25	1.24	1.22	1.21	1.19					
60	1	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.27	1.25	1.24	1.22	1.21	1.19	1.17	1.15					
120	1	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.26	1.24	1.22	1.21	1.19	1.18	1.16	1.13	1.10					
∞	1	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.22	1.21	1.19	1.18	1.14	1.12	1.08	1.00					

Fuente: Adaptada con autorización de *Biometrika Tables for Statisticians*, Vol. 1, 3a edición, por E. S. Pearson y H. O. Hartley, Cambridge University Press, Cambridge, 1966.

Puntos porcentuales de la distribución F (continuación)

F_{10, v1, v2}

v ₁ \ v ₂	Grados de libertad para el numerador (v ₁)																				∞
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120			
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33		
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49		
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13		
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76		
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10		
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72		
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47		
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29		
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16		
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06		
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97		
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90		
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85		
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80		
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76		
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72		
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69		
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66		
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63		
20	2.97	2.59	2.39	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61		
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59		
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57		
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55		
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53		
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52		
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50		
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49		
28	2.89	2.50	2.29	2.15	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48		
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47		
30	2.88	2.49	2.28	2.14	2.03	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46		
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38		
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29		
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19		
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00		

Grados de libertad para el denominador (v₂)

Puntos porcentuales de la distribución F (continuación)

F_{05, v1, v2}

v1	Grados de libertad para el numerador (v1)												v2						
	1	2	3	4	5	6	7	8	9	10	12	15		20	24	30	40	60	120
1	161.4	199.5	215.7	224.6	230.2	234.0	236.9	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.18	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.39	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.53	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Grados de libertad para el denominador (v2)



Puntos porcentuales de la distribución F (continuación)

F_{0.05, v1, v2}

v ₁	Grados de libertad para el numerador (v ₁)																				v ₂
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞		
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018		
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50		
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90		
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.58	8.51	8.46	8.41	8.36	8.31	8.26		
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02		
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85		
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14		
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67		
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33		
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08		
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88		
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72		
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60		
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49		
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.58	2.52	2.46	2.40		
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32		
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25		
18	5.98	4.58	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.57	2.51	2.44	2.38	2.32	2.26	2.19		
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13		
20	5.87	4.46	3.88	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09		
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04		
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00		
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97		
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94		
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91		
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88		
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85		
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83		
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81		
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79		
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64		
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.46		
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31		
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00		

Puntos porcentuales de la distribución F (continuación)

F_{01, v1, v2}

v ₁	v ₂	Grados de libertad para el numerador (v ₁)																			
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	
1	4052	4959.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6208	6235	6261	6287	6313	6339	6366		
2	96.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50		
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.00	26.50	26.41	26.32	26.22	26.13		
4	21.20	18.00	16.69	15.99	15.52	15.21	14.96	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46		
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02		
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88		
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65		
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.88		
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31		
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91		
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60		
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36		
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17		
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00		
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87		
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75		
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65		
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57		
19	8.18	5.90	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.50		
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42		
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36		
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31		
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26		
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21		
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17		
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13		
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10		
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06		
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03		
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01		
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80		
60	7.08	4.96	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60		
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38		
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00		

Grados de libertad para el denominador (v₂)

ANEXO G Distribución acumulativa de Poisson*

x	c = λt				
	.05	.10	.20	.30	.40
0	.990	.951	.904	.818	.740
1	.999	.998	.995	.982	.963
2		.999	.998	.996	.992
3			.999	.999	.998
4				.999	.999
5					.999
x	c = λt				
	.70	.80	.90	1.00	1.10
0	.496	.449	.406	.367	.332
1	.844	.808	.772	.735	.699
2	.965	.952	.937	.919	.900
3	.994	.990	.986	.981	.974
4	.999	.998	.997	.996	.994
5	.999	.999	.999	.999	.999
6		.999	.999	.999	.999
7			.999	.999	.999
8				.999	.999
x	c = λt				
	1.50	1.60	1.70	1.80	1.90
0	.223	.201	.182	.165	.149
1	.557	.524	.493	.462	.433
2	.808	.783	.757	.730	.703
3	.934	.921	.906	.891	.874
4	.981	.976	.970	.963	.955
5	.995	.993	.992	.989	.986
6	.999	.998	.998	.997	.996
7	.999	.999	.999	.999	.999
8	.999	.999	.999	.999	.999
9	.999	.999	.999	.999	.999
10	.999	.999	.999	.999	.999
x	c = λt				
	2.00	2.10	2.20		
0	.135	.122	.110		
1	.406	.379	.354		
2	.676	.649	.622		
3	.857	.838	.819		
4	.947	.937	.927		
5	.983	.979	.975		
6	.995	.994	.992		
7	.998	.998	.998		
8	.999	.999	.999		
9	.999	.999	.999		
10	.999	.999	.999		

Distribución acumulativa de Poisson (continuación)

x	7.50	8.00	8.50	9.00	9.50	10.0	15.0	20.0
0	.000	.000	.000	.000	.000	.000	.000	.000
1	.004	.003	.001	.001	.000	.000	.000	.000
2	.020	.013	.009	.006	.004	.002	.000	.000
3	.059	.042	.030	.021	.014	.010	.000	.000
4	.132	.099	.074	.054	.040	.029	.000	.000
5	.241	.191	.149	.115	.088	.067	.002	.000
6	.378	.313	.256	.206	.164	.130	.007	.000
7	.524	.452	.385	.323	.268	.220	.018	.000
8	.661	.592	.523	.455	.391	.332	.037	.002
9	.776	.716	.652	.587	.521	.457	.069	.005
10	.862	.815	.763	.705	.645	.583	.118	.010
11	.920	.888	.848	.803	.751	.696	.184	.021
12	.957	.936	.909	.875	.836	.791	.267	.039
13	.978	.965	.948	.926	.898	.864	.363	.066
14	.989	.982	.972	.958	.940	.916	.465	.104
15	.995	.991	.986	.977	.966	.951	.568	.156
16	.998	.996	.993	.988	.982	.972	.664	.221
17	.999	.998	.997	.994	.991	.985	.748	.297
18	.999	.999	.998	.997	.995	.992	.819	.381
19	.999	.999	.999	.998	.998	.996	.875	.470
20	.999	.999	.999	.999	.999	.998	.917	.559
21	.999	.999	.999	.999	.999	.999	.946	.643
22	.999	.999	.999	.999	.999	.999	.967	.720
23	.999	.999	.999	.999	.999	.999	.980	.787
24	.999	.999	.999	.999	.999	.999	.988	.843
25	.999	.999	.999	.999	.999	.999	.993	.887
26	.999	.999	.999	.999	.999	.999	.996	.922
27	.999	.999	.999	.999	.999	.999	.998	.947
28	.999	.999	.999	.999	.999	.999	.999	.965
29	.999	.999	.999	.999	.999	.999	.999	.978
30	.999	.999	.999	.999	.999	.999	.999	.986
31	.999	.999	.999	.999	.999	.999	.999	.991
32	.999	.999	.999	.999	.999	.999	.999	.995
33	.999	.999	.999	.999	.999	.999	.999	.997
34	.999	.999	.999	.999	.999	.999	.999	.998

Las entradas en la tabla son valores de $F(x) = P(C \leq x) = \sum_{i=0}^x e^{-c} c^i / i!$. Los espacios en blanco sobre la primera entrada en cualquier columna pueden leerse como 0.0.

ANEXO H

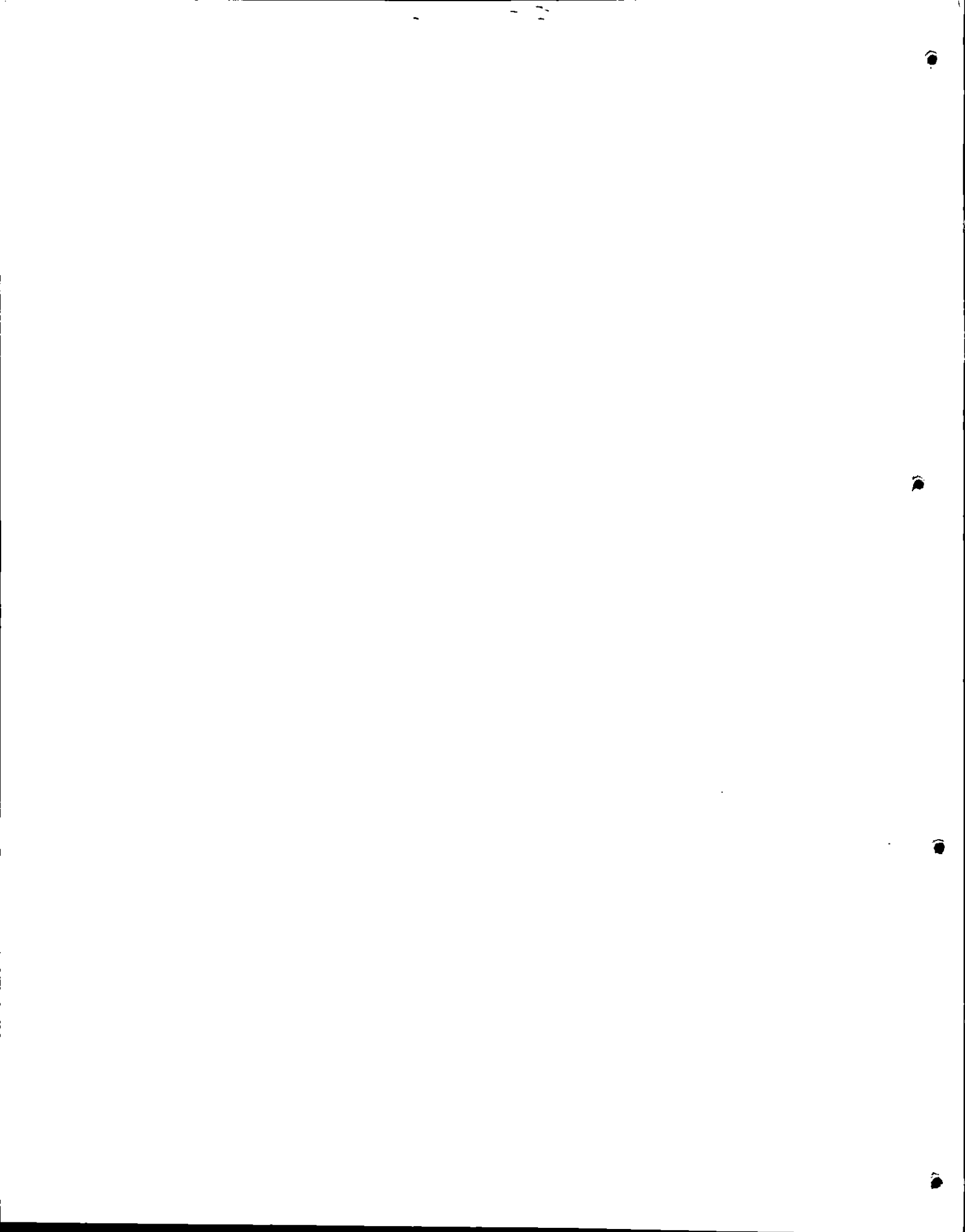
Función de distribución binomial

$$B(x; n, p) = \sum_{k=0}^x \binom{n}{k} p^k (1-p)^{n-k}$$

n	x	p																		
		0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500	0.2025	0.1600	0.1225	0.0900	0.0625	0.0400	0.0225	0.0100	0.0025
	1	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500	0.6975	0.6400	0.5775	0.5100	0.4375	0.3600	0.2775	0.1900	0.0975
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250	0.0911	0.0640	0.0429	0.0270	0.0156	0.0080	0.0034	0.0010	0.0001
	1	0.9927	0.9720	0.9393	0.8960	0.8438	0.7840	0.7183	0.6480	0.5748	0.5000	0.4252	0.3520	0.2818	0.2160	0.1563	0.1040	0.0607	0.0280	0.0073
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625	0.0410	0.0256	0.0150	0.0081	0.0039	0.0016	0.0005	0.0001	0.0000
	1	0.9860	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125	0.2415	0.1792	0.1265	0.0837	0.0508	0.0272	0.0120	0.0037	0.0005
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313	0.0185	0.0102	0.0053	0.0024	0.0010	0.0003	0.0001	0.0000	0.0000
	1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875	0.1312	0.0870	0.0540	0.0308	0.0156	0.0067	0.0022	0.0005	0.0000
6	0	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156	0.0083	0.0041	0.0018	0.0007	0.0002	0.0001	0.0000	0.0000	0.0000
	1	0.9672	0.8857	0.7765	0.6554	0.5339	0.4202	0.3191	0.2333	0.1636	0.1094	0.0692	0.0410	0.0223	0.0109	0.0046	0.0016	0.0004	0.0001	0.0000
7	0	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078	0.0037	0.0016	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
	1	0.9556	0.8503	0.7166	0.5767	0.4449	0.3294	0.2338	0.1586	0.1024	0.0625	0.0357	0.0188	0.0090	0.0038	0.0013	0.0004	0.0001	0.0000	0.0000
8	0	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039	0.0017	0.0007	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.9428	0.8131	0.6572	0.5033	0.3671	0.2553	0.1691	0.1064	0.0632	0.0352	0.0181	0.0085	0.0036	0.0013	0.0004	0.0001	0.0000	0.0000	0.0000

ANEXO.I Números aleatorios

10480	15011	01536	02011	81647	31646	59179	14194	62590
22368	46573	25595	85393	30995	89198	27982	53402	93965
24130	48360	22527	97265	76393	64809	15179	24830	49340
42167	93093	06243	61680	07856	16376	39440	53537	71341
37570	39975	81837	16656	06121	91782	60468	81305	49684
77921	06907	11008	42751	27756	53498	18602	70659	90655
99562	72905	56420	69994	98872	31016	71194	18738	44013
96301	91977	05463	07972	18876	20922	94595	56869	69014
89579	14342	63661	10281	17453	18103	57740	3437S	25331
35475	36857	53342	53988	53060	59533	38867	62300	08158
28918	69578	38231	33276	70997	79936	56865	05859	90106
63553	40961	48235	03427	49626	69445	18663	72695	52180
09429	93969	52636	92737	88974	33488	36320	17617	30015
10365	61129	37529	85689	48237	52267	67689	93394	01511
07119	97336	71048	08178	77233	13916	47564	81056	97735
51085	12765	51821	51259	77452	16308	60756	92144	49442
02368	21382	52404	60268	89368	19885	55322	44819	01188
01011	54092	33362	94904	31273	04146	18594	29852	71585
52162	53916	46369	58586	23216	14513	83149	98736	23495
07056	97628	33787	09998	42698	06691	76988	13602	51851
48663	91245	85828	14346	09172	30168	90229	04734	59193
54164	58492	22421	74103	47070	25306	76468	26384	58151
32639	32363	05597	24200	13363	38005	94342	28728	35806
29334	27001	87637	87308	58731	00256	45834	15398	46557
02488	33062	28834	07351	19731	92420	60952	61280	50001
81525	72295	04839	96423	24878	82651	66566	14778	76797
29676	20591	68086	26432	46901	20849	89768	81536	86645
00742	57392	39064	66432	84673	40027	32832	61362	98947
05366	04213	25669	26422	44407	44048	37937	63904	45766
91921	26418	64117	94305	26766	25940	39972	22209	71500
00582	04711	87917	77341	42206	35126	74087	99547	81817
00725	69884	62797	56170	86324	88072	76222	36086	84637
69011	65795	95876	55293	18988	27354	26575	08625	40801
25976	57948	29888	88604	67917	48708	18912	82271	65424
09763	83473	73577	12908	30883	18317	28290	35797	05998
91567	42595	27958	30134	04024	86385	29880	99730	55536
17955	56349	90999	49127	20044	59931	06115	20542	18059
46503	18584	18845	49618	02304	51038	20655	58727	28168
92157	89634	94824	78171	84610	82834	09922	25417	44137
14577	62765	35605	81263	39667	47358	56873	56307	61607
98427	07523	33362	64270	01638	92477	66969	98420	04880
34914	63976	88720	82765	34476	17032	87589	40836	32427
70060	28277	39475	46473	23219	53416	94970	25832	69975
53976	54914	06990	67245	68350	82948	11398	42878	80287
76072	29515	40980	07391	58745	25774	22987	80059	39911
90725	52210	83974	29992	65831	38857	50490	83765	55657
64364	67412	33339	31926	14883	24413	59744	92351	97473
08962	00358	31662	25388	61642	34072	81249	35648	56891
95012	68379	93526	70765	10592	04542	76463	54328	02349
15664	10493	20492	38391	91132	21999	59516	81652	27195



ANEXO J Valores críticos para la prueba Wilcoxon de dos muestras ^a

$R_{.05}^*$

$n_2 \backslash n_1$	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4			10											
5		6	11	17										
6		7	12	18	26									
7		7	13	20	27	36								
8	3	8	14	21	29	38	49							
9	3	8	15	22	31	40	51	63						
10	3	9	15	23	32	42	53	65	78					
11	4	9	16	24	34	44	55	68	81	96				
12	4	10	17	26	35	46	58	71	85	99	115			
13	4	10	18	27	37	48	60	73	88	103	119	137		
14	4	11	19	28	38	50	63	76	91	106	123	141	160	
15	4	11	20	29	40	52	65	79	94	110	127	145	164	185
16	4	12	21	31	42	54	67	82	97	114	131	150	169	
17	5	12	21	32	43	56	70	84	100	117	135	154		
18	5	13	22	33	45	58	72	87	103	121	139			
19	5	13	23	34	46	60	74	90	107	124				
20	5	14	24	35	48	62	77	93	110					
21	6	14	25	37	50	64	79	95						
22	6	15	26	38	51	66	82							
23	6	15	27	39	53	68								
24	6	16	28	40	55									
25	6	16	28	42										
26	7	17	29											
27	7	17												
28	7													

Fuente: Reproducida con autorización de "The Use of Ranks in a Test of Significance for Comparing Two Treatments," por C. White, *Biometrics*, 1952, Vol. 8, p. 37.

^aPara n_1 y n_2 grandes, R se distribuye aproximadamente en forma normal con media $\frac{1}{2}n_1(n_1 + n_2 + 1)$ y varianza $\frac{1}{12}n_1n_2(n_1 + n_2 + 1)$.

ANEXO K Valores críticos para la prueba del signo^a R_n^*

$n \backslash \alpha$.10	.05	.01
5	0		
6	0	0	
7	0	0	
8	1	0	0
9	1	1	0
10	1	1	0
11	2	1	0
12	2	2	1
13	3	2	1
14	3	2	1
15	3	3	2
16	4	3	2
17	4	4	2
18	5	4	3
19	5	4	3
20	5	5	3
21	6	5	4
22	6	5	4
23	7	6	4
24	7	6	5
25	7	7	5
26	8	7	6
27	8	7	6
28	9	8	6
29	9	8	7
30	10	9	7
31	10	9	7
32	10	9	8
33	11	10	8
34	11	10	9
35	12	11	9
36	12	11	9
37	13	12	10
38	13	12	10
39	13	12	11
40	14	13	11

^aPara $n > 40$, R se distribuye aproximadamente en forma normal con media $n/2$ y varianza $n/4$.

ANEXO L

Valores críticos para la prueba Wilcoxon del rango con signo^a

$n \backslash \alpha$.10	.05	.02	.01
4	0	0	0	0
5	2	2	0	1
6	3	3	1	1
7	5	5	3	3
8	8	8	5	5
9	10	10	8	8
10	13	13	10	10
11	17	17	13	13
12	21	21	17	17
13	25	25	21	21
14	30	30	25	25
15	35	35	29	29
16	41	41	34	34
17	47	47	40	40
18	53	53	46	46
19	60	60	52	52
20	67	67	58	58
21	75	75	65	65
22	83	83	73	73
23	91	91	81	81
24	100	100	89	89
25	110	110	98	98
26	119	119	107	107
27	130	130	116	116
28	140	140	126	126
29	151	151	137	137
30	163	163	147	147
31	175	175	159	159
32	187	187	170	170
33	200	200	182	182
34	213	213	195	195
35	227	227	208	208
36	241	241	221	221
37	256	256	235	235
38	271	271	249	249
39	286	286	264	264
40	302	302	279	279
41	319	319	294	294
42	336	336	310	310
43	353	353	327	327
44	371	371	343	343
45	389	389	361	361
46	407	407	378	378
47	426	426	396	396
48	446	446	415	415
49	466	466	434	434
50			397	373

Fuente: Adaptada con autorización de "Extended Tables of the Wilcoxon Matched Pair Signed Rank Statistic", por Robert L. McCormack, *Journal of the American Statistical Association*, Vol. 60, septiembre, 1965.

^aSi $n > 50$, R se distribuye aproximadamente en forma normal con media $n(n+1)/4$ y varianza $n(n+1)/24$.