

**UNIVERSIDAD DE EL SALVADOR  
FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE  
DEPARTAMENTO DE MATEMÁTICA**



**TRABAJO DE GRADUACIÓN:**

**MANUAL DE FUNDAMENTOS BÁSICOS DE ESTADÍSTICA  
DESCRIPTIVA E INFERENCIAL PARA ESTUDIANTES DE CIENCIAS  
NATURALES, ESPECIALIDAD BIOLOGÍA.**

**PARA OPTAR AL GRADO DE:  
LICENCIATURA EN ESTADÍSTICA**

**PRESENTADO POR:  
FÁTIMA ELENA ESTRADA TERESÓN  
ANA MARCELA LANDAVERDE SOSA**

**DOCENTE ASESOR:  
INGENIERO JOSÉ ELÍAS BARAHONA**

**JUNIO 2018  
SANTA ANA, EL SALVADOR, CENTRO AMÉRICA**

**UNIVERSIDAD DE EL SALVADOR**  
**AUTORIDADES CENTRALES**



**M.Sc. ROGER ARMANDO ARIAS ALVARADO**  
**RECTOR**

**DR. MANUEL DE JESÚS JOYA ÁBREGO**  
**VICE-RECTOR ACADEMICO**

**ING. NELSON BERNABÉ GRANADOS ALVARADO**  
**VICE-RECTOR ADMINISTRATIVO**

**LICDO. CRISTOBAL HERNÁN RÍOS BENÍTEZ**  
**SECRETARIO GENERAL**

**M.Sc. CLAUDIA MARÍA MELGAR DE ZAMBRANA**  
**DEFENSOR DE LOS DERECHOS UNIVERSITARIOS**

**LICDO. RAFAEL HUMBERTO PEÑA MARÍN**  
**FISCAL GENERAL**

**FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE**  
**AUTORIDADES**



**DR. RAUL ERNESTO AZCÚNAGA LÓPEZ**  
**DECANO**

**M.Ed. ROBERTO CARLOS SIGÜENZA CAMPOS**  
**VICE-DECANO**

**M.Sc. DAVID ALFONSO MATA ALDANA**  
**SECRETARIO DE LA FACULTAD**

**LICDA. ROSARIO CÁCERES AGUILAR**  
**JEFE DEL DEPARTAMENTO DE MATEMÁTICA**

## **AGRADECIMIENTO**

### **A DIOS**

Agradezco primeramente a Dios nuestro padre celestial por darme la vida y así haber tener la oportunidad de poder culminar mis estudios superiores.

### **A MIS PADRES**

Agradezco también a mis padres que con mucho esfuerzo y dedicación me apoyaron en cada momento tanto moral como espiritualmente y me animaron a seguir adelante porque siempre confiaron en mí.

### **A MI HERMANO**

Agradezco a mi hermano que a pesar de que no estamos juntos todos los días él siempre estuvo pendiente en lo que necesitara para seguir adelante y fue un apoyo económico muy importante para poder cumplir mis metas y anhelos.

*“No olvides quien eres, a quien perteneces y no pierdas jamás la esperanza...  
Dios apenas ha comenzado contigo”*

*Ana Marcela Landaverde Sosa*

<b>CONTENIDO</b>	<b>PAG.</b>
RESUMEN.....	XIII
CAPÍTULO 1: PLANTEAMIENTO DEL PROBLEMA.....	14
1.1. PROBLEMÁTICA .....	14
1.2. JUSTIFICACIÓN.....	15
1.3. OBJETIVOS .....	17
1.4. MARCO TEÓRICO CONTEXTUAL.....	18
1.4.1. Manual .....	18
1.4.2. Estadística descriptiva.....	18
1.4.3. Probabilidad.....	18
1.4.4. Estadística inferencial .....	19
1.4.5. Diseño de experimentos.....	19
1.4.6. Biología.....	20
1.5. LA ESTADISTICA EN LAS CIENCIAS BIOLOGICAS.....	20
1.6. VENTAJAS DE LA ELABORACION DEL MANUAL .....	21
CAPÍTULO 2: FUNDAMENTOS DE LA ESTADISTICA DESCRIPTIVA .....	22
2.1. Introducción a la estadística descriptiva .....	22
2.1.1. Conceptos básicos.....	22
2.2. Clasificación de variables .....	22
2.2.1. Variables por su medición .....	23
2.2.2. Clasificación según escalas de medición .....	25
2.2.3. Variables por causa.....	27

2.3.	Recopilación de datos .....	27
2.3.1.	Recolección indirecta.....	27
2.3.2.	Recolección directa.....	28
2.4.	Presentación de una sola variable .....	37
2.4.1.	Distribución de frecuencias .....	38
2.4.2.	Presentación gráfica de datos.....	45
2.5.	Análisis descriptivo de una sola variable.....	48
2.5.1.	Medidas de tendencia central.....	48
2.5.2.	Medidas de posición .....	54
2.5.3.	Medidas de dispersión. ....	60
2.5.4.	Medidas de forma .....	64
2.6.	Presentación y análisis de dos variables .....	67
2.6.1.	Tablas de contingencia.....	68
2.6.2.	Distribuciones marginales y condicionales.....	69
2.6.3.	Análisis descriptivo de dos variables.....	70
CAPÍTULO 3: NOCIONES DE PROBABILIDAD .....		74
3.1.	Técnicas de conteo.....	74
3.1.1.	Permutaciones.....	74
3.1.2.	Combinaciones.....	75
3.2.	Espacio muestral .....	76
3.2.1.	Tipo de espacios muestrales .....	76
3.3.	Propiedades de la probabilidad .....	77
3.4.	Independencia y probabilidad condicional .....	77

3.5.	Tipo de variables aleatorias .....	79
3.5.1.	Variables aleatorias discretas.....	79
3.5.2.	Esperanza matemática, varianza y covarianza.....	87
3.5.3.	Variables aleatorias continuas .....	90
CAPÍTULO 4: INFERENCIA ESTADISTICA.....		94
4.1.	Estadística Inferencial.....	94
4.1.1.	Definición de Estadística Inferencial .....	94
4.1.2.	Distribuciones muestrales .....	94
4.1.3.	Procedimientos de Muestreo.....	97
4.1.4.	Intervalo de Confianza.....	98
4.1.5.	Prueba de Hipótesis .....	99
4.1.6.	Procedimiento para Probar una Hipótesis.....	99
4.2.	Análisis de varianza y pruebas no paramétricas .....	103
4.2.1.	Análisis de Varianza (Anova).....	103
4.2.2.	Pruebas no Paramétricas .....	104
4.3.	Análisis de Regresión y Correlación .....	107
4.3.1.	Análisis de Regresión .....	107
4.3.2.	Análisis de Regresión Múltiple.....	112
4.3.3.	Análisis de Correlación.....	112
4.3.4.	Análisis de Correlación Múltiple.....	114
CAPÍTULO 5: DISEÑO DE EXPERIMENTOS.....		115
5.1.	Fundamentos de diseños experimentales.....	115
5.1.1.	Definición de un diseño experimental .....	115

5.1.2.	Propósitos de un diseño experimental .....	115
5.1.3.	Tipos de tratamientos.....	115
5.1.4.	Tipos de error experimental.....	116
5.1.5.	Principios básicos .....	116
5.1.6.	Forma de aleatorizar un experimento .....	118
5.1.7.	Directrices o procedimiento para el diseño experimental.....	118
5.1.8.	Utilización de los métodos estadísticos en la experimentación .....	120
5.2.	Diseños Unifactoriales.....	121
5.2.1.	Descripción.....	121
5.2.2.	Representación simbólica de los datos. ....	122
5.2.3.	Modelo estadístico .....	122
5.2.4.	Sumas y medias de cuadrados .....	123
5.2.5.	Análisis Residual .....	125
5.2.6.	Comparación de parejas de medias de tratamiento.....	125
5.3.	Bloqueo y factores extraños.....	128
5.4.	Diseño de cuadro latino .....	129
5.4.1.	Modelo Estadístico del cuadrado latino.....	130
5.4.2.	Sumas y Medias de Cuadrados.....	131
5.4.3.	Análisis Estadístico.....	132
5.5.	Diseño de experimentos factoriales y fraccionales.....	133
5.5.1.	Experimentos factoriales.....	133
5.6.	El diseño factorial $2^k$ .....	134
5.6.1.	El diseño $2^2$ .....	134



5.6.2. Una sola réplica del diseño $2^k$ .....	135
5.6.3. Adición de puntos centrales al diseño $2^k$ .....	136
CAPÍTULO 6: APLICACIONES DEL CAPITULO 4 Y CAPITULO 5 .....	138
Ejemplo 1:” Efecto antidepresivo y toxicidad del veneno de sapo sabanero (Rhinella marina) en ratones de laboratorio” .....	138
Ejemplo 2: “Estudio sobre la roya en El Salvador período 2016-2017” .....	148
CAPÍTULO 7: CONCLUSIONES Y RECOMENDACIONES DEL MANUAL....	176
7.1 CONCLUSIONES.....	176
7.2 RECOMENDACIONES.....	177
BIBLIOGRAFIA .....	179
ANEXOS .....	182
Anexo 1: Ejemplos prácticos del programa SPSS.....	182
Anexo 2: Distribución normal estandarizada, $N(0,1)$ .....	200
Anexo 3: Tabla de la distribución T de student.....	202
Anexo 4: tabla de distribucion F de fisher (0.01).....	203
Anexo 5: tabla de valores críticos de la prueba de tukey $qa(V1, V2)$ .....	204
Anexo 6: tabla de valores críticos de la prueba de Duncan $Ua(V1, V2)$ .....	207
Anexo 7: tabla de valores críticos de Dunnett.....	209

## **Indicé de gráficos**

Gráfico 1: Diagrama de barras para la tabla 1.....	45
Gráfico 2: Diagrama de sectores para la tabla 1 .....	47
Gráfico 3: histograma para la tabla 2 .....	48
Gráfico 4: Ejemplo de diagrama de dispersión. ....	68
Gráfico 5: peso de las botellas con agua y solución sacarosa por cada semana. ....	142
Gráfico 6: Tiempo de inmovilización ratones hembra (segundos) al finalizar las pruebas de estrés y Test de Nado Forzado. ....	144
Gráfico 7: Curvas de incidencia y severidad de roya y su relación con la precipitación en la finca Normandía (bajío). ....	155
Gráfico 8: Curvas de incidencia y severidad de roya y su relación con la humedad relativa en la finca Normandía (bajío). ....	156
Gráfico 9: Curvas de incidencia y severidad de roya y su relación con la temperatura en la finca Normandía (bajío). ....	157
Gráfico 10: Área Bajo la Curva de Progreso de la Incidencia y Severidad (ABCPI Y ABCPS) de la roya del café en la finca Normandía (bajío). ....	158
Gráfico 11: Curvas de incidencia y severidad de la enfermedad y su relación con la precipitación en la finca San Antonio (media altura).....	162
Gráfico 12: Curvas de incidencia y severidad de roya del café y su relación con la humedad relativa en la finca San Antonio (media altura).....	163
Gráfico 13: Curvas de incidencia y severidad de roya del café y su relación con la temperatura en la finca San Antonio (media altura). ....	164
Gráfico 14: Área Bajo la Curva de Progreso de la Incidencia y Severidad (ABCPI y ABCPS), en la finca San Antonio (media altura). ....	165

Gráfico 15: Curvas de incidencia y severidad y su relación con la precipitación en la finca El Olvido (estricta altura).....	169
Gráfico 16: Curvas de incidencia y severidad y su relación con la humedad relativa en la finca El Olvido (estricta altura).....	170
Gráfico 17: Curvas de incidencia y severidad y su relación con la humedad relativa en la finca El Olvido (estricta altura).....	171
Gráfico 18: Área Bajo la Curva de la Incidencia y Severidad (ABCPI y ABCPS) de la roya del café en la finca El Olvido (estricta altura). ....	172

### **Indicé de tablas**

Tabla 1: Distribución de frecuencia de tipos de orquídeas en un bosque.....	39
Tabla 2: Distribución de frecuencia del tiempo de vida de 100 moscas. ....	44
Tabla 3: Datos de la edad de 10 árboles en el cerro verde. ....	54
Tabla 4: Ejemplo de tabla de contingencia .....	68
Tabla 5: Ejemplo de tratamientos. ....	122
Tabla 6: Análisis de Varianza (ANOVA).....	125
Tabla 7 : Análisis de varianza (ANOVA). ....	132
Tabla 8 : Análisis de varianza (ANOVA). ....	133
Tabla 9: Promedio y relación entre pesos corporales (g) de ratones hembra al inicio y final de la prueba. ....	143
Tabla 10: Tiempos promedios (seg.) de los grupos de ratones hembra sometidos a evaluación de efecto antidepresivo del veneno de R. marina. ....	143

Tabla 11: Nivel de incidencia de la enfermedad en hojas viejas y nuevas en los meses de mayo de 2016 hasta abril de 2017 en finca de bajo.....	152
Tabla 12: Nivel de la severidad de la enfermedad en hojas viejas y nuevas durante los meses de mayo de 2011 a abril de 2012 en la finca de Bajo.....	153
Tabla 13: Nivel de incidencia de la enfermedad en hojas viejas y nuevas durante los meses de mayo de 2016 a abril de 2017 en la finca de media altura. .....	160
Tabla 14: Niveles de severidad de la enfermedad en hojas viejas y nuevas durante los meses de mayo de 2016 a abril de 2017 en la finca de media altura. .....	161
Tabla 15: Nivel de incidencia de la enfermedad en hojas viejas y nuevas en los meses de mayo de 2016 a abril de 2017 en finca de estricta altura. ....	167
Tabla 16: Nivel de severidad de la enfermedad en hojas viejas y nuevas, durante los meses de mayo de 2016 a abril de 2017 en la finca de estricta altura. ....	168

## **RESUMEN**

La Universidad de El Salvador Facultad Multidisciplinaria de Occidente, ofrece varias carreras para realizar estudios superiores; una de ellas es la de Licenciatura en Biología, con una carga académica desde el año 2003 en la cual se les imparten 35 materias divididas en diez ciclos y un trabajo de grado como requisito de graduación; dentro de esas materias se imparta la materia de estadística general y una materia optativa llamada bioestadística.

Tomando en cuenta, que la estadística es una ciencia de gran importancia para la realización de investigaciones ya sea de campo o de trabajos de grado, ya que por medio de ella se puede realizar la presentación de los datos recolectados y hacer análisis de ellos; es por eso que se ha elaborado este manual para los estudiantes de la carrera de Licenciatura en Biología.

En este manual se pretende que los alumnos puedan encontrar los conceptos básicos y esenciales de la estadística descriptiva e inferencial; así como ejemplos que les ayuden al lector a poder realizar cada uno de los cálculos con sus respectivas interpretaciones; también presenta el contenido de nociones de probabilidad y diseño de experimentos, contiene dos ejemplos de aplicaciones uno de inferencia estadística y otro de diseño de experimentos; para que el alumno o lector pueda comprender mejor de una forma práctica.

La parte de anexos contiene: El uso del programa informático SPSS en la parte de Estadística Descriptiva; tomando en cuenta los ejemplos que se han desarrollado de manera manual para que el lector pueda comparar esos resultados con los del programa. Al igual se le presentan las tablas a utilizar en las diferentes distribuciones o pruebas realizadas en este contenido.

## **CAPÍTULO 1: PLANTEAMIENTO DEL PROBLEMA**

### **1.1. PROBLEMÁTICA**

La Universidad de El Salvador Facultad Multidisciplinaria de Occidente, está conformada por varios departamentos, uno de ellos es el departamento de Matemática, al igual que el departamento de Biología en el cual se encuentra la carrera de Licenciatura en Biología.

En el desarrollo de esta carrera se le presentan al estudiante muchas investigaciones de campo, en las cuales los alumnos tienen que presentar un informe de todo el desarrollo de sus investigaciones, desde la creación de la idea hasta un análisis de los datos recolectados, es cuando la estadística hace su función ya que por medio de ella se puede organizar el proceso de desarrollo de las investigaciones, luego puede realizar la creación de un instrumento de recolección de datos , ya teniendo la información se pasa al análisis de estos datos y para finalizar se puede realizar estimar utilizando dicha información obtenida.

Es por lo anterior que la estadística se vuelve una herramienta muy importante, para el análisis de datos, es por eso que se tiene que realizar un estudio de toda la teoría de estadística para ponerla en práctica en las investigaciones que se realicen, para que los datos observados pueda tener una presentación y un análisis estadístico adecuado.

La mayoría de estudiantes de esta Licenciatura tiene dificultad, al utilizar la estadística a sus investigaciones, ya que solo se les imparten dos materias relacionadas con el área de estadística, una de ellas es Estadística general que es una materia obligatoria de cursar y Bioestadística como materia electiva; en las cuales solo se les imparte lo esencial ya que tienen un periodo de cuatro meses para abarcar todo el contenido, y es por lo cual los alumnos llegan a tener dificultades al realizar los análisis ya sea en investigaciones de campo que se les pidan en una materia o también en trabajos de grado al final de su carrera.

## 1.2. JUSTIFICACIÓN

La estadística es una ciencia que analiza datos (por ejemplo, edad de una población, altura de un equipo de baloncesto, temperatura de los meses de verano, etc.) y trata de extraer conclusiones sobre el comportamiento de estas variables. Es una de las ciencias que permite conocer, o al menos entender, la realidad en la que nos desenvolvemos. A través de la estadística podemos obtener información de gran valor que nos ayudará en la toma de decisiones en cualquier ámbito de nuestra vida. La Biología, se preocupa de los procesos vitales de cada ser. Como su nacimiento, desarrollo, procreación y muerte.

En la Facultad Multidisciplinaria de Occidente de la Universidad de El Salvador se imparte la carrera de Licenciatura en Biología, y en la carga académica desde el año 2003; se imparte la materia de Estadística general como asignatura obligatoria y Bioestadística como asignatura optativa, sin embargo en biología se diseñan experimentos en los cuales se utiliza la estadística descriptiva y la estadística inferencial para el tratamiento de información de una muestra para generalizar al resto de la población, además de analizar la variabilidad para contrastar hipótesis.

En un semestre no se logra la comprensión y el manejo de todos los conceptos estadísticos necesarios para poder utilizarlos. En la actualidad no hay manuales específicos para el área de Licenciatura en Biología en la FMO-UES<sup>1</sup> con respecto al área de estadística

Se consultó al Licenciado Carlos Linares (jefe del departamento de Biología año 2017) si existía en la carrera de Licenciatura en Biología un libro donde se encuentren los procedimientos estadísticos cuando se realiza una investigación científica, y él nos aseguró que no se cuenta con un recurso, donde se recopile todo lo que concierne a la estadística, y que a los estudiantes les cuesta encontrar bibliografía donde explique cómo obtener los parámetros estadísticos e interpretar y analizar los resultados.

---

1-FMO-UES: Facultad Multidisciplinaria de Occidente de la Universidad de El Salvador

Es por esto que resulta relevante la elaboración de un manual de estadística para los estudiantes de la Licenciatura en Biología, para que ellos puedan consultar y así procesar los datos y tener una mejor comprensión para el análisis y la interpretación de ellos.

También no se puede dejar de lado el aspecto tecnológico, ya que se quiere presentar una forma más rápida y concreta para poder realizar los cálculos estadísticos. Los programas más usados son SAS, STATA, MINITAB, S-PLUS, INSTAT, INFOSTAT, SPSS, R, PYTHON, STATGRAPHIC entre otros; en el manual se presentara en los anexos los pasos para utilizar el programa estadístico SPSS y las capturas de pantalla donde se mostraran paso a paso los procesos y los resultados de los cálculos realizados.



### **1.3. OBJETIVOS**

#### **Objetivo general**

Elaborar un manual de estadística descriptiva e inferencial para estudiantes de la Licenciatura en Biología en la Facultad Multidisciplinaria de Occidente de la Universidad de El Salvador.

#### **Objetivos específicos**

Definir los aspectos conceptuales de la Estadística Descriptiva e Inferencial

Explicar el uso del software estadístico SPSS en la Estadística Descriptiva, para facilitar el procesamiento de los datos en las investigaciones que se realizan en la carrera de Licenciatura en Biología.

Establecer los conceptos básicos y formulas necesarias para que el estudiante pueda presentar los datos de las investigaciones que realice, en gráficas adecuadas.

Dar a conocer la forma en la que se puede interpretar los resultados obtenidos de algunos procesos estadísticos realizados, para tomar decisiones adecuadas.

## **1.4. MARCO TEÓRICO CONTEXTUAL**

### **1.4.1. Manual**

Un manual es una publicación que incluye los aspectos fundamentales de una materia. Así, los manuales nos permiten comprender mejor el funcionamiento de algo, o acceder de manera ordenada y concisa, al conocimiento de algún tema o material. La palabra, como tal, proviene del latín manuālis, que significa: que se puede tomar con la mano o que se lleva con la mano. (Diccionario de la lengua española, 2017)

### **1.4.2. Estadística descriptiva**

Los métodos de la Estadística Descriptiva o Análisis Exploratorio de Datos ayudan a presentar los datos de modo tal que sobresalga su estructura.

Hay varias formas simples e interesantes de organizar los datos en gráficos que permiten detectar tanto las características sobresalientes como las características inesperadas. El otro modo de describir los datos es resumirlos en uno o dos números que pretenden caracterizar el conjunto con la menor distorsión o pérdida de información posible.

Explorar los datos, debe ser la primera etapa de todo análisis de datos.

En primer lugar, porque las computadoras no son demasiado hábiles (sólo son rápidas), hacen aquello para lo que están programadas y actúan sobre los datos que les ofrecemos. Datos erróneos o inesperados serán procesados de modo inapropiado y ni usted, ni la computadora se darán cuenta a menos que realice previamente un análisis exploratorio de los datos.

### **1.4.3. Probabilidad**

La teoría de probabilidad actualmente se ha convertido en una herramienta indispensable para aquellos estudios en los que se presenta la incertidumbre, la creación de esta teoría fue principalmente por los juegos al azar, ya que era una de las maneras para conocer si el jugador podía ganar o perder.

#### **1.4.4. Estadística inferencial**

La Estadística Inferencial hace referencia a un conjunto de métodos que permiten inferir acerca de características de un fenómeno sobre la base de información parcial acerca del mismo.

Los métodos de la inferencia nos permiten proponer el valor de una cantidad desconocida (estimación) o decidir entre dos propuestas cuál de ellas explica mejor los datos observados (test de hipótesis).

El fin de cualquier estudio es aprender sobre las poblaciones. Pero es usualmente necesario y más práctico, estudiar solo una muestra de cada una de las poblaciones.

La ventaja de los métodos estadísticos es que aplicados sobre datos obtenidos a partir de muestras aleatorias permiten cuantificar el error que podemos cometer en nuestra estimación o calcular la probabilidad de cometer un error al tomar una decisión en un test de hipótesis.

Finalmente, cuando existen datos para toda la población (CENSO) no hay necesidad de usar métodos de estadística inferencial, ya que es posible calcular exactamente los parámetros de interés. En el censo poblacional, por ejemplo, se registra el sexo de todas las personas censadas, que son prácticamente toda la población, así que es posible conocer exactamente la proporción de habitantes de los dos sexos.

#### **1.4.5. Diseño de experimentos**

Es una actividad crucial, Consiste en definir como se desarrollará la investigación para dar respuesta a las preguntas que motivaron la misma. La recolección de los datos requiere en general de un gran esfuerzo, por lo que, dedicar especial cuidado a la etapa de planificación de la investigación ahorra trabajo en las siguientes etapas.

Un estudio bien diseñado resulta simple de analizar y las conclusiones suelen ser obvias. Un experimento pobremente diseñado o con datos inapropiadamente

recolectados o registrados puede ser incapaz de dar respuesta a las preguntas que motivaron la investigación, más allá de lo sofisticado que sea el análisis estadístico.

Aún en los casos en que se estudian datos ya registrados, en que estamos restringidos a la información existente, los principios del buen diseño de experimentos, pueden ser útiles para ayudar a seleccionar un conjunto razonable de datos que esté relacionado con el problema de interés.

#### **1.4.6. Biología**

Estudia los seres vivos de manera integral, desde el nivel molecular hasta como integrante de los ecosistemas, a fin de conocer su estructura, función, diversidad, origen, evolución, e interrelaciones.

Su marco conceptual se basa en la teoría de la evolución y los conocimientos fundamentales del origen, la evolución y las funciones de los seres vivos. Así mismo se enfoca al estudio de los procesos hereditarios, la diversidad de los seres vivos, el desarrollo de los organismos, y las relaciones entre éstos y su medio.

En la Licenciatura en Biología se prepara al profesional conocido como biólogo, el cual debe ser capaz de comprender y valorar la vida en todas sus manifestaciones y niveles, adquiriendo la formación y la capacidad Científica técnica que le permita respetarla, conservarla y fomentarla.

### **1.5. LA ESTADISTICA EN LAS CIENCIAS BIOLOGICAS**

La principal ventaja del pensamiento estadístico interviniendo en la Biología es que no solo resuelve, sino que también comprende una compleja metodología para dar respuesta a las hipótesis, además de agilizar la cuestión de organización del sistema de investigación, desde el diseño general, el de muestreo, el control de la calidad de información y la presentación de los resultados.

## **1.6. VENTAJAS DE LA ELABORACION DEL MANUAL**

Las ventajas que se obtienen al utilizar este manual son:

- Facilitan al estudiante los conceptos básicos del área de estadística, para que pueda dar solución a sus problemas de investigaciones biológicas o de entendimiento del desarrollo de su materia de estadística.
- El estudiante tendrá una guía de referencia para los procedimientos estadísticos necesarios para el buen desarrollo de sus investigaciones.
- Motivar a los estudiantes a indagar más acerca de las diversas técnicas de análisis estadístico que les sean útiles en el desarrollo de sus investigaciones.
- Los ejemplos que se plantean en el contenido son aplicaciones del área de biología, para que el alumno pueda observar que la estadística se utiliza en el área de Biología.

## **CAPÍTULO 2: FUNDAMENTOS DE LA ESTADISTICA DESCRIPTIVA**

### **2.1. Introducción a la estadística descriptiva**

Es la parte de la estadística que se ocupa de recolectar, clasificar, organizar y representación de un conjunto de datos; para poder analizarlos y así obtener conocimiento del comportamiento de las variables en estudio.

#### **2.1.1. Conceptos básicos**

**Población:** conjunto de individuos con propiedades comunes sobre los que se realiza una investigación de tipo estadístico.

**Muestra:** subconjunto de la población.

**Tamaño muestra:** número de individuos que forman la muestra.

**Muestreo:** proceso de obtención de muestras representativas de la población.

**Variable:** propiedad o cualidad que puede manifestarse bajo dos o más formas distintas en un individuo de una población.

**Modalidades, categorías o clases:** distintas formas en que se manifiesta una variable.

**Dato:** es el valor de una variable asociada a un elemento de la población o muestra.

**Censo:** es una investigación estadística que consiste en el recuento de la totalidad de los elementos que componen la población por investigar.

### **2.2. Clasificación de variables**

Las variables pueden adquirir diferentes valores o clasificarse en diferentes categorías según la naturaleza o tipo de estudio que se lleve a cabo. Entre éstas tenemos las siguientes clasificaciones.

### 2.2.1. Variables por su medición

**Variables cualitativas:** Son aquellas que se expresan en forma verbal como categorías, atributos o describen a un elemento de una población. (Johnson & Kuby, 2008, p.9)

#### **Ejemplo 1:**

- a) Sexo
- b) Color
- c) afiliación política
- d) nacionalidad
- e) motivación, área académica o profesión de una persona.

Las variables cualitativas pueden ser.

**Variables nominales:** son categorías o nombres que no guardan ningún orden.

#### **Ejemplo 2:**

- a) estado civil
- b) sexo
- c) lugar de residencia
- d) comida favorita
- e) color de ojos, etc.

**Variables ordinales:** son categorías o nombres que representan un orden. **Ejemplo**

#### **3:**

- a) nivel escolar (primaria, secundaria,..)
- b) días de la semana
- c) nivel de desnutrición (inferior, medio, superior)
- d) intensidad de un color
- e) grado de satisfacción, etc.

**Variables cuantitativas:** Son las que varían en términos de cantidad y se registran o expresan en forma numérica, es decir cuantifican un elemento de la población. (Johnson & Kuby, 2008, p.10)

**Ejemplo 4:**

- a) Edad
- b) promedio académico
- c) puntuaciones de exámenes
- d) frecuencia de delitos
- e) temperatura, ingresos anuales o salarios por hora.

Las variables cuantitativas pueden ser.

**Variables discretas:** son aquellas que sólo adquieren valores enteros.

**Ejemplo 5:**

- a) número de hijos
- b) número de bacterias
- c) cantidad de plaquetas
- d) cantidad de crías en una especie
- e) número habitantes en una casa, etc.

**Variables continuas:** son las que pueden asumir cualquier valor.

**Ejemplo 6:**

- a) la edad
- b) altura
- c) peso
- d) salario
- e) velocidad de un avión, etc.



### 2.2.2. Clasificación según escalas de medición

**Escala nominal:** Una variable puede ser tratada como nominal cuando sus valores representan categorías que no obedecen un orden, solo existe diferencia entre ellas. Son aquellas donde se clasifican los objetos o individuos en categorías cualitativas distintas y dentro de ella no es posible utilizar puntajes estrictamente numéricos. (Sampieri, 2010, p.214)

#### **Ejemplo 7:**

- a) con la variable sexo( uno = mujer, dos = hombre)
- b) religión (uno=católico, dos= protestante, ...)
- c) color de ojos (uno= azules, dos= marron,...)
- d) clasificación de animales por especie
- e) color de piel.

**Escala ordinal:** Una variable puede ser tratada como ordinal cuando sus valores representan categorías que obedecen un orden de rango o puntaje de mayor a menor. Son aquellas que clasifican a las personas, eventos u objetos con una relación a cierto atributo, pero sin indicar la distancia que hay entre las posiciones. (Sampieri, 2010, p.215)

#### **Ejemplo 8:**

- a) estado de salud (uno = no saludable, dos = poco saludable, tres = medio saludable y cuatro = saludable).
- b) tamaño de un objeto (uno= pequeño, dos= mediano, tres= grande).
- c) niveles de escolaridad (uno= preescolar, dos= educación básica, tres= educación media, cuatro= educación superior).
- d) calidad de servicio (uno= .malo, dos= regular, tres= bueno, cuatro= excelente).
- e) nivel de edificio (uno= primera planta, dos=segunda planta,...).

**Escala de intervalo:** Esta escala, mantienen las propiedades de las escalas ordinales y nominales, es decir, los números designan categorías y permiten ordenarlas; pero además permiten decir a qué distancia está una de otra, por medio de operaciones aritméticas para comparar o igualar entre intervalos. El cero es arbitrario, no es real, ya que se asigna arbitrariamente a una categoría el valor de cero y a partir de ésta se construye la escala; no indica la ausencia de atributo. (Sampieri, 2010, p.216)

**Ejemplo 9:**

- a) la temperatura Celsius donde cero es la temperatura a la que se congela el agua
- b) la altura de las ciudades utilizando como referencia el nivel del mar
- c) rendimiento académico medio en una escala de 0 a 20 donde cero
- d) años (2001, 2002,...),
- e) ubicación en la carretera respecto a un punto de referencia
- f) sobrepeso de una persona respecto a un patrón de comparación.

**Escala de razón:** Esta escala es similar a la escala de intervalo ya que posee todas sus características, solo que en esta el cero indica ausencia de atributo, el cero es real y es absoluto (no es arbitrario). Cero absoluto implica que hay un punto en la escala donde está ausente o no existe la propiedad medida. (Sampieri, 2010, p.216)

**Ejemplo 8:**

- a) la escala de temperatura kelvin donde cero indica la ausencia de calor
- b) distancia
- c) peso
- d) estatura
- e) tiempo requerido para realizar una tarea.

### 2.2.3. Variables por causa

**Variables independientes:** Es aquella característica o propiedad que se supone ser la causa del fenómeno estudiado. Es decir que son las características controladas por el investigador y que se supone tendrán efecto sobre otra variable. (Sampieri, 2014, p.131)

**Variables dependientes:** Es aquella propiedad o característica que se trata de cambiar mediante la manipulación de la variable independiente. Los cambios sufridos por los sujetos como consecuencia de la manipulación de la variable independiente por parte del experimentador. En este caso el nombre lo dice de manera explícita, va a depender de algo que la hace variar (Sampieri, 2014, p.131)

**Variables intervinientes:** Son aquellas características o propiedades que de una manera u otra afectan el resultado que se espera y están vinculadas con las variables independientes y dependientes (Sampieri, 2014, p.112)

## 2.3. Recopilación de datos

La recolectar los datos es elaborar un plan detallado de procedimientos que nos conduzcan a reunir datos con un propósito específico. (Sampieri, 2010, p.198)

### 2.3.1. Recolección indirecta

La base para el trabajo estadístico es decir los datos, se pueden obtener de manera indirecta, eso sucede cuando la información es proporcionada por instituciones u organizaciones; es decir, al utilizar la información de los censos nacionales, de algún tipo de organización o institución. Existen fuentes las cuales pueden ser:

➤ **Fuentes:**

- **Fuentes primarias:** cuando la persona o institución ha recolectado los datos.
- **Fuentes secundarias:** cuando la persona o institución no es la que ha efectuado la investigación.
- **Fuentes oficiales:** cuando los datos son dependencia gubernamental.

• **Fuentes privadas:** es cuando los datos son provistos por fuentes no gubernamentales.

### **2.3.2. Recolección directa**

Los datos también se pueden obtener de manera directa. Es decir cuando los datos se recolectan por medio de una investigación de campo o un experimento, esto se da por medio de llenando de fichas, cédulas de observación, cuestionarios, entrevistas, encuestas.

#### ***Pasos para la recolección directa.***

Cuando los datos se obtienen de manera directa, es necesario apoyarse en el muestreo, diseñar el instrumento de recopilación (cuestionario, encuesta o entrevista), coordinar la recopilación de datos y su procesamiento, para finalmente hacer el análisis de información y realizar el informe.

#### **Paso 1: Métodos de muestreo**

Existen dos tipos de muestreo: el muestreo no probabilístico y el muestreo probabilístico.

##### **a. - Muestreo no Probabilístico**

Es conocido también como muestreo no aleatorio o de juicio, en él interviene solamente la opinión personal del investigador, para identificar así los elementos de la población que deben incluirse en la muestra, evitando el análisis estadístico requerido para realizar muestras de probabilidad.

##### **a.1. - Muestreo por Juicio**

El investigador toma la muestra seleccionando los elementos que a él le parecen representativos o típicos de la población, por lo que depende del criterio del investigador. (Johnson & Kubby, 2008, p.21)

#### **Ejemplo 9:**

En un estudio en donde un investigador quiere saber qué es necesario para graduarse con los más altos honores en la universidad, las únicas personas que podrán

brindarle el mejor asesoramiento serán las personas que se graduaron con los más altos honores.

#### **a.2. - Muestreo Casual o Fortuito**

Se usa en los casos en donde no es posible seleccionar los elementos, y deben sacarse conclusiones con los elementos que estén disponibles. (Fidias Arias, 2012, p.85)

##### **Ejemplo 10:**

Se realiza una encuesta de opinión sobre un producto de higiene personal que se ha lanzado al mercado recientemente. La empresa contrata a una serie de entrevistadores que realizan las encuestas en la calle. Éstos van entrevistando a los que se van encontrando accidentalmente por la calle.

#### **a.3. - Muestreo de Cuota**

Se utilizan en estudios de opinión del mercado. Los enumeradores, reciben instrucciones de obtener cuotas específicas, a partir de las cuales se constituye una muestra relativamente proporcional a la población. (Fidias Arias, 2012, 86p)

##### **Ejemplo 11:**

En un estudio en donde el investigador quiere comparar el rendimiento académico de los diferentes niveles de clases del secundario, su relación con el género y la situación socioeconómica, el investigador identifica primero los subgrupos.

Por lo general, los subgrupos son las características o variables del estudio. El investigador divide a toda la población en niveles de clase, cruzados con el género y el nivel socioeconómico. Luego, toma nota de las proporciones de estos subgrupos en toda la población y a continuación hace un muestreo de cada subgrupo

#### **a.4. - Muestreo de Poblaciones Móviles**

Este tipo de muestreo utiliza métodos de captura, marca y recaptura. Se utiliza mucho en el estudio de migración de poblaciones de animales y otras características. (Lastra, Ramos & Ramos, 1995, p.7).

**Ejemplo 12:**

Se ha aplicado en forma probabilística para obtener estimaciones del volumen de los flujos migratorios de mexicanos en la frontera norte del país

**b. - Muestreo Probabilístico o Aleatorio**

Es un tipo de muestreo en donde se elige una muestra, de modo que cada integrante de una población tenga una probabilidad conocida de ser incluida en la muestra. Este muestreo se basa en datos previamente obtenidos, de los cuales se sacan conclusiones, y se clasifica en cuatro tipos:

**b.1. - Muestreo Aleatorio Simple**

Es un tipo de muestreo en donde intervienen las leyes de la probabilidad, y por tal motivo, cada elemento de la población tiene la misma posibilidad de ser elegido para formar parte de la muestra. (Johnson & Kuby, 2008, p.22)

En este muestreo existen dos modalidades:

**b.1.1. - Muestreo Aleatorio Simple sin Reposición**

Consiste en que una vez seleccionado un elemento de la población para formar parte de la muestra y, hecho el estudio correspondiente, dicho elemento no puede volver a formar parte de la población de origen. (Johnson & Kuby, 2008, p. 22).

**Ejemplo 13:**

Seleccionar a 50 de los 858 Centros de Enseñanzas Medias que existían en la Comunidad Autónoma Andaluza en el curso 1992/93 para llevar a cabo un estudio sobre este nivel educativo, y los hemos numerado del 1 al 858, no tendría sentido extraer una muestra en la que se repitiera algún Centro, puesto que en la práctica por cada repetición tendríamos un Centro menos del número que deseábamos seleccionar. En esta situación, cada Centro seleccionado debería ser excluido de la población antes de seleccionar a otro

### **b.1.2. - Muestreo Aleatorio Simple con Reposición**

Aquí, una vez que es seleccionado un elemento de la población para formar parte de la muestra y, ya anotadas sus características, dicho elemento regresa a la población de origen. (Johnson & Kuby, 2008, p.22).

#### **Ejemplo 14:**

Si en la población constituida por los 6 niveles de Educación Primaria queremos determinar la composición de un equipo de 4 alumnos de Educación Primaria, tendríamos que seleccionar una muestra de 4 elementos a partir de dicha población. Asignando a cada uno de los niveles papeletas con los números 1 al 6, los introduciríamos en una urna y extraeríamos cuatro papeletas. Pero si queremos contemplar la posibilidad de que en el equipo existan alumnos del mismo nivel, tendríamos que llevar a cabo un muestreo con reposición. Así, tras extraer el primer número volveremos a introducirlo en la urna, de forma que pueda ser objeto de nuevas extracciones. Muestras como {3,4,1,6}, {1,3,5,1}, {5,1,3,1} ó {2,6,2,2} en las que se repite algún elemento o cambia el orden de los mismos, se encontrarían entre las muestras ordenadas posibles.

### **b.2. - Muestreo Sistemático**

En este tipo de muestreo, los elementos que van a formar parte de la muestra, son seleccionados de manera ordenada, tomando en cuenta el tamaño de la población y el tamaño de la muestra. Se debe determinar un intervalo regular o una razón. Se elige un punto de partida, representando éste el primer elemento de la muestra, el cual es al azar y debe de estar comprendido o incluido en el valor de “k”. A este punto de partida, se le suma el valor de “k” o intervalo regular, obteniendo de esta manera el segundo elemento de la muestra, al cual se le suma nuevamente el valor de “k” para encontrar el tercero y así sucesivamente hasta completar el tamaño de la muestra. (Johnson & Kuby2008, p.23)

#### **Ejemplo 15:**

El investigador tiene una población total de 100 individuos y necesita 12 sujetos. Primero elige su número de partida, 5.

Luego, el investigador elige su intervalo, 8. Los miembros de su muestra serán los individuos 5, 13, 21, 29, 37, 45, 53, 61, 69, 77, 85, 93.

### **b.3. - Muestreo Estratificado**

El muestreo estratificado, se encarga de separar a la población en diferentes estratos o grupos, para después elegir una muestra aleatoria de cada uno, lo cual dará como resultado una muestra global. (Johnson & Kuby, 2008, p.24)

#### **Ejemplo 16:**

En una fábrica que consta de 600 trabajadores queremos tomar una muestra de 20. Sabemos que hay 200 trabajadores en la sección A, 150 en la B, 150 en la C y 100 en la D.

Para calcular que parte de cada grupo se necesita se utiliza la formula

$$\text{trabajadores por Sección} = \frac{\text{muestra}}{\text{poblacion}} * \text{total de cada seccion}$$

$$\text{Sección A} = \frac{20}{600} * 200 = 6.6$$

$$\text{Sección B} = \frac{20}{600} * 150 = 5$$

$$\text{Sección C} = \frac{20}{600} * 150 = 5$$

$$\text{Sección D} = \frac{20}{600} * 100 = 3.3$$

La muestra a tomar por cada sección es: de la sección A=7trabajadores, B= 5 trabajadores, C=5 trabajadores y D= 3 trabajadores.

### **b.4. - Muestreo por Conglomerado, de Racimo o de Agregados**

El muestreo por conglomerado es un tipo de muestreo aleatorio, en el cual la población se divide en grupos o conglomerados de elementos, para luego seleccionar una muestra aleatoria de esos grupos. Este tipo de muestreo su usa en áreas geográficas grandes, permitiendo así reducir el costo de muestrear una población dispersa (Johnson & Kuby, 2008, p.25).



### **Ejemplo 17:**

Se planea hacer una encuesta entre universitarios de primer curso de un país. Se quieren entrevistar 5.000 universitarios. Ante la imposibilidad de acceder (de acuerdo con los costos) a un muestreo estratificado, se piensa en una muestra de 200 grupos de 25 alumnos cada uno, luego se selecciona una muestra de cada grupo por medio del muestreo estratificado.

### **Paso 2: *Formula para conocer el tamaño de muestra***

Al conocer el tipo de muestra que se quiere utilizar dependiendo del tipo de investigación que se lleve a cabo, entonces se realiza el paso 2; con el cual teniendo el total de la población conoceremos el tamaño de la muestra que se necesitara para representar a la población.

Antes de calcular el tamaño de la muestra necesitamos determinar varias cosas:

**Tamaño de la población.** Una población es una colección bien definida de objetos o individuos que tienen características similares. Hablamos de dos tipos: población objetivo, que suele tener diversas características y también es conocida como la población teórica. La población accesible es la población sobre la que los investigadores aplicaran sus conclusiones.

**Margen de error (intervalo de confianza).** El margen de error es una estadística que expresa la cantidad de error de muestreo aleatorio en los resultados de una encuesta, es decir, la falla que se produce al extraer la muestra de la población. Generalmente, oscila entre 1% y 5%.

**Nivel de confianza.** Son intervalos aleatorios que se usan para acotar un valor con un buen nivel de confianza. Un intervalo de confianza de 95% el coeficiente es igual a 2, entonces el valor de zeta crítico es  $2^2 = 4$ ; significa que los resultados de una acción probablemente cubrirán las expectativas el 95% de las veces. para un nivel de confianza del 99% el coeficiente es igual a 3, y zeta crítico es igual a  $3^2 = 9$ .

Los datos usados en: (Arias, 2012)

Para población desconocida:

$$muestra = \frac{Z^2PQ}{e^2}$$

Para población conocida:

$$muestra = \frac{NZ^2PQ}{e^2(N - 1) + Z^2PQ}$$

Dónde:

$Z$  = nivel de confianza,

$P$  = probabilidad de éxito, o proporción esperada

$Q$  = probabilidad de fracaso  $Q = (P - 1)$

$e$  = precisión (error máximo admisible en términos de proporción)

$N$  = total de la población

**Ejemplo 18:** Se desea determinar el tamaño de la muestra para una población de 1000 profesores, con un nivel de confianza del 95%, un error del 5%, un valor de  $p = 40$  y  $q = 60$ .

**Solución:** Se utilizará la fórmula para población conocida, ya que conocemos el tamaño de la población la cual es de 1000 profesores.

$$\begin{aligned} muestra &= \frac{NZ^2PQ}{e^2(N - 1) + Z^2PQ} = \frac{1000 * (2)^2 * 40 * 60}{(5)^2 * (1000 - 1) + (2)^2 * 40 * 60} \\ &= \frac{9600000}{34575} \end{aligned}$$

$muestra = 277.657267$  Se aproxima a 278

Por lo tanto el tamaño de muestra para este estudio será de 278 profesores. (p.89)

### **Paso 3: Instrumentos y técnicas para la recolección de datos.**

(Arias, 2012). Se entenderá por técnica de investigación, el procedimiento o forma particular de obtener datos o información.

La aplicación de una técnica conduce a la obtención de información, la cual debe ser guardada en un medio material de manera que los datos puedan ser recuperados, procesados, analizados e interpretados posteriormente. A dicho soporte se le denomina instrumento. (p.63)

Un instrumento de recolección de datos es cualquier recurso, dispositivo o formato (en papel o digital), que se utiliza para obtener, registrar o almacenar información. Las técnicas de recolección son:

### **Fichas**

Las fichas en la investigación, nos pueden informar en dónde se encuentra cada una de las fuentes de información que utilizamos o los pasos seguidos en dicha investigación.

### **La entrevista**

(Arias, 2012) La entrevista, más que un simple interrogatorio, es una técnica basada en un diálogo o conversación “cara a cara”, entre el entrevistador y el entrevistado acerca de un tema previamente determinado, de tal manera que el entrevistador pueda obtener la información requerida.(p.73)

El instrumento utilizado es; cámaras fotográficas, cámaras de video y grabadoras de audio

### **La encuesta**

(Arias, 2012) Se define la encuesta como una técnica que pretende obtener información que suministra un grupo o muestra de sujetos acerca de sí mismos, o en relación con un tema en particular.

Las encuestas son empleadas frecuentemente en áreas específicas como la investigación de mercado, estudios del consumidor, encuestas electorales y estudios de opinión en general.

El instrumento utilizado es: las grabadoras y cámaras de video, de uso muy común en las encuestas que realizan en la calle los medios de comunicación. (p.72)

## **El cuestionario**

(Arias, 2012) Es la modalidad de encuesta que se realiza de forma escrita mediante un instrumento o formato en papel contentivo de una serie de preguntas. Se le denomina cuestionario autoadministrado porque debe ser llenado por el encuestado, sin intervención del encuestador. (p.74)

El instrumento utilizado puede ser: CD, papel, el internet, entre otros

Las preguntas de un cuestionario no se inventan a capricho. Éstas deben tener una correspondencia con los objetivos específicos de la investigación. Además, son producto de la operacionalización de las variables y de la definición de los indicadores. (Fidias Arias, 2012, p.75).

### **Operacionalización de variables**

La operacionalización de variables consistió en determinar el método a través del cual las variables serán medidas o analizadas.

Una definición operacional puede señalar el instrumento por medio del cual se hará la medición de la variable.

### **Pasos para operacionalizar las variables**

- Conocer el objetivo.
- Variables que responden a ese objetivo.
- Tipo de variable: puede ser cualitativa o cuantitativa
- Definición contextual: puede pensarse como la definición que nos da un diccionario de determinado concepto. A esta definición también se le suele denominar constitutiva.
- Indicador: está constituida por una serie de procedimientos o indicaciones para realizar la medición de una variable definida.
- Dimensiones de una variable: son subvariables o variables con un nivel más cercano al indicador.

- Escala de medición: la cual puede ser nominal, ordinal, de intervalo o de razón.
- Se formula la pregunta utilizando la información anterior.

### **Ejemplo 19:**

#### **Operacionalización de variable**

**Variable:** Edad

**Tipo de variable:** cuantitativa

**Definición contextual:** Cantidad de años, meses y días cumplidos a partir de la fecha de aplicación del estudio

**Indicador:** Calculo a partir de la fecha de nacimiento en un documento de identidad

**Dimensiones:** El número de años cumplidos

**Escala de medición:** De razón

**Instrumento:** Encuesta

**Pregunta:** ¿Edad?

Luego las preguntas se ordenan de lo general a lo particular y de encabezado se utiliza el logo de la institución y los datos respectivos de ella. (Fidias Arias, 2012, p.76).

## **2.4. Presentación de una sola variable**

Al realizar una recolección de datos ya sea de forma directa, por medio de un instrumento o de forma indirecta, se puede realizar el análisis de cada una de las variables de interés.

Las representaciones de los datos puede ser, a través de cuadros (distribución de frecuencia) y a través de gráficos (diagrama de barras, diagrama de pastel, histograma, entre otros) en este manual se explicaran los primeros tres que son los gráficos más utilizados.

### 2.4.1. Distribución de frecuencias

La Distribución de Frecuencias (Simples o agrupadas en intervalos), comúnmente llamada tabla de frecuencias, se utiliza para hacer la presentación de datos provenientes de las observaciones realizadas en el estudio ya sean de variables cualitativas o cuantitativas, estableciendo un orden mediante la división en clases y registro de la cantidad de observaciones correspondientes a cada clase. (Johnson & Kuby, 2008, p.56).

#### *Distribución de frecuencia para datos no agrupados*

La siguiente tabla presenta las principales características de una distribución de frecuencias simple o no agrupada.

Datos	Frecuencias Absolutas		Frecuencias Relativas	
	Simples	Acumuladas	Simples	Acumuladas
$x_1$	$n_1$	$N_1 = n_1$	$f_1 = n_1 / n$	$F_1 = N_1 / n$
$x_2$	$n_2$	$N_2 = n_1 + n_2$	$f_2 = n_2 / n$	$F_2 = N_2 / n$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
$x_k$	$n_k$	$N_k = \sum_{i=1}^k n_i$	$f_k = n_k / n$	$F_k = N_k / n$
Total	$\sum_{i=1}^k n_i = n$		$\sum_{i=1}^k f_i = 1$	

La primera columna representa los distintos valores de esos datos y la segunda la frecuencia simple, es decir, el número de veces que se ha observado el correspondiente valor; la tercera columna representa la frecuencia acumulada (número de veces que se han observado valores menores o iguales que el que corresponde a dicha fila). Las frecuencias relativas se obtienen a partir de las frecuencias absolutas, dividiendo por el tamaño de la muestra.

**Ejemplo 20:** En la Universidad de San Carlos en Guatemala, los estudiantes de Biología realizaron una investigación en un bosque de la región, con el objetivo de

conocer cuáles son los tipos de orquídeas y la cantidad de cada tipo de orquídeas que posee dicho bosque.

a) Realizar una distribución de frecuencia, con los siguientes datos:

TIPO DE ORQUIDEA	CANTIDAD
ARPOPHYLLUM	10
BRASSIA	20
CALANTHE	5
DICHAEA	30
EPIDENDRUM	5
ISOCHILUS	10

Solución:

**Tabla 1:** Distribución de frecuencia de tipos de orquídeas en un bosque.

TIPO DE ORQUIDEA	Frecuencias Absolutas		Frecuencias Relativas	
	Simples	Acumuladas	Simples	Acumuladas
ARPOPHYLLUM	10	10	1/8	1/8
BRASSIA	20	30	1/4	3/8
CALANTHE	5	35	1/16	7/16
DICHAEA	30	65	3/8	13/16
EPIDENDRUM	5	70	1/16	14/16
ISOCHILUS	10	80	1/8	1
<b>Total</b>	<b>80</b>		<b>1</b>	

Fuente: creación propia.

Dónde:

$n_i$  = Frecuencia Absoluta Simple

$N_i$  = Frecuencia Absoluta Acumulada

$f_i$  = Frecuencia Relativa simple

$F_i$  = Frecuencia Relativa Acumulada

Obtener las frecuencias absolutas de cada una de las categorías.

$n_1 = 10$   $n_2 = 20$   $n_3 = 5$   $n_4 = 30$   $n_5 = 5$   $n_6 = 10$

Frecuencia absoluta acumulada

$$N_1 = n_1 \quad N_2 = N_1 + n_2 \quad N_3 = N_2 + n_3 \quad N_4 = N_3 + n_4 \quad N_5 = N_4 + n_5$$

$$N_6 = N_5 + n_6$$

Calcular las frecuencias relativas y mostrar la distribución de frecuencias

$$f_1 = \frac{n_1}{N} = \frac{10}{80} = \frac{1}{8}, f_2 = \frac{n_2}{N} = \frac{20}{80} = \frac{1}{4}, f_3 = \frac{n_3}{N} = \frac{5}{80} = \frac{1}{16}, f_4 = \frac{n_4}{N} = \frac{30}{80} = \frac{3}{8}, f_5 = \frac{n_5}{N} = \frac{5}{80} = \frac{1}{16}, f_6 = \frac{n_6}{N} = \frac{10}{80} = \frac{1}{8}$$

Nótese que:

$$f_1 + f_2 + f_3 + f_4 + f_5 + f_6 = 1$$

Calcular frecuencia relativa acumulada

$$F_1 = f_1 \quad F_2 = F_1 + f_2 \quad F_3 = F_2 + f_3 \quad F_4 = F_3 + f_4 \quad F_5 = F_4 + f_5 \quad F_6 = F_5 + f_6$$

Interpretación de la tabla de distribución de frecuencia:

#### **Frecuencia Absoluta Simple (Fa):**

La interpretación de esta frecuencia es, decir cual observación es la que más se repitió; hasta llegar la observación que obtuvo menos frecuencia.

**Ejemplo 20.1:** Al realizar el recorrido en el bosque el tipo de orquídea que más se encontró fue la Dichaea ya que de 80 orquídeas que se encontraron, 30 fueron de este tipo y el tipo de orquídea que fue menos encontrado fueron dos tipos; Calanthe y Epidendrum ya que de cada uno se encontraron 5.

#### **Frecuencia Absoluta Acumulada (Faa):**

Esta frecuencia es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado.

**Ejemplo 20.2:** 30 orquídeas de las que se encontraron en el bosque son de tipo Arpophyllum y Brassia.

#### **Frecuencia Relativa Simple (Fr):**

Esta frecuencia indica el porcentaje que representa cierto dato de la población total.

**Ejemplo 20.3:** el 37.5 por ciento de las orquídeas encontradas en el bosque donde se realizó el estudio son de tipo Dichaea, el 6.25 por ciento son de tipo Calanthe y



el 6.25 por ciento son de tipo Epidendrum, estos dos tipos fueron las menos encontradas.

**Frecuencia Relativa Acumulada (Fra):**

Es la suma de las frecuencias relativas simples la cual nos presenta un porcentaje de los datos.

**Ejemplo 20.4:** el 37.5 por ciento de las orquídeas encontradas en el bosque de estudio son de tipo Arpophyllum y Brassia.

***Distribución de Frecuencia para Datos Agrupados***

Cuando en la muestra existen muchos valores cuantitativos diferentes y mucha variabilidad se recomienda, aún a costa de perder información, agrupar los datos en clases, en lo que se denomina distribución de frecuencias agrupada en intervalos.

Clase	Marca	Frecuencias Absolutas		Frecuencias relativas	
		Clase	Simple	Acumuladas	Simple
$L_0$				$N_1 = n_1$	
$< L_1$	$x_1$	$n_1$		$N_2 = n_1 + n_2$	$f_1 = n_1 / n$
$L_1$	$x_2$	$n_2$		.	$f_2 = n_2 / n$
$< L_2$	.	.		.	.
.	.	.		.	.
.	.	.		.	.
.	$x_k$	$n_k$		$N_k = \sum_{i=1}^k n_i$	$f_k = n_k / n$
$L_{k-1} < L_k$					$F_k = N_k / n$
Total			$\sum_{i=1}^k n_i = n$		$\sum_{i=1}^k f_i = 1$

Fuente: (Gorgas, Gardiel & Zamorano, 2011, p.15)

A cada uno de los intervalos se les denomina clase y al punto medio marca de clase. A efectos de cálculo la marca de clase se elige como representante del intervalo. El número de clases en que se dividen los datos no debe ser excesivo. A modo orientativo, el número de clases se puede obtener mediante la siguiente fórmula empírica, llamada de Sturges:

Número de clases =  $[ 1 + 3.322 \log(N) ]$  (aproximación)

**Ejemplo 10:** realizar una distribución de frecuencia para datos agrupados del tiempo de vida de una población de moscas.

Datos: tiempo de vida de 100 moscas:

26,15,14,7,5,0,17,4,19,13,19,7,19,12,23,19,8,18,10,19,0,17,5,11,23,2,14,11,22,26,  
5,19,14,28,2,9,8,1,23,25,17,19,16,24,7,28,22,10,24,4,11,2,15,22,19,14,0,21,14,22,  
10,8,17,0,1,21,22,16,17,21,1,30,2,9,21,21,30,25,8,15,30,21,11,25,0,6,8,11,0,21,29,  
23,27,28,27,11,15,9,20,28.

Solución:

**Paso 1: Número de clases.**

$N=100$ , valor inferior = 0, valor superior = 30

Numero de clases  $[ 1 + 3.322 \log(100) ] = 7.644$

Se tiene que aproximar, es decir 8 clases se forman

**Paso 2: Amplitud del intervalo.**

Rango= valor superior – valor inferior

Rango =  $30-0 = 30$

Ya conociendo el rango y el número de clases, se puede conocer la amplitud de cada intervalo.

Amplitud = rango / número de clases

Amplitud =  $30/8 = 3.75$  aproximaremos a 4

**Paso 3: Intervalos.**

Estos intervalos representan los días que vivieron 100 moscas

Clase
$0 < 4$
$4 < 8$
$8 < 12$
$12 < 16$
$16 < 20$
$20 < 24$
$24 < 28$
$28 < 32$

**Paso 4: Marca de clase (Pm)**

La marca de clase es: (límite superior + límite inferior) / 2; es decir su promedio.

Clase	Marca de clase
$0 < 4$	2
$4 < 8$	6
$8 < 12$	10
$12 < 16$	14
$16 < 20$	18
$20 < 24$	22
$24 < 28$	26
$28 < 32$	30

### Paso 21: Distribución de Frecuencia

La distribución de frecuencia para datos agrupados se trabaja de la misma forma que para datos simples (**Tabla 1**).

**Tabla 2:** Distribución de frecuencia del tiempo de vida de 100 moscas.

Clase	Marca de clase	Frecuencias Absolutas		Frecuencias Relativas	
		Simples	Acumuladas	Simples	Acumuladas
0 < 4	2	14	14	0.14	0.14
4 < 8	6	8	22	0.08	0.22
8 < 12	10	17	39	0.17	0.38
12 < 16	14	11	50	0.11	0.50
16 < 20	18	16	66	0.16	0.66
20 < 24	22	18	84	0.18	0.84
24 < 28	26	9	93	0.09	0.93
28 < 32	30	7	100	0.07	1
	Total	100		1	

Fuente: creación propia.

Interpretación de los datos de la tabla de distribución de frecuencia.

#### **Frecuencia Absoluta Simple (Fa):**

**Ejemplo 21.1:** Al observar el tiempo de vida de las moscas se puede decir, que el mayor grupo de moscas muere entre 20 y 24 días ya que de 100 moscas murieron 18 y el tiempo en que menos mueren las moscas fue de 28 a 32 días ya que se obtuvo que 7 moscas murieron en ese tiempo ya que es el tiempo máximo en que pueden vivir las moscas.

#### **Frecuencia Absoluta Acumulada (Faa):**

**Ejemplo 21.2:** 21 moscas tuvieron un periodo de vida entre 0 y 8 días

#### **Frecuencia Relativa Simple (Fr):**

**Ejemplo 21.3:** El 18 por ciento de las moscas mueren entre 20 y 24 días y el 7 por ciento mueren entre 28 y 32 días

### Frecuencia Relativa Acumulada (Fra):

**Ejemplo 21.4:** El 66 por ciento de las moscas mueren entre 0 y 20 días.

#### Utilizando el programa SPSS

Las distribuciones de frecuencia simples o con datos agrupados, se pueden realizar con el software SPSS puede ver se en: (Anexo 1.1: Distribución de frecuencia)

### 2.4.2. Presentación gráfica de datos

Para una mejor interpretación de los datos de una tabla de frecuencia se utilizan gráficos de acuerdo a la necesidad o al tipo de dato.

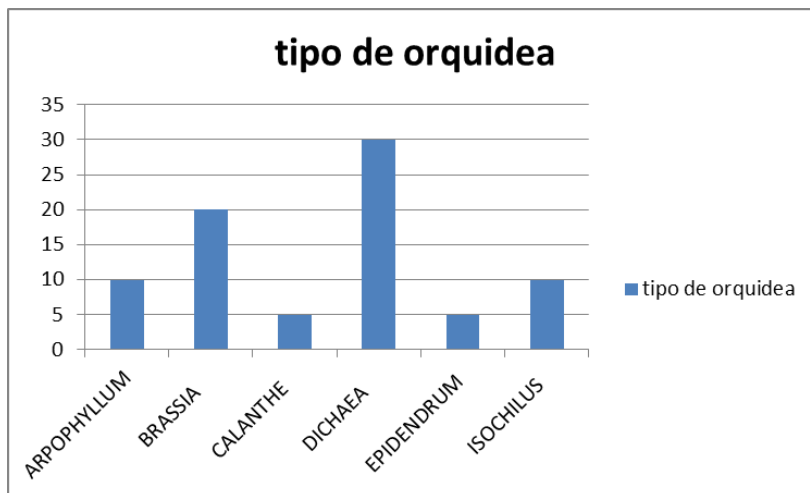
#### *Diagrama de barras*

Permite visualizar de forma sencilla la distribución de una variable. Se dibuja sobre cada categoría una barra (o rectángulo) cuya altura coincida con la frecuencia absoluta o relativa de dicha clase. (Johnson & Kuby, 2008, p.41).

**Ejemplo 22:** retomando el ejemplo 20. Realizar un diagrama de barras, utilizando la frecuencia absoluta simple.

Solución:

**Gráfico 1:** Diagrama de barras para la tabla 1



Interpretación: Al realizar el recorrido en el bosque el tipo de orquídea que más se encontró fue la Dichaea ya que de 80 orquídeas que se encontraron, 30 fueron de este tipo y el tipo de orquídea que fue menos encontrado fueron dos tipos ; Calanthe y Epidendrum ya que de cada uno se encontraron 5.

### **Gráfico de sectores**

En el caso de variables cualitativas el diagrama circular se utiliza con mucha frecuencia. Consiste en representar sobre un círculo los diferentes atributos, mediante un sector circular de ángulo proporcional a la correspondiente frecuencia. El ángulo de cada sector circular se calcula multiplicando por  $360^\circ$  la frecuencia relativa.

Si el número de categorías es excesivamente grande, la imagen proporcionada por el gráfico de sectores no es lo suficientemente clara y por lo tanto la situación ideal es cuando hay alrededor de cuatro categorías. En este caso se pueden apreciar con claridad dichos subgrupos. (Johnson & Kuby, 2008, p.41).

Para calcular que porcentaje le toca a cada uno se realiza de la siguiente manera:

Cada porción será  $= 360 * n_i/N$ .

Dónde:

$n_i =$  Frecuencia Absoluta Simple

$N =$  la sumatoria de la Frecuencia Absoluta Simple

**Ejemplo 23:** retomando el ejemplo 20. Realizar un diagrama de sectores.

Solución:

Para conocer cada parte del diagrama se debe utilizar la siguiente formula:

Cada porción será  $= 360 * n_i/N$ .

**grados del gráfico de sectores:**

$s_i =$  grados que tendrá la posición en el gráfico

$$S_1 = 360 \frac{n_1}{N} = 360 \frac{10}{80} = 45, \text{ representa el } 12.5\%$$

$$S_2 = 360 \frac{n_2}{N} = 360 \frac{20}{80} = 90, \text{ representa el } 25\%$$

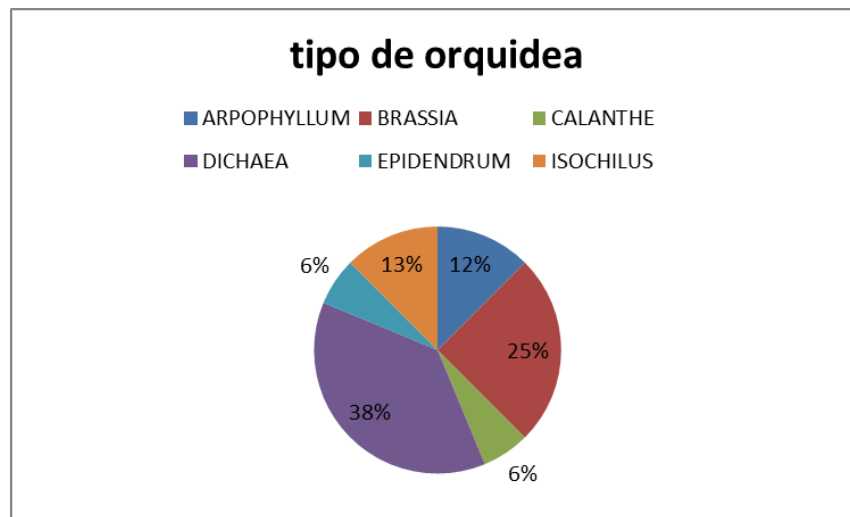
$$S_3 = 360 \frac{n_3}{N} = 360 \frac{5}{80} = 22.5, \text{ representa el } 6.25\%$$

$$S_4 = 360 \frac{n_4}{N} = 360 \frac{30}{80} = 135, \text{ representa el } 37.5\%$$

$$S_5 = 360 \frac{n_5}{N} = 360 \frac{5}{80} = 22.5, \text{ representa el } 6.25\%$$

$$S_6 = 360 \frac{n_6}{N} = 360 \frac{10}{80} = 45, \text{ representa el } 12.5\%$$

**Gráfico 2:** Diagrama de sectores para la tabla 1



Interpretación: El 37.5 aproximado a 38 por ciento de las orquídeas encontradas en el bosque de estudio son de tipo Dichaea y del tipo Brassia se encontró 6% al igual que del tipo Epidendrum.

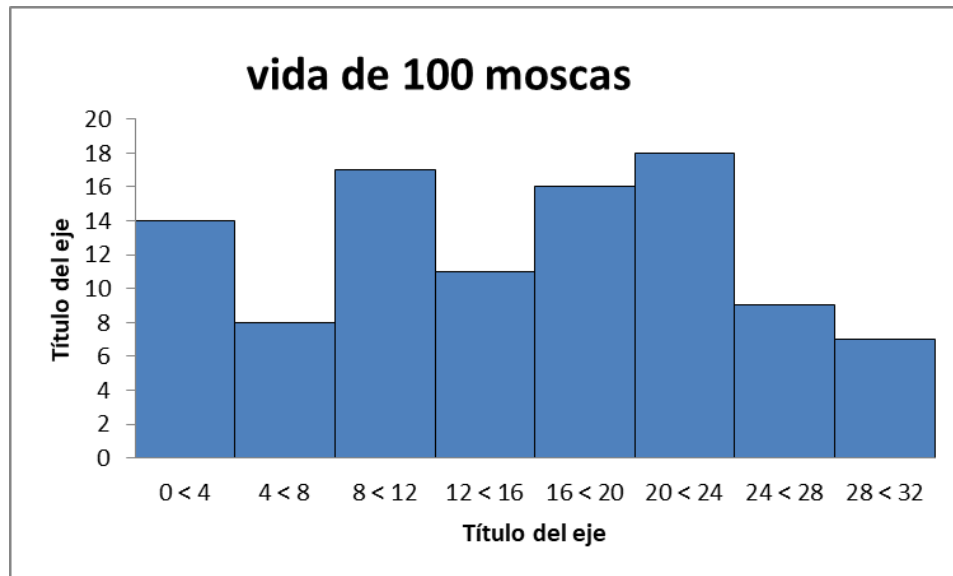
### ***Histograma***

Un histograma de frecuencias es un gráfico de barras verticales u horizontales. Describe la distribución de valores de tal forma que el área de cada barra es proporcional al número de objetos en la categoría o clase representada por la barra, es recomendable utilizarlo para datos agrupados. Así, un histograma de datos continuos sirve para el mismo fin que los gráficos de barras presentadas anteriormente. (Johnson & Kuby, 2008, p.59).

**Ejemplo 24:** utilizando el ejemplo 21, realizar un histograma.

Solución:

**Gráfico 3:** histograma para la tabla 2



Interpretación: Este gráfico muestra que las moscas mueren con más frecuencia entre 20 y 24 días, y con menor frecuencia mueren 28 a 32 días.

### Utilizando el programa SPSS

Los gráficos explicados teóricamente como lo son: el gráfico de barras, el gráfico de sectores y el histograma se pueden realizar con el programa informático SPSS.

Ver (Anexo 1.2: Creación de gráficos)

## **2.5. Análisis descriptivo de una sola variable**

Cuando los datos ya están organizados, se puede proceder con el análisis descriptivo de cada una de las variables de la investigación o el experimento.

### **2.5.1. Medidas de tendencia central**

Las medidas de tendencia central o medidas de posición proporcionan valores típicos o representativos de las variables en estudio.



### **Media aritmética**

Es el promedio de los datos en estudio. Y se calcula sumando cada uno de los valores y dividiendo entre el total de los datos; la media muestra se representa con  $\bar{x}$  “y la media poblacional con “ $\mu$ ”. (Johnson & Kuby, 2008, p.73)

Calculo de la media:

#### **Datos no agrupados:**

Si los datos no están ordenados en una tabla o se encuentran en tablas de frecuencia simple, se utilizará la formula siguiente:

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

**Ejemplo 25:** se tomaron la edad de 10 árboles en el cerro verde.

2, 4, 5, 5, 7, 4, 8, 9, 3, 1.

Encontrar la media de edades de estos 10 árboles.

Solución:

$$\bar{x} = \frac{(2+4+5+5+7+4+8+9+3+1)}{10} = \frac{48}{10} = 4.8$$

Interpretación: El promedio de estos 10 árboles del cerro verde es de 4.8 años.

#### **Datos agrupados:**

Si los datos están organizados en tabla de frecuencia de datos agrupados, se utilizará la siguiente formula:

$$\mu = \bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n}$$

Donde los  $f_i$  son la frecuencia absoluta simple.

**Ejemplo 26:** se retomará el ejemplo 21. Conocer el promedio de vida de las moscas.

Solución:

Se utilizará la formula dada anteriormente, con los datos de la tabla 2 del ejemplo 21.

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n}$$

$$\bar{x} = \frac{(2*14)+(6*8)+\dots+(26*9)+(30*7)}{100} = \frac{1528}{100} = 15.28$$

Interpretación: el promedio de vida de una mosca es de 15.28 días

### **Mediana**

(Peña, 2005) La mediana es el valor medio tal que, al ordenar los datos de menor a mayor, el 50% de los datos es menor que ella y el otro 50% es mayor que ella.

(p.60)

Calculo de la mediana:

#### **Datos no agrupados:**

Cuando los datos son pares: la mediana será el término medio de los valores centrales. La fórmula es la siguiente:

$$\text{Posición de la mediana } Md = \frac{n+1}{2}$$

Cuando los datos son impares: en este caso la media coincide con el dato central. La fórmula es la siguiente:

$$\text{Posición de la mediana } Md = \frac{n+1}{2}$$

**Ejemplo 27:** Se conoce que, en 6 zoológicos de los distintos países de Centro América, se tienen 100, 500, 200, 600, 250 y 150 especies en cada uno respectivamente. ¿Cuál es la mediana de especies de estos zoológicos?

Solución:

Se ordenan los datos de menor a mayor:

100, 150, 200, 250, 500, 600

$$\text{Posición de la mediana } Md = \frac{6+1}{2} = \frac{7}{2} = 3.5$$

La mediana se encuentra entre la tercera y cuarta posición, por lo que se suman ambos valores y se dividen entre dos

$$Md = \frac{200+250}{2} = \frac{450}{2} = 225$$

El valor de Md = 225 especies

Interpretación: el 50% de los zoológicos de Centro América tienen menos de 225 especies y el otro 50% tiene más de 225 especies.

**Datos agrupados:**

Para el caso de distribuciones agrupadas en intervalos se utiliza una fórmula aproximada. Se comienza calculando el intervalo donde se encuentra la mediana, para lo cual se calculan las frecuencias acumuladas, y se escoge el primer intervalo cuya frecuencia acumulada sea igual o superior a  $n/2$ . (Gorgas, Gardiel & Zamorano, 2011, p.27). A continuación se aplica la fórmula:

$$Mediana = L_i + \left( \frac{\frac{n}{2} - F_{i-1}}{F_i} \right) C$$

Dónde:

$L_i$  = Es el límite inferior de la clase donde se encuentra la mediana.

$F_{i-1}$  = Es la frecuencia acumulada anterior a la clase mediana.

$F_i$  = Es la frecuencia absoluta del intervalo mediano.

$C$  = Es la amplitud de los intervalos.

**Ejemplo 28:** Encontrar la mediana utilizando los datos del tiempo de vida de 100 moscas (Tabla 2 del ejemplo 21).

Solución:

Se calculará la mediana aproximada.

Frecuencia acumulada mayor o igual a:  $n/2$

$$100/2 = 50$$

Se trabajará con estos datos:

La posición de la tabla donde se encuentra la frecuencia acumulada igual o mayor a 50 es  $i = 4$  en la frecuencia absoluta simple.

$$F_{i-1} = 39$$

$$L_i = 12$$

$$C = 4(\text{amplitud})$$

$$F_i = 11$$

$$\text{Mediana} = L_i + \left( \frac{\frac{n}{2} - F_{i-1}}{F_i} \right) C$$

$$\text{Mediana} = 12 + \left( \frac{50 - 39}{11} \right) * 4 = 12 + 4 = 16$$

Interpretación: El 50% de las moscas muere en menos de 16 días y el otro 50% muere después de 16 días

### ***Moda***

La moda es una medida de tendencia central que indica cual es la puntuación, categoría o modalidad que más se repite. Se define como aquel valor de la variable o del atributo que presenta la mayor frecuencia.

Si se tiene un atributo o una variable con máxima frecuencia, la distribución es unimodal. Si hay dos valores en la variable con la misma frecuencia máxima, la distribución es bimodal. Si hay más de dos, la distribución es multimodal. Cuando ninguno de los valores que toma la variable se repite, no existe moda. (Johnson & Kuby, 2008, p.76)

Calculo de la moda:

### **Datos no agrupados:**

Teniendo una serie de datos, se ordenan y se calcula la frecuencia de cada uno de ellos y luego se verifica cuál de los datos tiene más frecuencia y ese dato será la moda del estudio.

**Ejemplo 29:** Haciendo uso de la tabla 1 del ejemplo 20. Encontrar la moda

Solución:

Ya teniendo ordenados los datos en tabla se verifica cual tipo de orquídea tiene la mayor frecuencia.

Interpretación: el tipo de orquídea que más se encontró en el bosque en el cual se está estudiando es: Dichaea, ya que este tipo tiene la frecuencia más alta.

**Datos agrupados:**

Cuando los datos están agrupados en intervalos para conocer la moda se requiere utilizar una formula ya que, no es fácil de conocer a simple vista. (Gorgas, Gardiel & Zamorano, 2011, p.29). . La fórmula es la siguiente:

$$Moda = L_i + \left( \frac{F_i - F_{i-1}}{(F_i - F_{i+1}) + (F_i - F_{i-1})} \right) C$$

Dónde:

$L_i$  = límite inferior del intervalo que presenta mayor frecuencia.

$F_i$  = Frecuencia absoluta maxima.

$F_{i+1}$  = Frecuencia absoluta del intervalo posterior al modal.

$F_{i-1}$  = Frecuencia absoluta del intervalo anterior al modal.

$C$  = la amplitud del intervalo.

**Ejemplo 30:** Haciendo uso de la tabla 2 del ejemplo 21. Encontrar la moda para datos agrupados.

Solución:

Teniendo la tabla de datos agrupados, se verifica cual es el intervalo con mayor frecuencia y con ese se trabajará; para aplicar la formula.

$$Moda = L_i + \left( \frac{F_i - F_{i-1}}{(F_i - F_{i+1}) + (F_i - F_{i-1})} \right) C$$

La posición de la tabla donde se encuentra la mayor frecuencia es  $i = 6$  en la frecuencia absoluta simple.

$$L_i = 20$$

$$F_i = 18$$

$$F_{i+1} = 9$$

$$F_{i-1} = 16$$

$$C = 4$$

$$Moda = 20 + \left( \frac{18-16}{9+2} \right) *4=20+0.727273=20.727273$$

Interpretación: los días en que más moscas mueren es a los 20.727273 días.

### 2.5.2. Medidas de posición

Estas medidas descriptivas permiten ubicar la posición que ocupa un valor dentro de un conjunto de datos.

#### *Cuartiles*

Son valores que dividen los datos ordenados en cuatro partes iguales. Existen tres cuartiles, por lo tanto, dicha medida hace referencia a un porcentaje de casos por debajo del cuartil y otro porcentaje por encima del cuartil, El primer cuartil, es un número tal que a lo sumo 25% de datos son menores en valor que él y a lo suma 75% son mayores. El segundo cuartil es la mediana. El tercer cuartil, es un número tal que a lo sumo 75% de los datos son menores en valor que él y a lo sumo 25% son mayores (Johnson & Kubly, 2008, p.93).

Fórmula para calcular los cuartiles:

#### **Para datos no agrupados:**

Se ordenan los datos de menor a mayor, luego se localiza el cuartil deseado mediante las siguientes formulas:

**Posición del primer cuartil:**  $1 * (n + 1)/4$  representa el 25% de los datos.

**Posición del segundo cuartil:** el mismo que la mediana. Representa el 50% de los datos.

**La posición del tercer cuartil:**  $3 * (n + 1)/4$  representa el 75% de los datos.

**Ejemplo 31:** se tomaron la edad de 10 árboles en el cerro verde 2, 4, 5, 5, 7, 4, 8, 9, 3, 1. ¿Encontrar los 3 cuartiles?

Solución:

Primeramente, se ordenarán los datos de menor a mayor.

**Tabla 3:** Datos de la edad de 10 árboles en el cerro verde.

Dato	1	2	3	4	4	5	5	7	8	9
Posición	1	2	3	4	5	6	7	8	9	10

**Primer cuartil:**

$Q_1 = 1*(n+1)/4 = 1*(10+1)/4 = 2.75$  se tomara el entero próximo que es 3

Interpretación: el 25% de los árboles en el cerro verde tienen menos de 3 años de edad

**Segundo cuartil:**

$Q_2 = 2*(n+1)/4 = 2*(10+1)/4 = 5.5$  por ser el número de datos par se saca el promedio de los dos números de la posición 5 y 6.  $(4 + 5)/2 = 5.5$

Interpretación: el 50% de los árboles en el cerro verde tienen menos de 5.5 años de edad

**Tercer cuartil:**

$Q_3 = 3*(n+1)/4 = 3*(10+1)/4 = 8.25$  se tomara la parte entera porque es menor que 8.5 que sería 8

Interpretación: el 75% de los árboles en el cerro verde tienen menos de 7 años de edad

**Para datos agrupados:**

Cuando los datos están agrupados en intervalos se utiliza la siguiente fórmula:

$$Q_{j.4} = L_i + \left( \frac{\frac{j * n}{4} - F_{i-1}}{F_i} \right) C \text{ Donde } (j = 1, 2, 3)$$

Dónde:

$Q_{j.4}$  = Cuartil deseado.

$j$  = Cuartil deseado.

$n$  = Suma de los datos.

$F_i$  = Frecuencia absoluta simple.

$F_{i-1}$  = Frecuencia absoluta acumulada anterior.

$L_i$  = Límite inferior del intervalo de la posición del cuartil.

$C$  = Amplitud del intervalo.

Para poder utilizar esta fórmula primero se conoce en donde está ubicado el cuartil deseado.

Posición del  $Q_1 = \frac{N}{4}$ , posición del  $Q_2 = \frac{N*2}{4}$ , posición del  $Q_3 = \frac{N*3}{4}$

Ya conociendo la posición se debe buscar la frecuencia absoluta acumulada mayor o igual que el valor de la posición y con esos datos se trabajará.

**Ejemplo 32:** Utilizando la tabla 2 del ejemplo 21. Encontrar el intervalo en que mueren el 75% de las moscas.

Solución:

Lo que se quiere saber el 75%, esto quiere decir encontrar el tercer cuartil para datos agrupados; utilizaremos la fórmula antes mencionada

$$Q_{j.4} = L_{i-1} + \left( \frac{\frac{j \cdot n}{4} - F_{i-1}}{F_i} \right) C_i \text{ Donde } (j = 1, 2, 3)$$

Encontrar la posición:  $(3 \cdot 100) / 4 = 75$ .

Luego se ubica la posición encontrada, en la frecuencia simple acumulada (tabla 2 del ejemplo 9).

Los datos son los siguientes:

$$n = 100$$

$$F_{i-1} = 66$$

$$F_i = 18$$

$$L_{i-1} = 20$$

$$C = 4$$

$$Q_3 = 20 + \left( \frac{\frac{3 \cdot 100}{4} - 66}{18} \right) * 4 = 20 + 2 = 22$$

Interpretación: el 75% de las moscas viven menos de 22 días.

### ***Deciles***

Son valores que dividen los datos ordenados en diez partes iguales. Existen nueve deciles, dicha medida deja un porcentaje de datos por debajo del decil y otro porcentaje por encima. Entre dos deciles consecutivos cualesquiera se encuentran un 10% o una décima partes de los elementos. (Spiegel & Stephens, 2009, p.66)

Fórmula para calcular los deciles:



### Para datos no agrupados:

Se ordenan los datos de menor a mayor, luego se localiza el decil deseado mediante las siguientes formulas:

**Posición del decil:**  $i * (n + 1)/10$ , donde

$i$ = al decil deseado

$n$ = al total de los datos

**Ejemplo 33:** utilizando la tabla 3 del ejemplo 31. ¿Encontrar el tercer decil?

Solución:

$$D_3 = 3 * (n + 1)/10 = 3 * (10)/10 = 3$$

Interpretación: el 30% de los árboles en el cerro verde tienen menos de 3 años de edad

### Para datos agrupados

Cuando los datos están agrupados en intervalos se utiliza la siguiente formula:

$$D_{j.10} = L_{i-1} + \left( \frac{\frac{j * n}{10} - F_{i-1}}{F_i} \right) C, (j = 1, 2, 3, 4, 5, 6, 7, 8, 9)$$

Dónde:

$D_{j.10}$ = Decil deseado.

$j$ = decil deseado.

$n$ =suma de datos.

$F_i$ =frecuencia absoluta simple.

$F_{i-1}$ = frecuencia absoluta acumulada anterior.

$L_i$  = límite inferior del intervalo de la posición del decil.

$C$ = amplitud del intervalo.

Para poder utilizar esta fórmula primero se conoce en donde está ubicado el cuartil deseado.

**Ejemplo 34:** utilizando la tabla 2 del ejemplo 21. Encontrar en que día ya ha muerto el 70% de las moscas.

Solución:

Lo que se quiere saber es: 70%, esto quiere decir encontrar el séptimo decil. Se utilizará la siguiente fórmula.

$$D_{j.10} = L_i + \left( \frac{\frac{j * n}{10} - F_{i-1}}{F_i} \right) C_i \text{ Donde } k = 10 \text{ (j = 1, 2, 3, 4, 5, 6, 7, 8, 9)}$$

Encontrar la posición:  $j * n / 10 = (7 * 100) / 10 = 70$ .

Luego se ubica la posición encontrada, en la frecuencia simple acumulada (tabla 2 del ejemplo 9).

Los datos son los siguientes:

$$n = 100$$

$$F_{i-1} = 66$$

$$F_i = 18$$

$$L_i = 20$$

$$C = 4$$

$$D_7 = 20 + \frac{\frac{7 * 100}{10} - 66}{18} * 4 = 20 + .889 = 20.889$$

Interpretación: el 70% de las moscas viven menos de 20.889 días

### ***Percentiles***

Son aquellos valores que dividen los datos ordenados en cien partes iguales. Existen noventa y nueve percentiles, dicha medida hace referencia a un porcentaje de casos por debajo del percentil y otro porcentaje por encima. Entre dos percentiles consecutivos cualesquiera se encuentran un 1% o una centésima parte de los datos. (Johnson & Kuby, 2008, p.93).

Fórmula para calcular los percentiles:

#### **Para datos no agrupados:**

Se ordenan los datos de menor a mayor. Y luego se localiza el percentil deseado mediante las siguientes fórmulas:

**Posición del percentil:**  $i * (n + 1) / 100$ , donde

$i$  = percentil deseado

$n$  = al total de los datos

**Ejemplo 35:** Utilizando la tabla 3 del ejemplo 31. ¿Encontrar el percentil 30?

Solución:

$$P_{30} = 30*(n)/100 = 30*(10)/100 = 3$$

Interpretación: el 30% de los árboles en el cerro verde tienen menos de 3 años de edad

**Para datos agrupados:**

Cuando los datos están agrupados en intervalos se utiliza la siguiente fórmula:

$$P_{j,100} = L_i + \left( \frac{\frac{j * n}{100} - F_{i-1}}{F_i} \right) C, \text{ donde } (j = 1, 2, 3, \dots, 99)$$

Dónde:

$P_{j,100}$  = percentil a encontrar

$j$  = percentil deseado

$n$  = suma de datos

$F$  = frecuencia absoluta simple

$F_{-}$  = frecuencia absoluta acumulada anterior

$L_i$  = límite inferior del intervalo de la posición del percentil

$C$  = amplitud del intervalo

Para poder utilizar esta fórmula primero se conoce en donde está ubicado el percentil deseado.

**Ejemplo 36:** Utilizando la tabla 2 del ejemplo 21. Encontrar en que día ya ha muerto el 70% de las moscas.

Solución:

Lo que se quiere saber es 70%, esto quiere decir encontrar el percentil 70, utilizando la fórmula.

$$P_{j,100} = L_i + \left( \frac{\frac{j * n}{100} - F_{i-1}}{F_i} \right) C \text{ Donde } k = 100 (j = 1, 2, 3, \dots, 99)$$

Encontrar la posición:  $(70*100)/100 = 70$ .

Luego se ubica la posición encontrada en la frecuencia simple acumulada (tabla 2 del ejemplo 9).

Se tienen los siguientes datos:

$$n=100$$

$$F_{i-1}=66$$

$$F_i=18$$

$$L_i=20$$

$$C=4$$

$$P_{70} = 20 + \frac{\frac{70 \times 100}{100} - 66}{18} * 4 = 20 + = 20.889$$

Interpretación: el 70% de las moscas viven menos de 20.889 días

### 2.5.3. Medidas de dispersión.

Las medidas de dispersión estudian la separación existente entre los diversos valores que toma la variable.

#### *Varianza*

La varianza mide la mayor o menor dispersión de los valores de la variable respecto a la media aritmética. Una varianza “grande” es indicativa de mayor dispersión y por lo tanto, menor representatividad tendrá la media aritmética, mientras que una varianza “pequeña” indica menor dispersión y por lo tanto mayor representatividad tendrá la media aritmética de los datos. Viene definida como la media de las diferencias cuadráticas de las puntuaciones respecto a su media aritmética. (Anderson & Sweeney & Williams, 2008, p.93)

La fórmula para encontrar la varianza poblacional es:

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^K (x_i - \mu)^2$$

En muchas situaciones es preciso estimar la varianza de una población a partir de una muestra.

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

**Para datos no agrupados:**

La varianza poblacional

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 F_i$$

La varianza muestral

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

Dónde:

$x_i$  = Cada uno de los datos

$\bar{x}$  = La media muestral

$\mu$  = La media poblacional

$n$  = Total de datos

$F_i$  = la frecuencia absoluta simple de cada dato

**Ejemplo 37:** se tomaron la edad de 5 árboles en el cerro verde 2, 4, 5, 7, 8. Encontrar la varianza.

Solución:

Encontrar la media de los datos

$$\bar{x} = \frac{2+4+5+7+8}{5} = \frac{26}{5} = 5.2$$

Luego se sustituye en la formula

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

$$S_{n-1}^2 = \frac{1}{5} * ((2 - 5.2)^2 + \dots + (8 - 5.2)^2) = \frac{1}{5} * (22.8) = 4.56$$

Interpretación: se puede decir que la media es poco representativa ya que el valor esta grande.

**Para datos agrupados:**

Varianza población:

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 F_i$$

Varianza muestral:

$$S_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 F_i$$

Donde  $n_i$  es la frecuencia.

**Ejemplo 38:** Utilizando la tabla 2 del ejemplo 21. Encontrar la varianza

Solución:

Se debe encontrar la media para datos agrupados. (Véase el ejemplo 26).

$$\bar{x} = 15.28$$

Luego se encontrará la varianza con la fórmula:

$$S_{n-1}^2 = \sum_{i=1}^n (x_i - \bar{x})^2 F_i$$

$$S_{n-1}^2 = \frac{1}{100} * (((2 - 15.28)^2 * 14) + \dots + ((30 - 15.28)^2 * 7))$$

$$S_{n-1}^2 = \frac{7132.16}{100-1} = 72.0420$$

Interpretación: esto indica que la media no es representativa ya que el valor es muy alto.

### ***Desviación típica***

La desviación típica, en un conjunto de datos, es una medida de dispersión, que nos indica cuánto pueden alejarse los valores respecto a la media aritmética. Está definida como la raíz cuadrada de la varianza, denotada por  $S$  cuando es muestral y por  $\sigma$  cuando es poblacional y se usa para una mejor comprensión pues esta expresada en las mismas unidades que la variable en estudio. (Anderson, sweeney & Williams, 2008, 95p)

La fórmula para calcular la desviación típica muestral es:

$$S = \sqrt{\text{varianza}} = \sqrt{S^2}$$

Fórmula para la desviación típica poblacional:

$$\sigma = \sqrt{\text{varianza}} = \sqrt{\sigma^2}$$

**Ejemplo 39:** Retomando el ejemplo 38. Encontrar la desviación típica.

Solución:

Conociendo que la varianza es:

$$S^2 = 72.0420$$

Para conocer la desviación solo se le aplica raíz cuadrada.

$$S = \sqrt{S^2} = \sqrt{72.0420} = 8.4878$$

Interpretación: en promedio, la cantidad de días para que muera cada una de las moscas del estudio se separa de la media, 8.4878 (días).

### ***Coefficiente de variabilidad***

El coeficiente de variación es una medida relativa de variabilidad. En ocasiones puede interesar comparar la dispersión de dos muestras y la desviación típica no ser válida, si las dos muestras tienen unidades diferentes. Para obviar este inconveniente se define el coeficiente de variación; si este coeficiente es mayor que 1.5, conviene investigar posibles fuentes de heterogeneidad (medidas con distinto instrumento: personas con distinto sexo, distintos tiempos...). (Anderson, Sweeney & Williams, 2008, p.95).

La fórmula para calcular el coeficiente de variación:

Coeficiente de variación poblacional:

$$CV = \frac{\text{Desviación Típica}}{|\text{Media}|} = \frac{\sigma}{|\mu|} * 100\%$$

Coeficiente de variación muestral:

$$CV = \frac{\text{Desviación Típica}}{|\text{Media}|} = \frac{S}{|\bar{x}|} * 100\%$$

Grado de representatividad de la media aritmética, mediante el coeficiente de variación tanto poblacional como muestral:

Valor del CV Grado en que la "  $\bar{x}$  o  $\mu$  " representa a la serie:

De 0% a menos de 10% Altamente representativa  
 De 10% a menos de 20% Bastante representativa  
 De 20% a menos de 30% Tiene representatividad  
 De 30% a menos de 40% Representación dudosa  
 De 40% o más Media carente de representatividad

**Ejemplo 40:** retomando el ejemplo 39. Encontrar el coeficiente de variación

$$CV = \frac{\text{Desviación Típica}}{|\text{Media}|} = \frac{S}{|\bar{x}|}$$

$$CV = \frac{8.4878}{|15.28|} = 0.55548$$

Interpretación: A mayor valor del coeficiente de variación mayor heterogeneidad de los valores de la variable; y a menor C.V., mayor homogeneidad en los valores de la variable. Es decir que los datos recolectados en la investigación de tiempo de vida de una mosca presentan mayor homogeneidad ya que el valor es pequeño.

#### 2.5.4. Medidas de forma

Este tipo de medidas permite conocer la forma de la distribución sin necesidad de recurrir a su representación gráfica. Existen dos tipos de medidas de forma: Asimetría y curtosis.

##### *Coefficiente de asimetría*

Una variable se dice que es simétrica si al “doblar” la distribución respecto a un eje (centro de simetría), las frecuencias coinciden. Caso contrario se dice que la distribución es asimétrica positiva o negativa, dependiendo que la rama larga de la distribución se encuentra en el sentido positivo o negativo del eje de las x. La medida más popular de asimetría es el coeficiente de asimetría de Fisher. (Spiegel & Stephens, 2009, p.125)

La fórmula para encontrar el coeficiente de asimetría es:

$$CA = \frac{m_3}{S^3} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 F_i}{\left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 F_i \right]^{3/2}}$$



Este coeficiente es invariante frente a cambios de origen y de escala; cuya interpretación es:

Si  $CA = 0$ , la distribución es simétrica.

Si  $CA > 0$ , la distribución es asimétrica positiva

Si  $CA < 0$ , la distribución es asimétrica negativa

**Ejemplo 41:** Utilizando la tabla 2 del ejemplo 21. Encontrar el coeficiente de asimetría.

Solución:

Se encontrara el  $m_3$  y el  $S^3$ , para luego encontrar el coeficiente de asimetría.

$$m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 F_i$$

$$m_3 = \frac{1}{100} * (((2 - 15.28)^3 * 14) + \dots + ((30 - 15.28)^3 * 7))$$

$$m_3 = \frac{-2509.2096}{100} = -25.092096$$

$$S^3 = \left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 F_i \right]^{3/2}$$

$$S^3 = \left[ \frac{1}{100} * (((2 - 15.28)^2 * 14) + \dots + ((30 - 15.28)^2 * 7)) \right]^{3/2}$$

$$S^3 = \left[ \frac{7132.16}{100} \right]^{3/2} = 602.3260$$

Luego se aplica la formula

$$CA = \frac{m_3}{S^3} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 F_i}{\left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 F_i \right]^{3/2}}$$

$$CA = \frac{-25.092096}{602.3260} = -0.04166$$

Interpretación: la asimetría es negativa es decir tiende a la izquierda, lo que significa que las moscas tardan varios días en morir.

### ***Curtosis***

El coeficiente de curtosis trata de estudiar la concentración de frecuencias en la zona central de la variable, de modo que variables con curtosis alta (leptocúrticas) tienen forma alargada y variables con curtosis baja tienen forma aplanada (platicúrticas). La comparación se realiza respecto a una distribución “moderada” como es la distribución normal (mesocúrtica). (Spiegel & Stephens, 2009, p.125).

La fórmula viene dada de la siguiente manera.

$$Curtosis = \frac{m_4}{S^4} - 3 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 F_i}{\left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 F_i \right]^2} - 3$$

Que se interpreta del siguiente modo:

Si  $Curtosis = 0$ , la distribución es mesocúrtica o normal.

Si  $Curtosis > 0$ , la distribución es leptocúrtica o por encima de lo normal.

Si  $Curtosis < 0$ , la distribución es platicúrtica o por debajo de la normal.

Al igual que el coeficiente de asimetría de Fisher, el coeficiente de curtosis es invariante frente a cambios de origen y de escala.

**Ejemplo 42:** Utilizando la tabla 2 del ejemplo 21. Encontrar la curtosis.

Solución:

Se encontrara el  $m_4$  y el  $S^4$ , para luego encontrar el coeficiente de asimetría.

$$m_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 F_i$$

$$m_4 = \frac{1}{100} * (((2 - 15.28)^4 * 14) + \dots + ((30 - 15.28)^4 * 7))$$

$$m_4 = \frac{993090.732}{100} = 9930.90732$$

$$S^4 = \left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 F_i \right]^2$$

$$S^4 = \left[ \frac{1}{100} * (((2 - 15.28)^2 * 14) + \dots + ((30 - 15.28)^2 * 7)) \right]^2$$

$$S^4 = \left[ \frac{7132.16}{100} \right]^2 = 5086.770627$$

Luego se aplica la formula

$$Curtosis = \frac{m_4}{S^4} - 3 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 F_i}{\left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 F_i \right]^2} - 3$$

$$Curtosis = \frac{9930.90732}{5086.770627} - 3 = -1.047699$$

Interpretación: la curtosis es negativa, por lo tanto, la distribución es platicúrtica o por debajo de la normal. Ya que hay un intervalo en el que las moscas mueren con más frecuencia.

### Utilizando el programa SPSS

El análisis descriptivo que está compuesto por todas las medidas que se le pueden realizar a una variable también se pueden realizar de una manera más rápida en el programa SPSS. Ver (Anexo 1.3: Análisis descriptivo)

## **2.6. Presentación y análisis de dos variables**

Teniendo una población y un conjunto de “n” individuos, se desea medir dos variables X e Y.

### 2.6.1. Tablas de contingencia

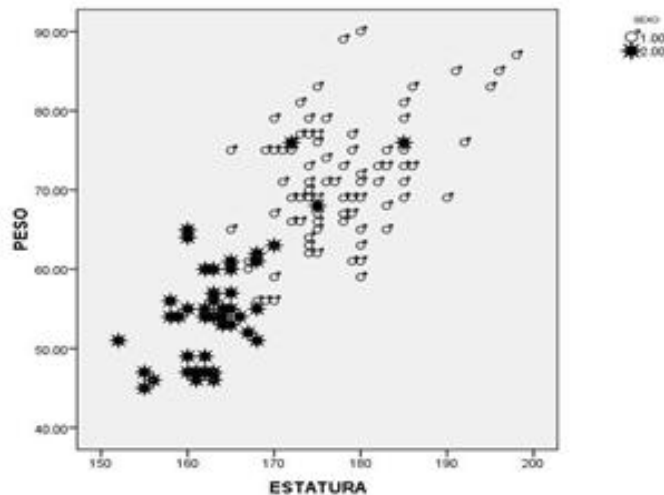
Muy a menudo los datos de estudio estadístico suelen tener variables bidimensionales, donde los valores de X e Y pueden venir dados en intervalos, y esto serian tablas de doble entrada o tablas de contingencia. (Peña, 2001, p.90)

**Tabla 4:** Ejemplo de tabla de contingencia

X/Y	$y_1$	$y_2$	...	$y_j$	...	$y_k$	Total
$x_1$	$n_{11}$	$n_{12}$	...	$n_{1j}$	...	$n_{1k}$	$n_{1+}$
$x_2$	$n_{21}$	$n_{22}$	...	$n_{2j}$	...	$n_{2k}$	$n_{2+}$
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
$x_l$	$n_{l1}$	$n_{l2}$	...	$n_{lj}$	...	$n_{lk}$	$n_{l+}$
Total	$n_{+1}$	$n_{+2}$	...	$n_{+j}$	...	$n_{+k}$	n

La representación gráfica más utilizada para la representación de estos datos es la nube de puntos o diagrama de dispersión.

**Gráfico 4:** Ejemplo de diagrama de dispersión.



### 2.6.2. Distribuciones marginales y condicionales

Dada una variable estadística bidimensional, las distribuciones marginales permiten estudiar de un modo aislado cada una de las componentes.

#### Distribuciones marginales de X e Y.

X	$n_{+}$	Y	$n_{+j}$
$x_1$	$n_{1+}$	$y_1$	$n_{+1}$
$x_2$	$n_{2+}$	$y_2$	$n_{+2}$
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
$x_l$	$n_{l+}$	$y_k$	$n_{+k}$

Las distribuciones condicionadas permiten estudiar el comportamiento de una de las variables cuando la otra permanece constante. Vienen dadas por:

#### Distribuciones condicionadas de X e Y.

X/Y = $y_j$	$n_{i/j}$	Y/X = $x_i$	$n_{j/i}$
$x_1$	$n_{1j}$	$y_1$	$n_{11}$
$x_2$	$n_{2j}$	$y_2$	$n_{22}$
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
$x_l$	$n_{lj}$	$y_k$	$n_{lk}$

### 2.6.3. Análisis descriptivo de dos variables.

Se trabaja de la misma forma que para una variable solo que para datos agrupados ya que para encontrarlo, se trabaja con las distribuciones marginales o condicionales dependiendo de los que se desea conocer.

Medida	Formula
Media	$\mu = \bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n}$
Mediana	$Mediana = L_i + \left( \frac{\frac{n}{2} - F_{i-1}}{F_i} \right) C$
Moda	$Moda = L_i + \left( \frac{F_i - F_{i-1}}{(F_i - F_{i+1}) + (F_i - F_{i-1})} \right) C$
Varianza Poblacional	$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^K (x_i - \mu)^2 F_i$
Muestral	$S^2 = \frac{1}{n-1} \sum_{i=1}^K (x_i - \bar{x})^2 F_i$
Desviación típica Poblacional	$S = \sqrt{\text{varianza}} = \sqrt{S^2}$
Muestral	$\sigma = \sqrt{\text{varianza}} = \sqrt{\sigma^2}$

**Ejemplo 43:** Se realiza una investigación con 30 bacterias que producen los alimentos de distinto tipo, para conocer a que tiempo (X) y a que temperatura (Y) se reproducen.

Tiempo para reproducirse	Temperatura para reproducirse			total
	20°C	30°C	40°C	
1 hora	1	5	2	8
2 horas	3	3	6	12
3 horas	3	2	5	10
Total	7	10	13	30

Se desea conocer:

la media, varianza y desviación típica marginales.

Hallar el número medio de tiempo en que se reproducen las bacterias condicionado a una temperatura 20°C .

Solución:

Para conocer la media, varianza y desviación típica primero se conocen las distribuciones marginales.

Distribución marginal con respecto a la tiempo del ejemplo anterior	
Tiempo para reproducirse	total
1 hora	8
2 horas	12
3 horas	10

La media de X es:

$$\bar{x} = \frac{(1*8)+(2*12)+(3*10)}{30} = \frac{62}{30} = 2.067$$

Interpretación: el promedio de tiempo que se tarda una bacteria en reproducirse es de 2.067 horas.

**La varianza**

$$S^2 = \frac{1}{n - 1} \sum_{i=1}^K (x_i - \bar{x})^2 F_i$$

$$S^2 = \frac{1}{30 - 1} * ((1 - 2.067)^2 * 8 + (2 - 2.067)^2 * 12 + (3 - 2.067)^2 * 10$$

$$S^2 = \frac{17.86667}{30 - 1} = 0.61609$$

Interpretación: la media es representativa ya que el valor de la varianza es muy pequeño.

**Desviación típica**

$$S = \sqrt{\text{varianza}}$$

$$S = \sqrt{0.61609} = 0.78491$$

Interpretación: en promedio, la cantidad de horas para que se reproduzca una bacteria se separa de la media, 0.78491 (horas).

para este caso encontraremos la distribución conjunta.

Tiempo para reproducirse	Temperatura para reproducirse
	20°C
1 hora	1
2 horas	3
3 horas	3
Total	7

**La media**

Donde

Y= es la temperatura de 20°C.



$$\bar{x} / Y = \frac{(1*1)+(2*3)+(3*3)}{7} = \frac{16}{7} = 2.2857$$

Interpretación: 2.287 horas es el tiempo promedio de la reproducción una bacteria a 20°C.

### **Utilizando el programa SPSS**

Para realizar el análisis de dos variables o como se conoce tablas de doble entrada, se puede hacer uso del programa informático SPSS. Ver (Anexo 1.4. Tablas de contingencia)

## CAPÍTULO 3: NOCIONES DE PROBABILIDAD

La teoría de probabilidad actualmente se ha convertido en una herramienta indispensable para aquellos estudios en los que se presenta la incertidumbre, la creación de esta teoría fue principalmente por los juegos al azar, ya que era una de las maneras para conocer si el jugador podía ganar o perder.

### 3.1. Técnicas de conteo

Las técnicas de conteo son una herramienta de las matemáticas, que estudia los distintos arreglos que se pueden hacer con los elementos de un conjunto de datos.

#### 3.1.1. Permutaciones

Las permutaciones o también llamadas ordenaciones son aquellas formas de agrupar los elementos de un conjunto teniendo en cuenta, que el orden en que se colocan influye.

#### Permutaciones sin repetición

Cuando todos los elementos que disponemos son: (Walpole, Myers & Myers, 2012, p.48)

$${}^n P_r = \frac{n!}{(n-r)!}$$

**Ejemplo 43:** Sea un conjunto formado por las bacterias a, b, c. ¿de cuantas formas diferentes se pueden observar?

Solución:

$${}^n P_r = \frac{n!}{(n-r)!}$$

$${}^3 P_3 = \frac{3!}{(3-3)!} = 6$$

$\{a, b, c\} \rightarrow (a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b), (c, b, a) \rightarrow 6$

Interpretación: se pueden observar las 3 bacterias de 6 formas diferentes.

#### Permutaciones con repetición

Cuando se dispone de elementos repetidos (ese es el n° de veces que se repite el elemento en cuestión). (Walpole, Myers & Myers, 2012, p.49)

$${}^n P_{n_1, n_2, \dots, n_i} = \frac{n!}{n_1! * \dots * n_i!}$$

**Ejemplo 44:** Sea un conjunto formado por las bacterias a, a, b, c, c, c. ¿de cuantas formas diferentes se pueden observar?

Solución:

$${}^n P_{n_1, n_2, \dots, n_i} = \frac{n!}{n_1! * \dots * n_i!}$$

$${}^6 P_{2, 1, 3} = \frac{6!}{2! * 1! * 3!} = 60$$

Interpretación: se pueden observar las bacterias de 60 formas diferentes.

### 3.1.2. Combinaciones

Se llama combinación de “n” elementos, a los diferentes grupos que se puedan formar en donde el orden no importa y no nos interesa el lugar que ocupan los mismos dentro del arreglo.

#### Combinaciones sin repetición

Se definen como las distintas agrupaciones formadas con “m” elementos distintos, eligiéndolos entre los “n” elementos disponibles. (Walpole, Myers & Myers, 2012, p.50)

$${}^n C_m = \binom{n}{m} = \frac{n!}{m! (n - m)!}$$

**Ejemplo 45:** se tienen 12 plantas por ser sembradas ¿de cuantas formas diferentes se pueden seleccionar 3 plantas sin repetir?

Solución:

Se utilizará la fórmula sin repetición.

$${}^n C_m = \binom{n}{m} = \frac{n!}{m! (n - m)!}$$

$${}^{12} C_3 = \binom{12}{3} = \frac{12!}{3! (12 - 3)!} = \frac{12 * 11 * 10 * 9!}{3 * 2 * 1 * 9!} = \frac{12 * 11 * 10}{3 * 2 * 1} = 220$$

Interpretación: se pueden hacer 220 de 3 plantas

### Combinaciones con repetición

Se definen como las distintas agrupaciones formadas con “m” elementos que pueden repetirse, eligiéndolos de entre los “n” elementos disponibles. (Wilhelmi, 2004, P.54)

$${}_n C_m = \binom{n+m-1}{m} = \frac{(n+m-1)!}{m!(n-1)!}$$

**Ejemplo 46:** se tienen 12 plantas por ser sembradas ¿de cuantas formas diferentes se pueden seleccionar 3 con repetir?

Solución:

Se utilizará la formula con repetición.

$$\begin{aligned} {}_n C_m &= \binom{n+m-1}{m} = \frac{(n+m-1)!}{m!(n-1)!} \\ {}_{12} C_3 &= \binom{12+3-1}{3} = \frac{(12+3-1)!}{3!(12-1)!} = \frac{14 * 13 * 12 * 11!}{3 * 2 * 1 * 11!} = \frac{14 * 13 * 12}{3 * 2 * 1} \\ &= 364 \end{aligned}$$

Interpretación: se pueden hacer 364 de 3 plantas

## 3.2. Espacio muestral

Dado un experimento aleatorio, llamaremos espacio muestral  $\Omega$  al conjunto de todos los resultados posibles distintos de dicho experimento aleatorio. Los elementos de  $\Omega$  se denominan sucesos elementales. (Wilhelmi, 2004, p.122p)

### 3.2.1. Tipo de espacios muestrales

espacio muestral finito: se da cuando  $\Omega$  está formado por un número finito de elementos.

Espacio muestral infinito: se da cuando el experimento es repetido infinidad de veces hasta obtener el objetivo.

**Ejemplo 47:** una pareja de esposos está esperando que nazcan sus gemelos, pero no conocen su sexo, ni el orden en que nacerán. El espacio muestral asociado sería:  $\Omega = \{(niña, niña), (niña, niño), (niño, niña), (niño, niño)\}$

### 3.3. Propiedades de la probabilidad

Para todo suceso A, la probabilidad de A estará entre cero y uno.

$$0 \leq P(A) \leq 1$$

La probabilidad de un suceso seguro "E", que ocurra siempre es uno.

$$P(E) = 1$$

La probabilidad de un suceso imposible o del conjunto vacío es cero.

$$P(\emptyset) = 0$$

Si  $A^c$  es el suceso complementario de A, entonces:

$$P(A^c) = 1 - P(A)$$

Ley de la suma de probabilidad para dos sucesos cualquiera A y B.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

La suma de las probabilidades de los sucesos elementales es 1.

$$P(A_1) + P(A_2) + \dots + P(A_n) = 1$$

### 3.4. Independencia y probabilidad condicional

#### Independencia:

Sean A y B dos sucesos del espacio muestral. El suceso A se dice independiente del suceso B si el conocimiento de la ocurrencia de B no modifica la probabilidad de aparición de A es decir, si: (Johnson & Kubly, 2008, p.243)

$$P(A/B) = P(A) \text{ o } P(A) = P(A/B).$$

Propiedad: Si dos sucesos A, B son independientes, entonces se verifica:

$$A \text{ es independiente de } B \leftrightarrow P(A \cap B) = P(A) \cdot P(B)$$

De la definición, se dice que "n" sucesos son independientes, si se cumple que:

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2) \dots P(A_n)$$

**Ejemplo 48:** Se consideran dos sucesos, A y B, asociados a un experimento aleatorio con  $P(A) = 0.7$   $P(B) = 0.6$   $P(A^c \cap B^c) = 0.58$ . ¿Son independientes A y B?

Solución:

Para ver si son independientes, se comprueba que:  $P(A \cap B) = P(A) \cdot P(B)$

$$P(A^c \cap B^c) = P[(A \cap B)^c] = 1 - P(A \cap B)$$

Por lo tanto,  $P(A \cap B) = 1 - P(A^c \cup B^c) = 1 - 0.58 = 0.42$

Por otro lado,  $P(A) \cdot P(B) = 0.7 \cdot 0.6 = 0.42$

Luego A y B son independientes, pues  $P(A \cap B) = P(A) \cdot P(B) = 0.42$

### **Probabilidad condicional:**

La probabilidad condicional de un evento A, dada la probabilidad de que ya ocurrió un evento B, es igual a (Johnson & Kuby, 2008, p.223)

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \text{ si } P(B) \neq 0$$

“se lee la probabilidad de un evento A dada que ya ocurrió el evento B”

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \text{ si } P(A) \neq 0$$

“se lee la probabilidad de un evento B dada que ya ocurrió el evento A”

**Ejemplo 49:** Una mujer es portadora de la enfermedad de Dúchenne. Si esta mujer tiene un bebe.

¿si nace él bebe y es varón, que probabilidad hay que tenga la enfermedad?

Solución:

Según las leyes de Mendel, todos los posibles genotipos de un hijo de una madre portadora ( $xX$ ) y un padre normal ( $XY$ ) son  $xX, xY, XX, XY$  y tienen la misma probabilidad.

El espacio muestral es:  $\Omega = \{xX, xY, XX, XY\}$

El suceso A= {hijo enfermo} corresponde al genotipo  $xY$ , por tanto, según la definición clásica de probabilidad

$$P(A) = 1/4 = 0,25$$

El suceso B = {ser varón}

El espacio muestral es:  $\Omega = \{xY, XY\}$

$$P(B) = \frac{1}{2} = 0,5$$

El suceso es decir, que sea varón y este enfermo  $A \cap B = \{xY\}$ ;

$$P(A \cap B) = 0,25;$$

Se quiere conocer la probabilidad de que él bebe posea la enfermedad partiendo de que es varón.

$$P(A|B) = 0,25/0,5 = 0,5$$

Interpretación: La probabilidad de que esta mujer tenga un bebe varón y que contraiga la enfermedad de ella es de 0.5

### **3.5. Tipo de variables aleatorias**

Una variable aleatoria de un espacio muestral  $\Omega$ . Es una regla que asigna un valor numérico a cada resultado de  $\Omega$ , en otras palabras, una función de  $\Omega$  en el conjunto  $R$  de números reales.

#### **3.5.1. Variables aleatorias discretas**

Es cuando toma un numero de valores finitos, o infinitos numerables. Esta variable corresponde a experimentos en los que se cuenta el número de veces que ha ocurrido un suceso(Johnson & Kuby(2008); Estadística Elemental; 10° edición; México, 271p)..

Se llamara función de probabilidad,  $f(x)$ , a la función que indique la probabilidad de cada posible valor.

$$f(x_i) = P(x = x_i)$$

Una forma equivalente de caracterizar la distribución de una variable es mediante la función de distribución,  $F(x)$  definida en cada punto " $x_0$ " como la probabilidad de que la variable aleatoria  $x$  tome un valor menor o igual que " $x_0$ " escribimos:

$$F(x_0) = P(x \leq x_0)$$

## *Principales distribuciones discretas de probabilidad*

### **Distribución de Bernoulli**

Se denomina experimento de Bernoulli a todo experimento aleatorio en el que solo son posibles dos resultados. (Peña, 2001, p.166)

$X = \{0 \text{ probabilidad de éxito, } 1 \text{ probabilidad de fracaso}\}$

La función de probabilidad se escribe:

$$f(x) = P(X = x) = p^x q^{1-x}$$

Su media será:  $\mu = E(x) = p$

La varianza:  $Var(x) = pq$

**Ejemplo 50:** En las costas de Perú se ha realizado un estudio sobre las algas para conocer si son venenosas o no.

¿Si se selecciona un alga, cuál es la probabilidad que el alga seleccionada sea venenosa?

Solución:

$x = \text{Que el alga sea venenosa}$

Espacio muestral = {venenosa, no venenosa}

$$p = \frac{1}{2}$$

$$q = (1 - p) = (1 - 0.5) = 0.5$$

$x = 1$  (Éxito)

$x = 0$  (Fracaso)

Se utiliza la formula

$$P(X = x) = p^x q^{1-x}$$

$$P(x = 1) = (0.5)^1 (0.5)^{1-1} = 0.5$$

Interpretación: la probabilidad de que el alga encontrada sea venenosa es de 0.5



### Distribución Geométrica

Consideremos el mismo mecanismo de generación de sucesos que en el modelo binomial, pero en lugar de contar el número de defectos en una muestra de “n”, consideramos. (Peña, 2001, p.168)

X= al número de elementos hasta el primer defectuoso

La función de probabilidad se escribe:

$$f(x = n) = P(X = x) = p(1 - p)^{n-1}; n = 1,2,$$

$$\text{Su media será: } \mu = E(x) = \frac{1}{p}$$

$$\text{La varianza: } Var(x) = q/p^2$$

**Ejemplo 51:** un granjero tiene una vaca la cual ha tenido 7 partos de una cría y está cargada nuevamente, si los estudios dicen que la probabilidad de que tenga dos crías es de 2% y de que tenga solo una cría es de 98%. Que probabilidades tiene el granjero que el octavo parto de su baca sea de dos crías.

Solución:

$$X = 8 \text{ partos}$$

$$p = 0.02$$

$$q = (1 - p) = 0.98$$

Sustituyendo en la fórmula:

$$P(x = n) = P(1 - p)^{n-1}$$

$$P(x = 8) = 0.02(1 - 0.02)^{8-1} = 0.02*(0.98)^7 = 0.01736$$

Interpretación: la probabilidad de que la vaca del granjero tenga dos crías en su octavo parto es de 0.01736

## Distribución Hipergeométrica

Una variable tiene distribución hipergeométrica si procede de un experimento que cumpla las condiciones siguientes:

1) una muestra de tamaño “ $n$ ”, sin remplazamiento, de un conjunto finito de “ $N$ ” objetos.

2)  $K$  de los  $N$  objetos se pueden clasificar como éxitos y  $N - K$  como fracasos.

$X$ = número de éxitos del que se quiere conocer la probabilidad. (Anderson, sweeney & Williams, 2008, p.214)

La función de probabilidad se escribe:

$$h(x, N, n, k) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}} = \frac{{}_K C_x * {}_{N-K} C_{n-x}}{{}_N C_n}$$

Su media será:  $\mu = E(x) = n * \frac{K}{N}$

La varianza:  $\text{Var}(x) = \frac{N-n}{N-1} \cdot n \cdot \frac{K}{N} \cdot \left(1 - \frac{K}{N}\right)$

**Ejemplo 52:** una empresa está exportando flores. De un lote de 10 flores, 4 se seleccionan al azar y son trasplantadas, sí el lote contiene 3 flores defectuosas las cuales si se trasplantan se marchitan, ¿Cuál es la probabilidad que las 4 flores se puedan trasplantar?

Solución:

$N = 10$  flores en total

$K = 7$  flores que se pueden trasplantar

$n = 4$  flores seleccionadas

$X = 4$  la probabilidad que se puedan trasplantar

$$h(x, N, n, k) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}$$

$$h(4,10,4,7) = \frac{\binom{7}{4} \binom{10-7}{4-4}}{\binom{10}{4}} = \frac{\binom{7}{4} \binom{3}{0}}{\binom{10}{4}} = \frac{35 * 1}{210} = 0.1667$$

Interpretación: la probabilidad de que las 4 flores seleccionadas al azar puedan ser trasplantadas es de 0.1667

### **Distribución Binomial**

Es una distribución de probabilidad discreta que cuenta el número de éxitos en una secuencia de "n" ensayos de Bernoulli independientes entre sí, con una probabilidad fija "p" de ocurrencia del éxito entre los ensayos. (Spiegel & Stephens, 2009, p.172)

x= números de elementos éxitos al observar "n"

La función de probabilidad es:

$$f(X = x) = P(X = x) = \binom{n}{x} p^x q^{n-x}$$

Su media será:  $\mu = E(x) = np$

La varianza es:  $Var(x) = npq$

**Ejemplo 53:** Un hombre y una mujer, cada uno con un gen recesivo (Azul) y uno dominante (Marrón) para el color de los ojos, serán padres de tres hijos. La probabilidad de que un niño nazca con ojos azules es de 0.25.

¿Cuál es la probabilidad de que los 3 hijos, tengan ojos azules?

Solución:

$$p = 1/4$$

$$q = (1 - p) = \left(1 - \frac{1}{4}\right) = 3/4$$

$$n = 3$$

x = {probabilidad que los 3 hijos tengan ojos azules}

Sustituyendo en:

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

$$P(X = 3) = \binom{3}{3} (1/4)^3 (3/4)^{3-3} = 0.0156$$

Interpretación: la probabilidad que los tres hijos tengan ojos de color es de 0.015

### **Distribución Binomial Negativa**

Es una ampliación de las distribuciones geométricas, utilizada en procesos en los cuales se ve necesaria la repetición de ensayos hasta conseguir un número de casos favorables (primer éxito). (Walpole, Myers & Myers, 2012, p.158)

La función de probabilidad es:

$$f(X = x) = P(X = x) = \binom{x-1}{r-1} p^r q^{x-r} \quad r = 0, 1, 2, \dots$$

$$\text{Su media será: } \mu = E(x) = r * \frac{q}{p}$$

$$\text{La varianza: } Var(x) = r * \frac{q}{p^2}$$

**Ejemplo 54:** Se sabe que la probabilidad de que un niño expuesto a una enfermedad contagiosa la contraiga es de 0,4.

¿Calcula la probabilidad de que el décimo niño estudiado sea el tercero en contraer la enfermedad?

Solución:

$$p = 0.4$$

$$q = (1 - p) = (1 - 0.4) = 0.6$$

$$r = 3 \text{ (Tercero en contraer la enfermedad)}$$

$$x = 10 \text{ (Niños del estudio)}$$

$$P(X = x) = \binom{x-1}{r-1} p^r q^{x-r}$$

$$P(X = 10) = \binom{10-1}{3-1} (0.4)^3 (0.6)^{10-3} = 0.0645$$

Interpretación: La probabilidad de que el décimo niño estudiado sea el tercero en contraer la enfermedad es de 0.0645

### **Distribución Uniforme Discreta.**

Tenemos esta distribución cuando el resultado de una experiencia aleatoria puede ser un conjunto finito de “n” posibles resultados, todos ellos igualmente probables. (Walpole, 1999, p.114)

Función de probabilidad es:

$$f(X = n) = P(X = n) = 1/n$$

$$\text{Media: } \mu = E(x) = \frac{n+1}{2}$$

$$\text{La varianza: } Var(x) = \frac{n^2-1}{12}$$

**Ejemplo 55:** un señor compro 3 tipos de semillas, semilla de girasol, tomate y chile. Por error se le mezclaron las 3 semillas y el no conoce de jardinería.

¿Cuál es la función de probabilidad?

¿Cuál es la probabilidad de que la semilla que siembre sea de tomate?

¿Cuál es la probabilidad que la semilla sea de cebolla?

Solución:

a)

$$n = 3(\text{Total de semillas})$$

$$P(X = x) = 1/n = 1/3$$

Interpretación: La probabilidad de tener semilla de girasol, tomate o chile es de 1/3.

Otros tipos de semilla tienen cero probabilidades.

b)

$$P(X = \text{tomate}) = 1/3$$

Interpretación: la probabilidad de que la semilla sembrada sea de tomate es de 1/3

c)

$$P(X = \text{cebolla}) = 0$$

Interpretación: la probabilidad de que la semilla sembrada sea de cebolla es de cero ya que el señor no compro de ese tipo de semilla.

### **Distribución Poisson.**

La distribución de Poisson es una distribución de probabilidad discreta que se aplica a sucesos que pueden ocurrir en un intervalo de tiempo o espacio. En este tipo de experimentos los éxitos buscados son expresados por unidad de área, tiempo, piezas, etc. (Spiegel & Stephens, 2009, p.175)

La función de probabilidades es:

$$f(x, \lambda) = P(x, \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, 2, \dots$$

Su media será:  $\mu = E(x) = \lambda$

La varianza:  $\text{Var}(x) = \lambda$

**Ejemplo 56:** En un depósito de agua se ha detectado una concentración media de 0.25 bacterias nocivas por cada  $\text{cm}^3$  de agua. Supuesto que las bacterias están aleatoriamente distribuidas, determinar:

¿Cuál es la probabilidad de que en una extracción al azar de un volumen de  $10 \text{ cm}^3$  se encuentren al menos 3 bacterias nocivas?

Solución:

$$\lambda = 0.25 * 10 = 2.5$$

$x =$  al menos 3

$$P(x, \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$$

$$P(0, 2.5) = \frac{e^{-2.5} 2.5^0}{0!} = 0.082085$$

$$P(1, 2.5) = \frac{e^{-2.5} 2.5^1}{1!} = 0.205212$$

$$P(2, 2.5) = \frac{e^{-2.5} 2.5^2}{2!} = 0.256516$$

$$P(\text{al menos } 3) = 1 - [F(x = 2) + F(x = 1) + F(x = 0)] \\ = 1 - 0.543813 = 0.456187$$

Interpretación: la probabilidad de encontrar al menos 3 bacterias en  $10 \text{ cm}^3$  es de 0.4562

### 3.5.2. Esperanza matemática, varianza y covarianza

#### Esperanza matemática

La esperanza matemática de una variable aleatoria “X” es el momento respecto al origen de orden 1 y se conoce como media o valor esperado de “X”. (Walpole, Myers & Myers, 2012, p.111)

La fórmula es:

$$E(x) = \sum_{i=1}^n x_i * f(x_i)$$

#### Propiedades de la esperanza matemática.

La esperanza de una constante “k” es la propia constante:  $E[k] = k$

Si X está acotada  $a \leq X \leq b$  la esperanza se encuentra entre:

$$a \leq E[X] \leq b.$$

La esperanza  $E[a * X + b] = a * E[X] + b$  y en general, la esperanza de una combinación lineal es:

$$E[a * g(X) + b * h(X)] = a * E[g(X)] + b * E[h(X)]$$

Si  $g(x_i) \leq h(x_i)$  para toda i es

$$E[g(X)] \leq E[h(X)]$$

El valor absoluto de la esperanza de una función  $g(X)$ , es menor o igual que la esperanza del valor absoluto de la dicha función:

$$|E[g(X)]| \leq E[|g(X)|]$$

Si  $f(x) = f(-x)$  o sea si la función de densidad es simétrica, entonces:

$$E[f(x)] = 0.$$

**Ejemplo 57:** Cierta día un vendedor de una empresa de aparatos médicos tiene dos citas. Considera que en la primera cita tiene 70 por ciento de probabilidades de cerrar una venta, por la cual podría obtener una comisión de \$1000. Por otro lado, cree que en la segunda cita sólo tiene 40 por ciento de probabilidades de cerrar el trato, del cual obtendría \$1500 de comisión. ¿Cuál es su comisión esperada con base en dichas probabilidades? Suponga que los resultados de las citas son independientes.

**Solución:**

En primer lugar sabemos que el vendedor, en las dos citas, puede obtener 4 comisiones totales: \$0, \$1000, \$1500 y \$2500. Necesitamos calcular sus probabilidades asociadas. Mediante la independencia obtenemos

$$f(\$0) = (1 - 0.7)(1 - 0.4) = 0.18, f(\$2500) = (0.7)(0.4) = 0.28,$$

$$f(\$1000) = (0.7)(1 - 0.4) = 0.42, \text{ y } f(\$1500) = (1 - 0.7)(0.4) = 0.12.$$

Por lo tanto, la comisión esperada para el vendedor es

$$E(X) = (\$0)(0.18) + (\$1000)(0.42) + (\$1500)(0.12) + (\$2500)(0.28) = \$1300.$$

### **Varianza**

La varianza de una variable aleatoria es una característica numérica que proporciona una idea de la dispersión de la variable aleatoria respecto de su esperanza. Se dice que es un parámetro de dispersión. (Walpole, Myers & Myers, 2012, p.119)

$$Var(x) = E[(x - \mu)^2] = \sum_{i=1}^n (x_i - \mu)^2 f(x_i)$$

### **Propiedades de la varianza**

Fórmula alternativa muy útil

$$Var(X) = E[X^2] - (E[X])^2$$

La varianza de una constante “k” es cero.

$$Var(k) = 0$$

La varianza de una constante por X, la constante sale al cuadrado.

$$Var(a * X) = a^2 * Var(X)$$

$$Var(aX + b) = a^2 * Var(X)$$

**Ejemplo 58:** Suponga que la variable aleatoria X representa el número de automóviles que se utilizan con propósitos de negocios oficiales en un día de trabajo dado. La distribución de probabilidad para la empresa A es



x	1	2	3
$f(x)$	0.3	0.4	0.3

y para la empresa B es

x	0	1	2	3	4
$f(x)$	0.2	0.1	0.3	0.3	0.1

Demuestre que la varianza de la distribución de probabilidad para la empresa B es mayor que la de la empresa A.

Solución: Para la empresa A encontramos que

$$\mu = E(X) = (1)(0.3) + (2)(0.4) + (3)(0.3) = 2.0,$$

y entonces

$$\sigma_A^2 = \sum_{x=1}^3 (x - 2)^2 = (1 - 2)^2(0.3) + (2 - 2)^2(0.4) + (3 - 2)^2(0.3) = 0.6.$$

Para la empresa B tenemos

$$\mu = E(X) = (0)(0.2) + (1)(0.1) + (2)(0.3) + (3)(0.3) + (4)(0.1) = 2.0$$

y entonces

$$\begin{aligned} \sigma_B^2 &= \sum_{x=0}^4 (x - 2)^2 f(x) \\ &= (0 - 2)^2(0.2) + (1 - 2)^2(0.1) + (2 - 2)^2(0.3) + (3 - 2)^2(0.3) \\ &\quad + (4 - 2)^2(0.1) = 1.6. \end{aligned}$$

Interpretación: Es evidente que la varianza del número de automóviles que se utilizan con propósitos de negocios oficiales es mayor para la empresa B que para la empresa A. Una fórmula alternativa que se prefiere para calcular  $\sigma^2$ , que a menudo simplifica los cálculos, se establece en el siguiente teorema.

### Covarianza

Sean X e Y dos variables aleatorias, se define la covarianza entre estas dos variables como Ronald walpole(2012); probabilidad y estadística para ingeniería y ciencia; 9ª edición; Pearson; México; 119p):

$$Cov[X, Y] = E[(X - \mu_x)(Y - \mu_y)]$$

### **Propiedades de la covarianza**

Expresión alternativa:

$$Cov(X, Y) = E[X \cdot Y] - E[X] \cdot E[Y]$$

Si X e Y son variables aleatorias independientes, entonces:

$$Cov(X, Y) = 0$$

Si X e Y son variables aleatorias con esperanzas  $E[X]$  y  $E[Y]$  y se definen las funciones:  $U = a_1 * X + b_1$  y  $V = a_2 * Y + b_2$ , entonces:

$$Cov(U, V) = Cov(a_1 * X + b_1, a_2 * Y + b_2) = a_1 * a_2 Cov(X, Y)$$

### **3.5.3. Variables aleatorias continuas**

Una variable es continua cuando puede tomar cualquier valor en un intervalo. Por ejemplo el peso de una persona, el tiempo de duración de un suceso, etc. No es posible conocer el valor exacto ya que está en un intervalo. (Johnson & Kubby(2008, p.271p)

#### ***Distribuciones de probabilidad continua***

##### **Distribución Uniforme Continua.**

La distribución o modelo uniforme puede considerarse como proveniente de un proceso de extracción aleatoria. Dada una variable aleatoria continua, x, definida en el intervalo [a,b] de la recta real, diremos que x tiene una Distribución Uniforme en el intervalo [a,b] cuando su función de densidad para  $x \rightarrow U[a, b]$  sea.( Walpole, Mylers & Mylers, 2012, p.171)

Su función se define como:

$$f(x) = P(X = x) = \frac{1}{b-a} \text{ Para } x \in [a, b].$$

$$\text{La media: } \mu = E(x) = \frac{a+b}{2}$$

$$\text{Varianza: } Var(x) = \frac{(b-a)^2}{12}$$

**Ejemplo 59:** la temperatura aproximada en la que una bacteria de alimentos se reproduzca se encuentra, entre 5 y 37 grados Centígrados.

¿Cuál es la función de distribución y la temperatura media esperada?

¿Cuál es la probabilidad que de 20 a 30 grados centígrados se reproduzca una bacteria de alimento?

Solución:

$$a) P(x) = \frac{1}{b-a} = \frac{1}{37-5} = 0.03125$$

Es decir, si la temperatura esta entre 5 y 6 grados la probabilidad que una bacteria de alimentos se reproduzca es de 3.1%

$$E(x) = \frac{a+b}{2} = \frac{37+5}{2} = 21$$

Interpretación: La temperatura media estimada para la reproducción de bacterias de alimento es de 21 grados centígrados.

$$b) P(a \leq x \leq b) = f(x)(b-a) = \left(\frac{1}{32}\right)(30-20) = 0.3125$$

Interpretación: La probabilidad que una bacteria de alimento se reproduzca entre 20 y 30 grados centígrados e de 0.3125

### **Distribución Gaussiana o Normal**

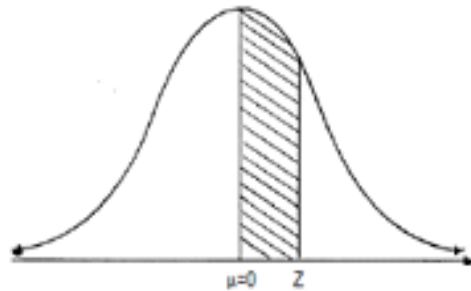
El modelo de distribución de probabilidad para variables continua más importante es la distribución normal, esta función depende de dos paramentos: la media  $\mu$  y la desviación típica  $\sigma$ . Esta distribución se utiliza para conocer: las medidas físicas del cuerpo humano en una población, errores de las observaciones astronómicas, etc. (Spiegel & Stephens, 2009, p.173)

La función es la siguiente:

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Para calcular probabilidades en el caso general, se transforma la variable aleatoria normal x en variable normal estándar z, mediante:

$$z = \frac{x - \mu}{\sigma}$$



Para conocer el área bajo la curva del valor z se utiliza la tabla de distribución normal tipificada que se encuentra en los anexos (anexo 2).

**Ejemplo 60:** La estatura de una persona depende de la genética de los padres en la formación del ADN, supóngase que en un colegio se les toma la estatura a los jóvenes de 3er año de bachillerato; y resultado que la estatura media es de 167cm, con una desviación de 4cm. El director quiere conocer el porcentaje de jóvenes que están entre 165cm y 169cm de estatura.

Solución:

Media: 167

Desviación: 4

$$x_1 = 165$$

$$x_2 = 169$$

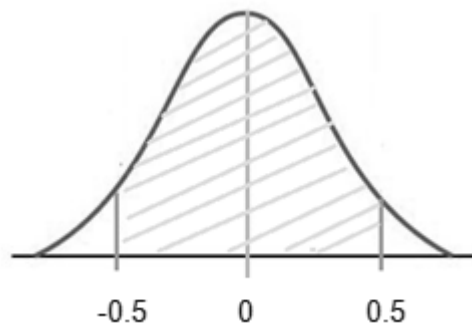
$$P(x_1 \leq X \leq x_2) = P(z_1 \leq z \leq z_2)$$

$$Z_1 = \frac{x_1 - \mu}{\sigma} = \frac{165 - 167}{4} = -0.5$$

$$Z_2 = \frac{x_2 - \mu}{\sigma} = \frac{169 - 167}{4} = 0.5$$

$$P(165 \leq X \leq 169) = P(-0.5 \leq z \leq 0.5)$$

Se realiza el grafico para conocer el área a encontrar en la tabla que se encuentra en anexos:



$$P(-0.5 \leq z \leq 0.5) = P(z \leq 0.5) + P(z \geq -0.5)$$

Los valores de  $z$  se buscan en la tabla de distribución normal tipificada.

$$P(-0.5 \leq z \leq 0.5) = 0.1915 + 0.1915 = 0.383$$

Interpretación: el 38.3% de los jóvenes estudiantes de 3er año de bachillerato tiene estaturas entre 165 y 169.

## **CAPÍTULO 4: INFERENCIA ESTADÍSTICA**

### **4.1. Estadística Inferencial**

De acuerdo con el diccionario de la Real Academia Española, inferir significa “Sacar una consecuencia o deducir algo de otra cosa”. Existen algunos eventos, que, para ser analizados, necesitan elementos estadísticos que van más allá de la Estadística Descriptiva, es decir, que no basta con recolectar, describir y resumir los datos, para presentarlos en gráficas, tablas o valores numéricos. Es por eso que surge la Estadística Inferencial, la cual se encarga de hacer deducciones de una población por medio de una muestra tomada a partir de ésta.

#### **4.1.1. Definición de Estadística Inferencial**

La Estadística Inferencial nos permite comprender las técnicas, con las que con base únicamente en una muestra sometida a observación, se toman decisiones sobre una población o proceso estadístico.

Ésta se dedica a la generación de los modelos, inferencias y estimaciones asociados a los fenómenos en cuestión, teniendo en cuenta lo aleatorio e incertidumbre en las observaciones. Esta Estadística trabaja con muestra, que no son más que subconjuntos formados de algunos individuos de la población; y a partir del estudio de la muestra se pretende inferir aspectos relevantes de toda la población. (Johnson & Kuby, 2008, p.398)

#### **4.1.2. Distribuciones muestrales**

Quizás a veces nos preguntamos: ¿Por qué se requiere muestrear? En muchos casos no es factible estudiar a la población entera, bien por el costo muy alto, o se necesita de mucho tiempo para contactar a toda la población, entre otras, son algunas de las razones por lo que es indispensable muestrear.

Para hacer dicho estudio, se necesitan tomar muestras, que no son más que una parte de la población. Por medio del muestreo, surge el uso de los métodos estadísticos inferenciales.

Una vez seleccionada la muestra, se pueden construir variables aleatorias alternativas, desprendiéndose de éstas, propiedades estadísticas de gran interés.

Las dos formas más comunes de estas variables corresponden a las distribuciones muestrales de la media y las distribuciones muestrales de la proporción. Cabe destacar, que cuando se usan valores muestrales o estadísticos para estimar parámetros o valores poblacionales, se puede correr el riesgo de que ocurran dos tipos de errores: el error muestral y el error no muestral (sesgo muestral).

Para (Webster, Allen, 1996, 296p), las distribuciones muestrales se refieren a la “lista de todos los valores posibles de un estadístico y la probabilidad asociada a cada valor.” La distribución muestral, se refiere a la distribución de los valores que tomará el estimador al escoger diferentes muestras de la población. Ésta distribución se basa fundamentalmente en dos medidas: la media, que indica el valor promedio del estimador; y la desviación típica o error típico de estimación, que se refiere a la desviación promedio que se puede esperar entre el estimador y el valor del parámetro.

Existen tres tipos de distribuciones muestrales, de las cuales se definirán brevemente: la distribución muestral de la media, la distribución muestral de la proporción y la distribución muestral de la varianza.

**Distribución Muestral de la Media:** puede definirse como aquella distribución que consta de todas las medias muestrales posibles de un tamaño de muestra dado. En donde cada muestra de tamaño “n”, extraída de una población proporciona una media, la cual es considerada como una variable aleatoria para estudiar su distribución. Si se denota a la media  $\mu_{\bar{x}}$  y la desviación estándar  $\sigma_{\bar{x}}$  de una distribución de medias y con media  $\mu$  y desviación estándar  $\sigma$  de una población, entonces. (Spiegel & Stephens, 2009, p.204)

$$\mu_{\bar{x}} = \mu \text{ y } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_P - N}{N_P - 1}}$$

Si la población es infinita,

$$\mu_{\bar{x}} = \mu \text{ y } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

Si el valor de  $N$  es grande ( $N \geq 30$ ), la distribución muestral de las medias es aproximada a la normal media  $\mu_{\bar{x}}$  y la desviación estándar  $\sigma_{\bar{x}}$ , independientemente de la población (siempre y cuando la media y la varianza poblacional sea finita) y si la población está distribuida normalmente, la distribución muestral de las medias también es normal aun cuando el valor de  $N$  sea pequeño (es decir,  $N < 30$ ).

**Distribución Muestral de la Proporción:** es una distribución que se da, cuando muchas veces se plantea estimar una proporción o porcentaje, en donde la variable aleatoria toma únicamente dos valores diferentes, que no son más, que el éxito o el fracaso, en otras palabras, sigue una distribución binomial. De esta manera se obtiene una distribución muestral de las proporciones cuya media  $\mu_p$  y cuya desviación estándar  $\sigma_p$  están dadas por. (Spiegel & Stephens, 2009, p.205)

$$\mu_p = p \text{ y } \sigma_p = \sqrt{\frac{pq}{N}} \sqrt{\frac{p(1-p)}{N}}$$

Si el valor de  $N$  es grande ( $N \geq 30$ ), esta distribución es aproximadamente normal.

**Distribución muestral de la varianza:** (Spiegel, 1999) Si se extrae una muestra aleatoria de tamaño  $n$  de una población normal con media  $\mu$  y varianza  $\sigma^2$ , y se calcula la varianza muestral, se obtiene un valor del estadístico  $S^2$ . Tenemos una variable aleatoria chi cuadrada con  $n$  grados de libertad dividida en dos componentes. Observe que en la distribución chi cuadrada es un caso especial. El segundo término del lado derecho es  $Z^2$ , que es una variable aleatoria chi cuadrada con 1 grado de libertad, y resulta que  $(n - 1)S^2/\sigma^2$  es una variable aleatoria chi cuadrada con  $n - 1$  grados de libertad. Formalizamos esto: si  $S^2$  es la varianza de una muestra aleatoria de tamaño  $n$  que se toma de una población normal que tiene la varianza  $\sigma^2$ , entonces el estadístico

$$\chi^2 = \frac{(n - 1)S^2}{\sigma^2} = \sum_{i=1}^n \frac{(X_i - \bar{x})^2}{\sigma^2}$$

Tiene una distribución chi cuadrado con  $n-1$  grados de libertad. (p.244)



#### **4.1.3. Procedimientos de Muestreo**

Antes de hablar en sí de los métodos de muestreo, es necesario mencionar los riesgos asociados al procedimiento de muestreo. Estos son: el error muestral y el sesgo muestral. (Spiegel, 1999, p.262).

##### ***Errores y Sesgos***

###### **a) Error Muestral**

(Chao, 1993, p.137). Opina que “es la diferencia entre el valor de una estadística obtenido mediante los datos muestrales y el valor correspondiente al parámetro de la población debido a variaciones fortuitas en la selección de las unidades”.

###### **b) Sesgo Muestral**

Para (Mason, Lind y Marchal, 2001, p263). “es cuando los resultados de la muestra no probabilística pueden no ser representativos de la población”. Cuando se usan valores muestrales o estadísticos para estimar parámetros o valores poblacionales, se puede correr el riesgo de que ocurran dos tipos de errores; en donde las muestras no son representativas a la población, debido a que los datos no son cien por ciento reales. Entendiéndose por muestras no representativas, aquellas que dan lugar a una estimación errónea del parámetro y a un error muestral.

Estos tipos de errores son: el error muestral, que se conoce como azar de la extracción, el cual se refiere a la variación natural que se da con las muestras tomadas de la misma población y por ende se encarga de determinar que esas muestras no contengan elementos que no sean característicos de la población; y el error no muestral, que es a lo que se refiere el sesgo muestral, aunque para algunos autores este es un tipo de error muestral; no es más que una tendencia inherente a un método de muestreo, que da estimaciones de un parámetro, las cuales pueden ser menores o mayores en promedio que el parámetro real (sesgo positivo y sesgo negativo), debido a factores que dependen de la recolección, análisis, interpretación y revisión de los datos; permitiendo así, llegar a conclusiones que son sistemáticamente diferentes de la verdad o incorrectas acerca de los objetivos de una

investigación.

#### **4.1.4. Intervalo de Confianza**

Con la normalización de estudios estadísticos por medio de las distribuciones muestrales, se pueden determinar parámetros de una población a través de sus valores estadísticos. Esto se conoce como estimación, la cual se clasifica en: estimación puntual y estimación por intervalo. La estimación puntual, es aquella donde se estima un parámetro poblacional por medio de un sólo estadístico o valor del estimador.

La estimación por intervalo o intervalo de confianza, es aquella que viene dada por un rango de valores dentro del cual se espera encontrar el valor del parámetro que se estime. En otras palabras, el intervalo de confianza, se refiere a todos los valores obtenidos a partir de los datos de una muestra, en el que hay una determinada probabilidad de que se encuentre el parámetro estudiado con cierta certeza.

Dicho intervalo viene dado por un Límite Inferior (Li) y un Límite Superior (Ls). Pero antes, de empezar a hablar, tanto del intervalo de confianza de la media como el de la proporción, es necesario definir que es un nivel de confianza, así como el coeficiente de confianza, y el teorema central del límite.

**Nivel de confianza:** es la máxima probabilidad de encontrar el valor del parámetro que se estima dentro del intervalo establecido, es decir, es el coeficiente de confianza expresado en porcentaje.

**Coefficiente de confianza:** es la probabilidad que existe, de que el intervalo de confianza contenga el parámetro poblacional.

**Teorema del límite central:** permite el uso de probabilidad normal para crear los intervalos de confianza de la media poblacional y realizar pruebas de hipótesis.

#### **4.1.5. Prueba de Hipótesis**

Una hipótesis es una afirmación o suposición respecto al valor de un parámetro poblacional.

Cuando las poblaciones de interés son tan grandes que no es factible estudiar todos sus elementos, una solución es tomar una muestra de la población de interés; y de esta manera se puede probar una aseveración para determinar si la evidencia soporta o no la afirmación.

Para (Mason, Lind & Marchal, 2001, p.353). Una prueba de hipótesis “Es un procedimiento estadístico que se basa en evidencias muestrales y se emplea para determinar si la declaración planteada acerca del parámetro poblacional es racional”.

La prueba de hipótesis, viene a ser un supuesto que se plantea el investigador antes de iniciar una determinada investigación. Supuesto que al final de la investigación puede ser cierto y aprobarse, así como también puede ser falso y rechazarse. La finalidad de la prueba de hipótesis, no es cuestionar el valor calculado del estadístico muestral, sino que se encarga de hacer juicio con respecto a la diferencia entre el estadístico muestral y un valor planteado del parámetro.

#### **4.1.6. Procedimiento para Probar una Hipótesis**

Para probar una hipótesis, se sigue un procedimiento sistemático que consta de una serie de pasos, los cuales le van a permitir al investigador tomar una decisión.

##### **Paso 1: Plantear la hipótesis nula ( $H_0$ ) y la hipótesis alternativa ( $H_a$ )**

Toda investigación estadística requiere de la existencia de hipótesis o afirmaciones acerca de la población que se estudia. El primer paso, es establecer la hipótesis a ser probada. Ésta es llamada hipótesis nula ( $H_0$ ), es una afirmación que no se rechaza, a no ser que los datos de la muestra proporcionen evidencia convincente de que es falsa; indica que no existe diferencia significativa entre los resultados obtenidos esperados en la investigación determinada. Si no se rechaza la hipótesis nula, se dice que la evidencia no es suficiente para rechazarla, pero tampoco se puede afirmar que es verdadera.

La hipótesis alternativa ( $H_a$ ), es una afirmación que se toma si se rechaza la hipótesis nula. Indica que existe diferencia significativa entre los resultados obtenidos y los resultados esperados en una determinada investigación. Se toma la hipótesis alternativa si la evidencia proporcionada por la muestra es suficiente para afirmar que la hipótesis nula es falsa. (Peña, 2001, p.380)

### **Paso 2: Seleccionar el nivel de significancia**

Se determina el criterio de contraste, especificando el nivel de significancia, el tipo de distribución y los valores críticos, como se muestra en la siguiente figura:



El nivel de significancia o de confianza, es la probabilidad de rechazar una hipótesis nula verdadera. Si la hipótesis planteada es verdadera, entonces, el nivel de confianza indicará la probabilidad de no aceptarla por estar en el área de aceptación. El error asociado  $(1-\alpha)$ , indica la probabilidad de aceptar la hipótesis planteada cuando es verdadera en la población.

El tipo de distribución va a depender de la naturaleza de la hipótesis y del tamaño de la muestra. Si la hipótesis es relativa a las medias poblacionales y las muestras son grandes ( $n > 30$ ) se usa la distribución normal. En caso contrario, de tratarse de una muestra pequeña ( $n \leq 30$ ) se utiliza la distribución "t" de Student.

Toda decisión tomada por medio de una prueba de hipótesis, puede conllevar a un error. Existen cuatro posibilidades cuando se va a tomar una decisión respecto a una hipótesis:

Decisión basada en la muestra	Verdad acerca de la población	
	<b><math>H_0</math> Es verdadera</b>	<b><math>H_0</math> Es falsa</b>
<b>No rechazar <math>H_0</math></b>	Decisión correcta(probabilidad = $1 - \alpha$ )	<b>Error tipo II</b> no rechazar $H_0$ cuando es falsa (probabilidad= $\beta$ )
<b>Rechazar <math>H_0</math></b>	<b>Error tipo I</b> rechazar $H_0$ cuando es verdadera (probabilidad= $\alpha$ )	Decisión correcta (probabilidad= $1 - \beta$ )

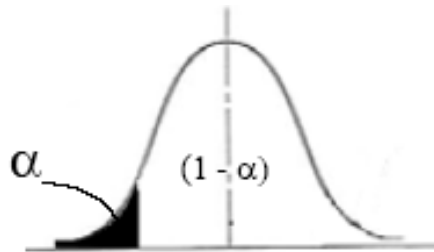
El error tipo I, que se denota con la letra griega  $\alpha$ , se da en el caso de que la hipótesis nula sea rechazada siendo esta verdadera. Mientras que el error tipo II, denotado con la letra griega  $\beta$ , existe cuando se no se rechaza la hipótesis nula siendo esta falsa. Los valores críticos, separan a la región de no rechazo de la de rechazo. Son aquellos valores de la variable de la distribución que limitan el área crítica, que no es más que la parte de la curva que corresponde al nivel de significancia.

La prueba de hipótesis está conformada por dos regiones. Una región de rechazo, conocida como región crítica; y una región de no rechazo, conocida como región de aceptación.

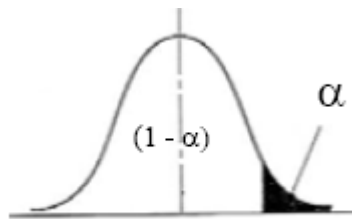
Si el estadístico de prueba cae dentro de la región de aceptación, no se puede rechazar la hipótesis nula. Sin embargo, la región de rechazo es considerada como el conjunto de valores de la prueba de hipótesis que no tienen posibilidad de presentarse si la hipótesis nula es verdadera. Por otra parte, es indispensable saber que existen dos tipos de prueba: la prueba de una cola y la prueba de dos colas.

La prueba unilateral, de una cola o de un extremo; es aquella en donde la hipótesis planteada indica una sola dirección, formulada con mayor o igual que ( $\geq$ ) o menor o igual que ( $\leq$ ).

Las siguientes figuras muestran, los diferentes tipos de prueba unilateral:



Prueba unilateral izquierda



Prueba unilateral derecha

La prueba bilateral, de dos colas o de dos extremos; es aquella en donde la hipótesis planteada no indica dirección.  $H_0$ , se formula con la igualdad (=) y,  $H_a$  con diferencia ( $\neq$ ).



### Paso 3. Calcular el valor estadístico de prueba

El estadístico de prueba, se refiere a un valor determinado por medio de la información de la muestra, el cual se debe comparar con el criterio de contraste, permitiendo así rechazar o aceptar la hipótesis. Este estadístico, va a variar con la cantidad de muestras que se tomen. Si las muestras a utilizar es mayores a treinta, se utiliza el estadístico “z”, de ser el caso contrario, es utilizado el estadístico “t”.

#### **Paso 4. Formular la regla de decisión**

Es cuando se establecen las condiciones en las que se rechaza o se acepta la hipótesis nula. En donde, la región de rechazo, define la ubicación de todos los valores que son tan grandes o demasiado pequeños, que la probabilidad que se presenten bajo la suposición de que una hipótesis nula es verdadera.

#### **Paso 5. Tomar decisión**

El último paso es tomar una decisión, para ver si se rechaza o se acepta la hipótesis nula. Si el estadístico de prueba queda dentro de la zona crítica, la hipótesis nula se tendrá que rechazar y se aceptará la hipótesis alternativa. Si dicho valor se encuentra fuera de la zona crítica, entonces, la hipótesis nula no deberá rechazarse.

### **4.2. Análisis de varianza y pruebas no paramétricas**

El análisis de varianza, es utilizado para comparar si los valores de un grupo de datos numéricos, son significativamente distintos a los valores de otro o más conjunto de datos. El procedimiento para comparar estos valores, está basado en la varianza global observada en los grupos de datos numéricos a comparar.

#### **4.2.1. Análisis de Varianza (Anova)**

Para (Levin & Rubin, 1996, p.631). El análisis de varianza, es una “técnica estadística utilizada para probar la equidad de tres o más medias de muestra y, de este modo, hacer inferencias sobre si las muestras provienen de poblaciones que tienen la misma media”.

El análisis de varianza, es una herramienta que se aplica para probar simultáneamente si las medias de varias poblaciones son iguales. Dicho análisis tiene una serie de pasos a seguir:

1. Se debe determinar una estimación de la varianza de la población, a partir de la varianza entre las medias muestrales.
2. Se determina una segunda estimación de la varianza de la población desde la varianza dentro de las muestras.

Se comparan las dos estimaciones, a través del cociente  $F$ ; si un valor es

aproximadamente igual, se acepta la hipótesis nula.

#### **4.2.2. Pruebas no Paramétricas**

En la práctica no con todas las poblaciones se puede asumir que sigan una distribución de probabilidad normal, se generan situaciones en donde no es posible formular una hipótesis segura sobre el valor de un parámetro o la forma de la distribución poblacional, para las que las pruebas paramétricas de “z” y “t” no son adecuadas; surgen las pruebas no paramétricas, las cuales no dependen de un solo tipo de distribución o de unos valores específicos de los parámetros.

Según (Webster, 1996, p.836), las pruebas no paramétricas, “son procedimientos estadísticos que se pueden utilizar para contrastar hipótesis cuando no es posible fijar ningún supuesto sobre parámetros o distribuciones poblacionales”.

Existen diferentes tipos de pruebas no paramétricas, que se pueden utilizar para una necesidad determinada, como lo son: la prueba de los signos, la prueba de rachas, la prueba de Mann-Whitney, la correlación de rangos de Spearman, la prueba de Kruskal y la prueba de ji- cuadrado, dividiéndose ésta en: prueba de bondad de ajuste y prueba de independencia o tablas de contingencias.

##### ***Prueba de los signos***

Es usada para hacer pruebas de hipótesis acerca de la mediana. La prueba estadística se basa en el estadístico de (Wilcoxon 1945), el cual se calcula de la siguiente manera: Se resta de cada dato el valor de la mediana que se considera en la hipótesis nula. Se calcula los rangos de las diferencias sin tomar en cuenta el signo de las mismas (o sea en valor absoluto). En el caso de haber empate se asigna un rango promedio a todas las diferencias empatadas, es decir; se les asigna el rango: (menor rango del grupo del empate + mayor rango del grupo del empate)/2.

El estadístico W de Wilcoxon será la suma de los rangos correspondientes a las diferencias positivas.

##### ***Prueba de rachas***

El contraste de rachas permite verificar la hipótesis nula de que la muestra es



aleatoria, es decir, si las sucesivas observaciones son independientes. Este contraste se basa en el número de rachas que presenta una muestra. Una racha se define como una secuencia de valores muestrales con una característica común precedida y seguida por valores que no presentan esa característica.

Así, se considera una racha la secuencia de " $k$ " valores consecutivos superiores o iguales a la media muestral (o a la mediana o a la moda, o a cualquier otro valor de corte) siempre que estén precedidos y seguidos por valores inferiores a la media muestral (o a la mediana o a la moda, o a cualquier otro valor de corte).

El número total de rachas en una muestra proporciona un indicio de si hay o no aleatoriedad en la muestra. Un número reducido de rachas (el caso extremo es 2) es indicio de que las observaciones no se han extraído de forma aleatoria, los elementos de la primera racha proceden de una población con una determinada característica (valores mayores o menores al punto de corte) mientras que los de la segunda proceden de otra población. De forma idéntica un número excesivo de rachas puede ser también indicio de no aleatoriedad de la muestra.

### ***Prueba de Mann-Whitney***

Se usa cuando se quiere comparar dos poblaciones usando muestras independientes, es decir; es una prueba alterna a la prueba de "t" para comparar dos medias usando muestras independientes.

La hipótesis nula es que las medianas de las dos poblaciones son iguales y la hipótesis alterna puede ser que la mediana de la población 1 sea mayor (menor o distinta) de la mediana de la población 2.

Cuando tanto " $n_1$ " como " $n_2$ " sean mayores que 10, se puede demostrar que si no hay empates, entonces  $W$  se distribuye aproximadamente como una normal con:

$$\text{Media} = n_1(n_1 + n_2 + 1)/2$$

$$\text{Varianza} = n_1n_2(n_1 + n_2 + 1)/12.$$

### ***Correlación de rangos de Spearman***

La correlación de Spearman mide el grado de asociación entre dos variables

cuantitativas que siguen una tendencia siempre creciente o siempre decreciente. Es más general que el Coeficiente de correlación de Pearson, la correlación de Spearman, en cambio se puede calcular para relaciones exponenciales o logarítmicas entre las variables.

### ***Prueba de Kruskal***

La prueba de Kruskal-Wallis, es una alternativa a la prueba “F” del análisis de varianza para diseños de clasificación simple. En este caso se comparan varios grupos, pero usando la mediana de cada uno de ellos, en lugar de las medias.

### ***Prueba de ji- cuadrado***

Esta prueba puede utilizarse incluso con datos medibles en una escala nominal. La hipótesis nula de la prueba Chi-cuadrado postula una distribución de probabilidad totalmente especificada como el modelo matemático de la población que ha generado la muestra.

Para realizar este contraste se disponen los datos en una tabla de frecuencias. Para cada valor o intervalo de valores se indica la frecuencia absoluta observada o empírica ( $O_i$ ). A continuación, y suponiendo que la hipótesis nula es cierta, se calculan para cada valor o intervalo de valores la frecuencia absoluta que cabría esperar o frecuencia esperada ( $E_i = n \cdot p_i$ , donde “n” es el tamaño de la muestra y  $p_i$  la probabilidad del i-ésimo valor o intervalo de valores según la hipótesis nula). El estadístico de prueba se basa en las diferencias entre la " $O_i$ " y " $E_i$ " y se define como:

$$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

Este estadístico tiene una distribución ji cuadrado con k-1 grados de libertad si “n” es suficientemente grande, es decir, si todas las frecuencias esperadas son mayores que 5. En la práctica se tolera un máximo del 20% de frecuencias inferiores a 5.

Si existe concordancia perfecta entre las frecuencias observadas y las esperadas el

estadístico tomará un valor igual a 0; por el contrario, si existe una gran discrepancia entre estas frecuencias el estadístico tomará un valor grande y, en consecuencia, se rechazará la hipótesis nula. Así pues, la región crítica estará situada en el extremo superior de la distribución Chi-cuadrado con  $k-1$  grados de libertad.

### **4.3. Análisis de Regresión y Correlación**

El análisis de Regresión y Correlación, son muy usados en la Investigación Científica, una herramienta muy útil cuando se trata de relacionar dos o más variables, relacionadas entre sí. La Correlación implica el grado de dependencia de una variable respecto a otra y la Regresión es otra técnica que ayuda en la investigación.

El análisis de Regresión y Correlación simple muestra la relación entre dos variables, la variable independiente y la variable dependiente. Al usar sólo una variable independiente estamos ignorando la relación que pudiera tener la variable dependiente con otras variables independientes. Al estudio de la influencia de dos o más variables independientes ( $x$ ) sobre la variable dependiente ( $y$ ) se le llama análisis de Regresión y Correlación múltiple.

Las técnicas de Regresión, permiten hacer predicciones sobre los valores de cierta variable “ $y$ ” (dependiente), a partir de los de otra “ $x$ ” (independiente), entre las que intuimos que existe una relación. (Montgomery, Peck & Vining, 2006, p.13)

#### **4.3.1. Análisis de Regresión**

La Regresión, es un procedimiento por medio del cual se trata de determinar si existe relación de dependencia o no entre dos o más variables. En un Análisis de Regresión simple existe una variable dependiente ( $y$ ) que puede ser el número de especies, la abundancia o la presencia-ausencia de una sola especie y una variable explicativa o independiente ( $x$ ). El propósito es obtener una función sencilla de la variable explicativa ( $x$ ), que sea capaz de describir lo más ajustadamente posible la variación de la variable dependiente ( $y$ ). Como los valores observados de la variable dependiente difieren generalmente de los que predice la función, ésta posee un error.

La función más eficaz es aquella que describe la variable dependiente con el

menor error posible o, dicho en otras palabras, con la menor diferencia entre los valores observados y predichos. La diferencia entre los valores observados y predichos (el error de la función) se denomina variación residual o residuos.

Para estimar los parámetros de la función se utiliza el ajuste por mínimos cuadrados. Es decir, se trata de encontrar la función en la cual la suma de los cuadrados de las diferencias entre los valores observados y esperados sea menor. Sin embargo, con este tipo de estrategia es necesario que los residuos o errores estén distribuidos normalmente y que varíen de modo similar a lo largo de todo el rango de valores de la variable dependiente.

Estas suposiciones pueden comprobarse examinando la distribución de los residuos y su relación con la variable dependiente. (Spiegel & Stephens, 2009, p.345)

Cuando la variable dependiente es cuantitativa (por ejemplo, el número de especies) y la relación entre ambas variables sigue una línea recta, la función es del tipo:

$$y_i^* = a + bx_i, \text{ donde:}$$

“a” es el intercepto o valor del punto de corte de la línea de regresión con el eje de la variable dependiente (una medida del número de especies existente cuando la variable ambiental tiene su mínimo valor), “b” es la pendiente o coeficiente de regresión (la tasa de incremento del número de especies con cada unidad de la variable ambiental considerada).

Si la relación no es lineal pueden transformarse los valores de una o ambas variables para intentar transformarla en lineal. Si no es posible convertir la relación en lineal, puede comprobarse el grado de ajuste de una función polinomial más compleja. La función polinomial más sencilla es la cuadrática:

$$Y = a + bx + bx^2$$

Que describe una parábola, pero puede usarse una función cúbica u otra de un orden aun mayor capaz de conseguir un ajuste casi perfecto a los datos. Cuando la

variable dependiente se expresa en datos cualitativos (presencia-ausencia de una especie) es aconsejable utilizar las regresiones logísticas:

$$Y = [\exp(a + bx)] / [1 + \exp(a + bx)].$$

De lo anteriormente dicho, se deduce, que el análisis de regresión simple, estudia el comportamiento de una variable dependiente (y) en función de una variable independiente (x) de manera tal que se pueda precisar la relación entre dichas variables, con el propósito de hacer pronósticos o predicciones. La relación entre dos variables se puede determinar mediante un gráfico o diagrama de dispersión, un modelo de línea recta a través del método de los mínimos cuadrados y mediante un contraste de hipótesis.

### ***Diagrama de Dispersión***

Cuando los valores de la variable son llevados a un eje de coordenadas cartesianas, se puede apreciar un conjunto de puntos que muestran a simple vista la relación entre las variables, el cual recibe el nombre de diagrama de dispersión; es decir, no es más que graficar los pares de puntos de la variable, lo cual le permite al investigador, aproximar que tipo de relación hay entre las variables. En otras palabras, el diagrama de dispersión, es un trazo en un sistema de coordenadas rectangulares o cartesianas de manera tal que se grafican los puntos apareados de las variables estudiadas.

Este gráfico permite precisar si existe relación o no entre las variables, y si la misma es positiva o negativa (ascendente o descendente).

**Ecuación de Regresión:** es una ecuación que define la relación lineal entre dos variables.

$$y_i^* = a + bx_i$$

Dónde:

$y_i^*$  : Es el valor pronosticado de la variable “y” para un valor seleccionado de “x”.

$a$  : Es la ordenada de la intersección con el eje “y”, o sea el valor estimado de “y” cuando  $x=0$ . Es decir, corresponde al valor estimado de “y”, donde la recta de

regresión cruza el eje “y”, cuando “x” es igual a cero.

$b$  : Es la pendiente de la recta, o sea, el cambio promedio en “y\*” por unidad de cambio (incremento o decremento) en la variable independiente “x”.

$x_i$  : Es cualquier valor seleccionado de la variable independiente.

Las fórmulas para  $b$  y  $a$  son:

Pendiente de la línea de regresión:

$$b = \frac{n\sum y_i x_i - \sum y_i \sum x_i}{n\sum x_i^2 - (\sum x_i)^2}$$

Intersección con el eje “x”:

$$a = \frac{\sum y_i - b\sum x_i}{n}$$

Dónde:

$x_i$  : Es un valor de la variable independiente.

$y_i$  : Es un valor de la variable dependiente.

$n$  : Es el número de elementos en la muestra.

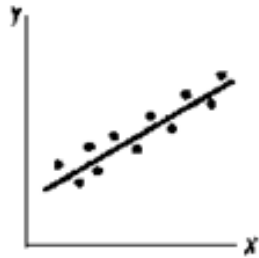
### ***Principio de Mínimos Cuadrados***

Es la técnica utilizada para lograr la ecuación de regresión, minimizando la suma de los cuadrados de las distancias verticales entre los valores verdaderos de "y" y los valores pronosticados "Y". Esta técnica nos permite seleccionar la línea recta que mejor se ajusta a los datos de manera tal que la suma de los cuadrados de la diferencia de cada valor observado y cada valor esperado (tendencia) sea la misma. (Montgomery, Peck & Vining, 2006, p.13)

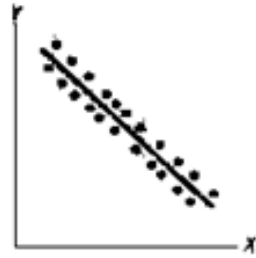
### ***Trazo de la Línea de Regresión***

El trazo de la línea recta, es una línea ajustada a un grupo de puntos para estimar la relación entre dos variables, como se puede observar en las figuras que se presentan a continuación: (Spiegel & Stephens, 2009, p.346)

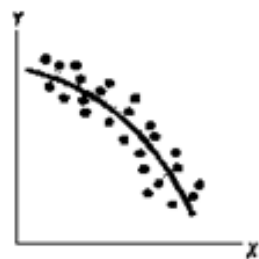
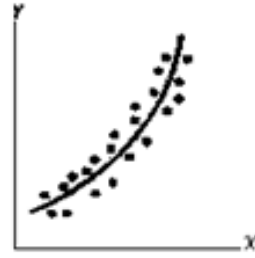
**Lineal directa**



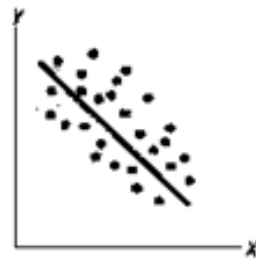
**Lineal inversa**



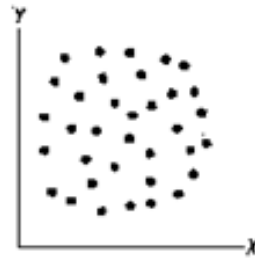
**Curvilínea directa**



**Curvilínea inversa**



**Lineal inversa con  
más dispersión**



**ninguna relación**

### ***Error Estándar de Estimación***

Es la medida de confiabilidad de la ecuación de estimación, que indica la variabilidad de los puntos observados alrededor de la línea de regresión, es decir hasta qué punto los valores observados difieren de sus valores predichos sobre la línea de regresión.

En otras palabras, el error estándar de estimación, es la medida de la dispersión de los valores observados, con respecto a la línea de regresión. (Spiegel & Stephens, 2009, p.347)

### 4.3.2. Análisis de Regresión Múltiple

Se puede definir como el proceso, a través de cual, se utilizan varias variables para predecir otra. Para este análisis, la ecuación tiene varias variables independientes:

$$y^* = a + b_1x_1 + b_2x_2 + \dots + b_kx_k$$

Dónde:

“  $x_1, x_2, \dots, x_k$  ” : Son las variables independientes.

“  $a$  ” : Es el punto donde la línea de regresión cruza el eje de las “ $y$ ”.

“  $b_1, b_2, \dots, b_k$  ” : Son los coeficientes de regresión (Douglas c. montgomery(2006); introducción al análisis de regresión lineal; 3ª edición; México; 61p).

### 4.3.3. Análisis de Correlación

La Correlación, es un conjunto de técnicas estadísticas utilizadas para medir la fuerza que existe entre la relación de dos variables.

Para (Levin & Rubin, 1996, p.680), “el análisis de correlación es una herramienta estadística que podemos usar para describir el grado hasta el cual una variable está linealmente relacionada con otra”.

El análisis de Correlación lineal o simple, permite medir y precisar la intensidad con que una variable se relaciona, es decir, su propósito primordial es encontrar que tan fuerte es la relación entre dos variables. Usualmente, el análisis de Correlación es usado junto con el análisis de Regresión para medir los cambios que explica la línea de regresión con respecto a la variable dependiente “ $y$ ”.

Pero, la Correlación, también se puede utilizar para medir el grado de asociación entre dos variables; y para realizar dicho análisis, es necesario contar con varias medidas estadísticas como lo son: el Diagrama de Dispersión, el Coeficiente de Correlación, el Coeficiente de Determinación y el Coeficiente de no Determinación.



## Diagrama de Dispersión

El Diagrama de Dispersión, no es más que una gráfica, en donde se describe en un plano cartesiano con una serie de puntos, la relación entre dos variables de interés. Donde, la variable dependiente se grafica sobre el eje vertical “y”, y la variable independiente sobre el eje horizontal “x”.

## Coefficiente de Correlación

El Coeficiente de Correlación, llamado también momento de Pearson, fue creado por Karl Pearson, aproximadamente en el año 1.900, el cual, es denotado con la letra “r”.

Dicho coeficiente, mide la intensidad de la relación entre dos variables y, tiene un campo de variabilidad o puede asumir valores entre -1 y +1, indicando éstos los puntos críticos de “r”.

Si  $r=1$ , la Correlación es perfecta positiva, implicando ésta una relación directamente proporcional, lo que quiere decir, que mientras una variable crece, la otra aumenta en la misma proporción. (Spiegel & Stephens, 2009, p.348)

Fórmula del Coeficiente de Correlación:

$$r = \frac{n\sum y_i x_i - \sum y_i \sum x_i}{\sqrt{\left((n\sum x_i^2) - (\sum x_i)^2\right) \left((n\sum y_i^2) - (\sum y_i)^2\right)}}$$

Dónde:

$n$  = Número de pares de observaciones

$\sum x$  = Suma de la variable independiente (x)

$\sum y$  = Suma de la variable dependiente (y)

$\sum x^2$  = Suma de los cuadrados de x

$(\sum x)^2$  = Cuadrado de la suma de x

$\sum y^2$  = Suma de los cuadrados de y

$(\sum y)^2$  = Cuadrado de la suma de y

### ***Coefficiente de Determinación***

Permite establecer si el modelo estudiado es confiable o no. Por medio del resultado de “r”, es posible calcular el Coeficiente de Determinación, el cual mide la proporción en que la variable dependiente explica los cambios ocurridos por la variable independiente. (Montgomery, Peck & Vining, 2006, p.37).

Su fórmula es la siguiente:

$$CD = r^2 * 100$$

El Coeficiente de Determinación puede solamente ser positivo y puede asumir valores entre 0 y 1 incluidos.

### **4.3.4. Análisis de Correlación Múltiple**

Al igual que el Análisis de Correlación Lineal o Simple, éste utiliza las mismas medidas estadísticas pero en este caso múltiples, para describir la relación entre las variables (el Diagrama de Dispersión, el Coeficiente de Correlación, el Coeficiente de Determinación y el Coeficiente de no Determinación).

### ***Coefficiente de Correlación Múltiple***

Se puede definir como una medida de la fuerza de la asociación entre la variable dependiente y dos o más variables independientes.

Este Coeficiente está representado con la letra “R”, y sólo puede tener valores comprendidos entre 0 y +1.

Si el Coeficiente está cercano a +1, quiere decir que hay una fuerte correlación entre la variable dependiente y las variables independientes. En caso de estar cercano a 0, indica que existe una débil correlación. (Spiegel & Stephens, 2009, p.384)

### ***Coefficiente de Determinación Múltiple***

Se simboliza con " R<sup>2</sup> " y, representa la proporción de la variación total en la variable dependiente “y”, que es explicada por las variables independientes. (Gujarati, 2010, p.196)

## **CAPÍTULO 5: DISEÑO DE EXPERIMENTOS**

### **5.1. Fundamentos de diseños experimentales**

El diseño de un experimento es la secuencia completa de los pasos que se deben tomar de antemano, para planear y asegurar la obtención de toda la información relevante y adecuada al problema bajo investigación, la cual será analizada estadísticamente para obtener conclusiones válidas y objetivas con respecto a los objetivos planteados. (Montgomery, 2004, p.1)

#### **5.1.1. Definición de un diseño experimental**

Un Diseño Experimental es una prueba o serie de pruebas en las cuales existen cambios deliberados en las variables de entrada de un proceso o sistema, de tal manera que sea posible observar e identificar las causas de los cambios que se producen en la respuesta de salida.

#### **5.1.2. Propósitos de un diseño experimental**

El propósito de cualquier Diseño Experimental es proporcionar una cantidad máxima de información pertinente al problema que se está investigando Y ajustar el diseño que sea lo más simple y efectivo; para ahorrar dinero, tiempo, personal y material experimental que se va a utilizar. Es de acotar, que la mayoría de los diseños estadísticos simples, no sólo son fáciles de analizar, sino también son eficientes en el sentido económico y en el estadístico.

De lo anterior, se deduce que el diseño de un experimento es un proceso que explica tanto la metodología estadística como el análisis económico.

#### **5.1.3. Tipos de tratamientos**

Es muy importante que cuando se elijan los tratamientos, éstos deben dar respuesta a una hipótesis de investigación. La hipótesis de investigación establece un conjunto de circunstancias y sus consecuencias. Los tratamientos deben ser una creación de las circunstancias para el experimento. Así, es necesario identificar los tratamientos con el papel que cada uno tiene en la evaluación de la hipótesis de

investigación. Por lo tanto, el investigador debe asegurarse que los tratamientos elegidos concuerden con la hipótesis de investigación.

#### **5.1.4. Tipos de error experimental**

Todos los experimentos están sujetos a posibles errores; los cuales se pueden disminuir, pero no controlar totalmente y pueden ser de los siguientes tipos:

Errores de experimentación.

Errores de observación

Errores de medición.

Variación natural en las unidades experimentales

La interacción de los tratamientos con las unidades experimentales.

Factores extraños que pueden influir en las características de la investigación.

La metodología estadística y un planeamiento adecuado del experimento permitirá reducir el error experimental, y esto hará posible que sean detectables las diferencias significativas entre los tratamientos que han sido usados. Si el investigador no hace el planeamiento del experimento en forma correcta, el error experimental del experimento va a ser grande y no se podrá detectar las diferencias significativas en el experimento. Por lo tanto, el investigador concluirá de manera equivocada que todos los tratamientos tienen el mismo efecto y que ningún tratamiento es ni mejor ni peor que los otros tratamientos.

#### **5.1.5. Principios básicos**

Los tres principios básicos del Diseño de un Experimento son: (Montgomery, 2004)

##### **1. Replicación (Obtención de Réplicas).**

Este principio se refiere al número de veces que se aplica un tratamiento a las unidades experimentales. El número de réplicas dependerá que tan costoso sea económicamente el introducir una nueva réplica en el experimento; pero se recomienda que sea un número lo más razonablemente posible; para poder así obtener inferencias estadísticas satisfactorias con respecto al efecto de los

tratamientos que se están estudiando. Entonces, el número de réplicas está determinado por las restricciones que se pueden asignar al problema.

**Aleatorización:** Consiste en el método por el cual las unidades experimentales reciben las aplicaciones de los tratamientos en forma aleatoria; es decir, que tanto la asignación del material experimental como el orden en que se realizan las pruebas individuales o ensayos se determinan aleatoriamente.

Al realizar la aleatorización adecuadamente en el experimento, se ayuda a "cancelar" los efectos de factores extraños que pudieran estar presentes; además simula el efecto de independencia y permite proceder como si las observaciones fueran independientes y con distribución normal; es decir, asegura el cumplimiento del supuesto de independencia del análisis de varianza. El método de aleatorización depende del Diseño de Experimento que será usado.

**Control Local:** Consiste en el uso de técnicas de bloqueo, balanceo y agrupamiento de las unidades experimentales para asegurar que el diseño usado sea eficiente; ya que los objetivos de la mayoría de los experimentos son las comparaciones claras y exactas entre los tratamientos a través de un conjunto apropiado de condiciones.

**Técnica de bloqueo:** Las unidades experimentales se distribuyen en grupos de unidades similares, con base en un factor o factores que se espera o se sabe que tienen alguna relación con la variable respuesta o con la medición que se supone responde de manera diferente a los diversos tratamientos.

**Técnica de balanceo:** Es el bloqueo y la asignación de los tratamientos a las unidades experimentales de modo que resulte una configuración balanceada. La comparación precisa entre los tratamientos requiere la selección de unidades experimentales uniformes para reducir el error experimental.

**Técnica de agrupamiento:** Es la colocación de un conjunto de unidades experimentales homogéneas en grupos, de modo que los diferentes grupos puedan sujetarse a distintos tratamientos. Estos grupos pueden constar de diferente número de unidades experimentales.

En los tres principios analizados anteriormente el objetivo principal es disminuir en gran medida la Variabilidad Natural o error experimental.(p.11)

#### **5.1.6. Forma de aleatorizar un experimento**

Se puede utilizar el siguiente método: (Montgomery, 2004)

Asignar números a las unidades experimentales.

Elaboran tarjetas de papel con los mismos números de las unidades experimentales y colocarse en un recipiente.

Sacar al azar una por una las tarjetas del recipiente.

Los primeros “n” números son las unidades experimentales asignadas al primer tratamiento.

Los segundos “n” números corresponden a las unidades experimentales que se asignan al segundo tratamiento. Y así sucesivamente, hasta sacar todas las tarjetas de “n” en “n”. (p.12)

#### **5.1.7. Directrices o procedimiento para el diseño experimental**

Se considera necesario que todos los estudiantes que realizan un enfoque estadístico en el cual se diseña y analiza un experimento tengan de antemano una idea clara de lo qué es exactamente lo que se va a estudiar; es decir, como se van a recopilar los datos, y como se van a analizar. (Montgomery, 2004)

**Primero:** Reconocimiento y Planeamiento del problema:

Es necesario desarrollar todas las ideas sobre los objetivos del experimento. Una clara comprensión y planteamiento del problema con frecuencia contribuye sustancialmente a un mayor entendimiento del fenómeno y a la solución final del problema.

**Segundo:** Elección de Factores y Niveles.

Se deben seleccionar los factores que van a ser investigados en el experimento, los intervalos de variación y los niveles específicos a los cuales se hará el experimento.

Además, debe considerarse la forma en que se controlarán estos factores para mantenerlos en los valores deseados y como se les medirá.

**Tercero:** Selección de la Variable Respuesta.

El experimentador debe seleccionar la variable respuesta o variable dependiente de tal forma que esté seguro que la respuesta, que se va a medir, realmente proporcione información útil a cerca del problema en estudio.

Las respuestas en un problema pueden ser múltiples y además con mayor frecuencia, el promedio o la desviación estándar (o ambas) de la característica medida serán la variable respuesta.

**Cuarto:** Elección del Diseño Experimental.

Para definir o determinar el diseño experimental a utilizar se debe considerar el tamaño muestral (número de repeticiones), seleccionar un orden adecuado para los ensayos experimentales y determinar si hay implicaciones de bloque u otras restricciones de aleatorización.

Además, tener presente los objetivos experimentales.

**Quinto:** Ejecución del Experimento.

En esta fase se lleva a cabo la recolección de los datos. Se debe observar cuidadosamente el proceso para asegurar que todo se realice conforme lo planteado; ya que los errores en el procedimiento suelen anular la validez del experimento.

**Sexto:** Análisis de los Datos.

Se deben utilizar métodos estadísticos para analizar los datos, para que los resultados y conclusiones sean objetivos más que apreciativos. Si todo se ha realizado correctamente los métodos estadísticos que se necesitan no son complicados. El análisis de residuos y la verificación de la idoneidad del modelo son también técnicas de análisis de gran utilidad.

**Séptimo:** Conclusiones y Recomendaciones.

Consiste en la interpretación de las inferencias estadísticas. Y para llevar a cabo la presentación de los resultados son muy útiles los métodos gráficos, en especial

cuando se presentan a otras personas. Es bien importante también realizar corridas de seguimiento o pruebas de confirmación para validar las conclusiones del experimento. (p.15)

#### **5.1.8. Utilización de los métodos estadísticos en la experimentación**

El uso de los métodos estadísticos puede incrementar la eficiencia de los experimentos y, ayudar a justificar las conclusiones que se obtienen. La utilización de las técnicas estadísticas en la experimentación requiere que el investigador considere los siguientes puntos: (Montgomery, 2004)

##### **Uso del conocimiento no estadístico del problema.**

Muchas veces se puede utilizar una gran cantidad de teoría para explicar las relaciones que hay entre los factores y la variable respuesta. Este tipo de conocimiento no estadístico se debe tomar en cuenta para elegir los factores y las respuestas, también al decidir el número de réplicas que se quieren realizar, al analizar los datos, etc.

##### **Mantener el Diseño y el Análisis tan simple como sea posible.**

Casi siempre, lo más adecuado son los métodos de diseño y análisis estadístico más simples. Por lo tanto, es recomendable el uso de técnicas estadísticas poco complejas y muy refinadas.

Si se realiza el diseño cuidadosamente y correctamente, el análisis se espera que sea relativamente sencillo.

##### **Reconocer la diferencia entre la significación práctica y estadística.**

No hay seguridad de que una diferencia sea suficientemente grande, desde el punto de vista práctico, por el sólo hecho de que dos condiciones experimentales producen respuestas medias, estadísticamente diferentes.

##### **Usualmente los experimentos son iterativos.**

Por lo general, al principio de un experimento no se está en capacidad de definir estos aspectos, pero es posible conocerlos a medida que el experimento avanza. Esto



favorece al empleo del enfoque iterativo o secuencial; pero por regla general, la mayoría de los experimentos son iterativos. (p.19)

## **5.2. Diseños Unifactoriales**

### **5.2.1. Descripción.**

En el análisis de los resultados de los experimentos se pueden observar diferentes aplicaciones de los Diseños Experimentales. Hay experimentos muy útiles en los cuales existe un sólo factor de interés; el cual se analiza por medio de la comparación de dos condiciones que intervienen en el Experimento (a menudo llamadas tratamientos o niveles del factor); a este tipo de experimentos se le denomina Experimentos de Comparación Simple.

El análisis de los datos de este tipo de Experimentos resulta ser sencillo, ya que se utilizan técnicas de la Inferencia Estadística, llamada Prueba de Hipótesis (o pruebas de significancia) que son las que ayudan al experimentador a comparar estas condiciones. Si en el tipo de Diseño Experimental planteado anteriormente se requiere más de dos niveles del factor que se analiza, éstos son considerados como "Diseños Unifactoriales". Teniendo en cuenta que para el análisis de éstos se utiliza el Análisis de Varianza, ya que se requiere probar la igualdad de varias medias, la cual se explicará posteriormente.

En los experimentos de los Diseños unifactoriales, el número de observaciones recolectadas en cada tratamiento pueden ser iguales o diferentes. Cuando el número de observaciones sea diferente se dice que el Diseño está Desequilibrado o Desbalanceado; en caso contrario el Diseño está Equilibrado o Balanceado.

### 5.2.2. Representación simbólica de los datos.

Bajo el supuesto que se tiene "a" niveles o "a" tratamientos de un único factor, la respuesta que se observa en cada uno de los "a" tratamientos es una variable aleatoria. La representación típica de los datos para un experimento Unifactorial, se presenta a continuación. (Walpole, 1999, p.509)

**Tabla 5:** Ejemplo de tratamientos.

Tratamiento (Nivel)	Observaciones	Totales (y <sub>i</sub> )	Promedios ( $\bar{y}_i$ )
1	y <sub>11</sub> y <sub>12</sub> .....y <sub>1n</sub>	y <sub>1.</sub>	$\bar{Y}_1$ .
2	y <sub>21</sub> y <sub>22</sub> .....y <sub>2n</sub>	y <sub>2.</sub>	$\bar{Y}_2$ .
3	y <sub>31</sub> y <sub>32</sub> .....y <sub>3n</sub>	y <sub>3.</sub>	$\bar{Y}_3$ .
.	.	.	.
.	.	.	.
.	.	.	.
A	y <sub>a1</sub> y <sub>a2</sub> .....y <sub>an</sub>	$\frac{y_{a.}}$	$\frac{\bar{Y}_a.}$
		Y <sub>...</sub>	$\bar{Y}_{...}$

### 5.2.3. Modelo estadístico

Sea "y" la variable que se va a medir en las distintas unidades experimentales y "y<sub>ij</sub>" el valor de la j-ésima observación del tratamiento "i". Se pueden describir las observaciones de la tabla anterior por medio del siguiente Modelo Estadístico Lineal. (Walpole, 1999, p.510)

$$y_{ij} = \mu + l_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases}$$

Dónde:

y<sub>ij</sub>: Es la observación ij-ésima

μ : Es un parámetro común para todos los tratamientos, llamado media general.

l<sub>i</sub> : Es un parámetro asociado con el tratamiento i-ésimo denominado efecto del tratamiento i-ésimo.

ε<sub>ij</sub> : Es la componente del error aleatorio (variabilidad natural).

$i$ : Varía de 1 hasta el número de Tratamientos ( $a$ ).

$j$ : Varía de 1 hasta el número de réplicas de cada tratamiento ( $n$ ).

El objetivo será probar hipótesis adecuadas con respecto a los efectos del tratamiento y hacer estimaciones de ellos. Para llevar a cabo esta prueba de hipótesis, se debe suponer que los errores del modelo son variables aleatorias independientes con distribución normal, con media cero y varianza  $\sigma^2$ . Se supone que " $\sigma^2$ " es constante para todos los niveles del factor. El Modelo Estadístico recibe el nombre de "Análisis de Varianza de Clasificación Unidireccional", debido a que se investiga un sólo factor.

#### 5.2.4. Sumas y medias de cuadrados

El término "Análisis de Varianza" consiste en la descomposición de la variabilidad total de los datos, en sus partes que la forman. (Walpole, 1999, p.511)

Sea:

$SS_T$ : Suma Total de Cuadrados Corregida.

$SS_{Tratamientos}$ : Suma de cuadrados debida a los tratamientos (entre tratamientos).

$SS_E$ : Suma de Cuadrados debida al Error (dentro de los tratamientos).

$$SS_T = SS_{Tratamientos} + SS_E$$

Dónde:

$SS_T$ : Tiene  $N-1$  grados de libertad porque existe un total de  $N = an$  observaciones, y un sólo parámetro a estimar que es  $\mu$ .

$SS_{Tratamientos}$ : Tiene  $a - 1$  grados de libertad porque existen " $a$ " niveles del factor (y " $a$ " medias de tratamientos) y sólo un parámetro a estimar que es  $\mu$ .

$SS_E$ : Tiene  $N - a$  grados de libertad porque existe " $n$ " réplicas dentro de cada tratamiento, las cuales proporcionan  $n-1$  grados de libertad para estimar el error experimental. Como hay " $a$ " tratamientos, se tiene  $a(n-1) = an - a = N - a$

$$SS_{Tratamientos} = \sum_{i=1}^a \frac{y_i^2}{n} - \frac{y_{..}^2}{N}$$

$$SS_T = \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - \frac{y_{..}^2}{N}$$

$$SS_E = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_i.)^2$$

Las Medias de Cuadrados no son más que la suma de cuadrados divididos por sus respectivos grados de libertad. Matemáticamente las medias de cuadrados se definen de la siguiente manera:

$$MS_F = \frac{SS_F}{a-1}$$

$$MS_E = \frac{SS_E}{a(n-1)}$$

Puede demostrarse que el cuadrado medio del error  $MS_E$  estima la varianza del error experimental. Además  $MS_F$  estima la varianza sólo si todas las medias de los niveles del factor son iguales, pero si son diferentes,  $MS_F$  será mucho mayor que la varianza  $\sigma^2$ . Para probar lo anterior se usa la prueba F de varianzas.

$$F_0 = \frac{MS_F}{MS_E}$$

Si,  $F_0 > F_{a, a-1, a(n-a)}$  se puede concluir que las medias de los niveles del factor son diferentes. Si se hace el análisis en computadora, también se puede concluir observando el valor de P. A continuación se muestra la tabla de Análisis de Varianza (ANOVA)

**Tabla 6:** Análisis de Varianza (ANOVA).

<b>Fuente de Variación</b>	<b>Suma de cuadrados</b>	<b>Grados de libertad</b>	<b>Cuadrado medio</b>	<b>Fo .</b>
Entre Niveles Del factor	$SS_F$	$a - 1$	$MS_F$	$Fo = MS_F / MS_E$
Error (Dentro de Niveles de Factores)	$SS_E$	$a(n - 1)$	$MS_E$	
<b>Total</b>	$SS_T$	$an - 1$		

### 5.2.5. Análisis Residual

El análisis de varianza (ANOVA) asume que las observaciones son normales e independientemente distribuidas con la misma varianza en cada nivel del factor. Esto puede verificarse por inspección de los residuos. En este caso el residuo es  $e_{ij} = y_{ij} - \bar{y}_i$ , o sea la diferencia entre una observación y la media correspondiente al nivel del factor.

La suposición de normalidad se checa al graficar los residuos en papel de probabilidad normal. Para checar el supuesto de igual varianza en cada nivel del factor, graficar los residuos contra los niveles del factor y comparar la dispersión de los residuos. Es útil graficar los residuos para verificar con  $\bar{y}_i$ , (valor ajustado), la variabilidad de los residuos debe ser aleatoria.

### 5.2.6. Comparación de parejas de medias de tratamiento.

En un experimento el investigador puede estar interesado en comparar todas las parejas de “a” medias de tratamiento y que la hipótesis nula que se desea probar es  $H_0: \mu_i = \mu_j$  y las alternativas  $H_1: \mu_i \neq \mu_j$  para toda  $i \neq j$ , esta comparación se lleva a cabo a través de las siguientes pruebas. (Montgomery, 2004, p.96)

### ***Prueba de intervalos múltiples de Duncan***

Con este método se prueba las verdaderas diferencias que existen entre los pares de medias. Este método es muy eficiente para detectar diferencias entre medias siendo estas diferencias en realidad existen. (Walpole, 1999, p.527)

#### **Pasos para realizarla:**

**Paso 1:** se colocan los promedios de tratamiento en orden ascendente y se determina el error estándar de cada promedio de

$$S_{\bar{y}_i} = \sqrt{\frac{MS_E}{n}}$$

**Paso 2:** se calculan los intervalos de significancia ( anexo 6)

$$r_\alpha = (p, f)$$

Donde

$$p = 1, 2, \dots, a,$$

$\alpha$  = nivel de significancia

$f$  = numero de grados de libertad del error

Paso 3: calcular los intervalos de significancia

$$R_p = r_\alpha = (p, f) * S_{\bar{y}_i}$$

**Paso 4:** realizar las comparaciones de medias y las diferencias de medias

Se realiza la diferencia del valor más alto menos el valor pequeño y luego el segundo más pequeño y así sucesivamente hasta realizar todas las comparaciones. Se concluye que el par de medias en estudio es significativamente diferente si la diferencia observada es mayor que el intervalo significativo correspondiente.

### ***Prueba de tukey***

Este método está basado en el de intervalos. El procedimiento consiste en el uso de  $q_\alpha(a, f)$  para encontrar el valor crítico de todas las comparaciones por pares, importar cuantas medias estén en el grupo. (Walpole, 1999, p.526)

Por lo tanto, se declaran dos medias significativamente diferentes si:

$$\| \bar{y}_i - \bar{y}_j \| > T_\alpha, \text{ donde } T_\alpha = q_\alpha(a, f) * S_{\bar{y}_i} \text{ con } S_{\bar{y}_i} = \sqrt{\frac{MSE}{n}}$$

Se calcula el valor crítico  $q_\alpha(a, f)$ , de igual manera que en la prueba de Duncan solo que en este caso se utiliza la tabla (anexo 5)

Dónde:

$\alpha$  = nivel de significancia

$f$  = numero de grados de libertad del error

$a$  = el numero de tratamientos

El proceso es parecido a la prueba de Duncan, la diferencia es que en esta prueba solo se compara con  $T_\alpha$  para verificar si existe diferencia significativa.

### ***Prueba de dunnet***

En muchos problemas científicos y de ingeniería no nos interesa hacer inferencias acerca de todas las comparaciones posibles entre las medias de los tratamientos del tipo  $\mu_i - \mu_j$ . En vez de ello es frecuente que el experimento dicte la necesidad de comparar simultáneamente cada tratamiento con un control. Un procedimiento de prueba desarrollado por C. W. Dunnett determina diferencias significativas entre cada media de tratamiento y el control, con un solo nivel conjunto de significancia  $\alpha$ . (Walpole, 1999, p.528)

**Pasos 1:** calcular los valores

$$d_i = \frac{\bar{y}_i - \bar{y}_j}{\sqrt{\frac{2S^2}{n}}} \quad i = 1, 2, \dots, a$$

Dónde:

$S^2$ : Representa la varianza muestral se obtiene a partir del cuadrado medio del error en el análisis de varianza

**Paso 2:** criterio para considerar si tienen diferencia significativa

$$\|d_i\| = d_{\alpha/2}(k, f)$$

Dónde:

$\alpha$  = nivel de significancia

$f$  = numero de grados de libertad del error

$a$  = el numero de tratamientos

### 5.3. Bloqueo y factores extraños

En muchos casos es necesario diseñar el experimento de tal forma que la variabilidad que surge de variables extrañas pueda ser controlada. El procedimiento general para el diseño de bloques totalmente aleatorizado consiste en seleccionar "**b**" bloques y correr una réplica completa de experimentos en cada uno de los bloques. Habrá "**a**" observaciones (una por cada nivel del factor) en cada bloque, y el orden de realización de los experimentos será aleatorio dentro de cada bloque. Las observaciones pueden ser representadas por el modelo. (Walpole, 1999, p.535)

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \end{cases}$$

Dónde:

$\mu$  = Media general

$\tau_i$  = Efecto del nivel **i-ésimo** del factor. La suma de sus desviaciones respecto a la media general es cero.

$\beta_j$  = Efecto del bloque **j-ésimo**. La suma de sus desviaciones respecto a la media general es cero.

$\varepsilon_{ij}$  = NID(0,  $\sigma^2$ ) Es el término de error.

Se trata de probar la Hipótesis de igualdad de efectos del factor, o sea:

$$H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0$$

$$H_1: \tau_i \neq 0, \text{ para al menos una "i"}$$



La suma de cuadrados total de donde se pueden desglosar las sumas de cuadrados de los factores, bloques y error es:

$$SST = \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2 = b \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + a \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{.j} - \bar{y}_{i.} + \bar{y}_{..})^2$$

Con los siguientes grados de libertad correspondientes:

$$ab - 1 = (a - 1) + (b - 1) + (a - 1)(b - 1)$$

La Hipótesis nula de que no hay diferencia entre los niveles del factor se prueba con la prueba  $F_0 = MS_{factor} / MS_{Error}$ .

#### 5.4. Diseño de cuadro latino

En este diseño, la distribución aleatoria de los tratamientos se restringe más ampliamente mediante la agrupación de los mismos, tanto en columnas como en hileras (bloques). Así resulta posible eliminar la variabilidad del error experimental asociada con ambos efectos. Cada tratamiento ocurre el mismo número de veces (usualmente una vez) en cada hilera y columna y proporcionará una comparación más precisa de los efectos del tratamiento.

Un cuadro Latino requiere al menos tantas repeticiones como tratamientos existan; por tanto, no resulta práctico para experimento con un gran número de tratamientos, con una sola unidad experimental por tratamiento en cada columna e hilera.

Este diseño se ha usado con ventaja en muchos campos de investigación donde hay dos fuentes principales de variación en la realización de un experimento.

Los cuadros latinos más comunes van de 5x5 a 8x8; cuadros mayores de 12x12 se usan muy rara vez. En los cuadros latinos, como en los bloques al azar, a medida que aumenta el tamaño del bloque, el error experimental por unidad probablemente aumente. Los cuadros latinos pequeños proporcionan pocos grados de libertad para estimar el error experimental y así debe lograrse una disminución sustancial en el error para compensar el corto número de grados de libertad.

La aleatorización en el cuadro latino consiste en elegir un cuadro latino al azar entre todos los cuadros latinos posibles. Fisher y Yates dan el conjunto completo de cuadros

latinos desde 4x4 hasta 6x6 y muestran cuadros hasta de tamaño 12x12. Cochran & Cox dan cuadros latinos demuestra desde 3x3 hasta 12x12. (Montgomery, 2004, p.144)

La representacion de losdatos para este tipo de diseño es la siguiente:

renglones	columnas			
	1	2	.....	$y_{i..}$
1	$A=y_{111}$	$B=y_{122}$	.....	$y_{1..}$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
$y_{..k}$	$y_{..1}$	$y_{..2}$	.....	$y_{...}$

Este cuadro puede ser de “p x p”.

Dónde:

$y_{..k}$  : Total por columnas

$y_{.j}$ : Total de tratamientos

$y_{i..}$  : Toral por renglón

$y_{...}$ : Total general

### 5.4.1. Modelo Estadístico del cuadrado latino

El modelo estadístico lineal que resulta de un diseño de cuadrado latino es el siguiente:

$$y_{ijk} = \mu + \alpha_i + \tau_j + \varepsilon_{ijk} \begin{cases} i = 1, 2, \dots, p \\ j = 1, 2, \dots, p \\ k = 1, 2, \dots, p \end{cases}$$

Dónde:

$y_{ijk}$ = Es la observación ijk-esima.

$\mu$  = Media general.

$\alpha_i$ = Es el i-ésimo efecto del renglón.

$\tau_j$  = Es el j-esimo efecto del tratamiento.

$\beta_j$  = Es el k-ésimo. Efecto de columna.

$\varepsilon_{ijk}$  = Es el termino usual NID(0,  $\sigma^2$ ) Es el término de error.

#### 5.4.2. Sumas y Medias de Cuadrados

$$SS_T = \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p y_{ijk}^2 - \frac{y_{...}^2}{N}$$

$$SS_{\text{Reglones}} = \sum_{i=1}^p \frac{y_{i..}^2}{p} - \frac{y_{...}^2}{N}$$

$$SS_{\text{Tratamientos}} = \sum_{j=1}^p \frac{y_{.j.}^2}{p} - \frac{y_{...}^2}{N}$$

$$SS_{\text{Columnas}} = \sum_{k=1}^p \frac{y_{..k}^2}{p} - \frac{y_{...}^2}{N}$$

$$SS_E = SS_T - SS_{\text{Reglones}} - SS_{\text{Tratamiento}} - SS_{\text{Columna}}$$

$$MS_{\text{Tratamiento}} = \frac{SS_{\text{Tratamiento}}}{p - 1}$$

$$MS_{\text{Reglones}} = \frac{SS_{\text{Reglones}}}{p - 1}$$

$$MS_{\text{Columnas}} = \frac{SS_{\text{Columnas}}}{p - 1}$$

$$MS_E = \frac{SS_E}{(p - 2)(p - 1)}$$

### 5.4.3. Análisis Estadístico

La hipótesis que se prueba siempre tiene que ser en relación al a los tratamientos (letras latinas).

Las hipótesis a probar serán:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_a = 0$$

$$H_1: \mu_i \neq 0, \text{ para al menos un par } "(i, j)"$$

La hipótesis nula " $H_0$ " es frecuentemente la de interés central, por lo tanto se prueba de la siguiente manera:

$$F_0 = \frac{MS_{Tratamiento}}{MS_E}$$

De modo tal que la hipótesis nula se rechaza si:

$$\text{Si } F_0 > F_{\alpha, (p-1), (p-2)(p-1)}$$

Donde  $F_0$  se obtiene a través de la varianza y  $F_{\alpha, (p-1), (p-2)(p-1)}$  se obtiene a través de la tabla F.

**Tabla 7 :** Análisis de varianza (ANOVA).

<b>Fuente de Variación</b>	<b>Suma de cuadrados</b>	<b>Grados de libertad</b>	<b>Cuadrado medio</b>	<b>Fo.</b>
Tratamientos	$SS_{Tratamientos}$	$p - 1$	$MS_{Tratamientos}$	$F_0 = \frac{MS_{Tratamiento}}{MS_E}$
Renglones	$SS_{Renglones}$	$p - 1$	$MS_{Renglones}$	
Columnas	$SS_{Columnas}$	$p - 1$	$MS_{Columnas}$	
Error	$SS_E$	$(p - 2)(p - 1)$	$MS_E$	
<b>Total</b>	$SS_T$	$p^2 - 1$		

## 5.5. Diseño de experimentos factoriales y fraccionales

Muchos experimentos para mejoramiento involucran varias variables, existen diversos diseños para esas situaciones. A continuación, se tratarán algunos.

### 5.5.1. Experimentos factoriales

Se utilizan cuando hay varios factores de interés en un experimento. En los experimentos factoriales se deben utilizar en cada réplica, todas las combinaciones de los factores que se están investigando. Si se tienen dos factores A y B con niveles **a** y **b** respectivamente, cada réplica contendrá todas las posibles combinaciones **ab**. (Montgomery, 2004, p.170)

El efecto de un factor se define como el cambio en la respuesta como resultado de un cambio en el nivel del factor. A esto se denomina efecto principal ya que se refiere a los factores primarios del estudio.

Para el caso de experimentos con dos factores la suma de cuadrados ahora se va a formar de la suma de cuadrados del factor A, la suma de cuadrados del factor B, la suma de cuadrados de la interacción AB y la suma de cuadrados del error, como sigue:

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

La tabla de Análisis de varianza (ANOVA) queda como sigue:

**Tabla 8 :** Análisis de varianza (ANOVA).

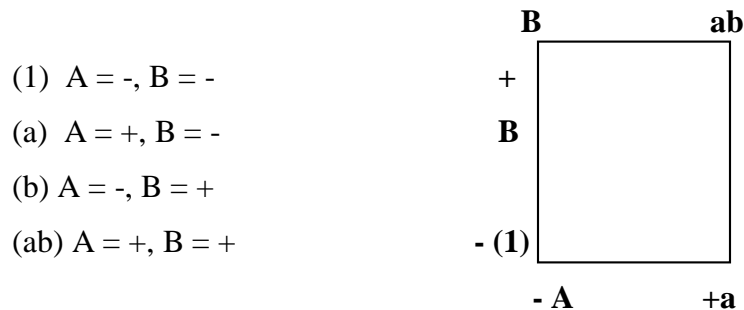
<b>Fuente de Variación</b>	<b>Suma de cuadrados</b>	<b>Grados de libertad</b>	<b>Cuadrado medio</b>	<b>Fo .</b>
Factor A	$SS_A$	$a - 1$	$MS_A$	$Fo = MS_A / MS_E$
Factor B	$SS_B$	$b - 1$	$MS_B$	$Fo = MS_B / MS_E$
Interacción AB	$SS_{AB}$	$(a - 1)(b - 1)$	$MS_{AB}$	$Fo = MS_{AB} / MS_E$
Error	$SS_E$	$ab(n - 1)$	$MS_E$	
<b>Total</b>	<b><math>SS_T</math></b>	<b><math>abn - 1</math></b>		

## 5.6. El diseño factorial $2^k$

Un diseño muy común es el que considera varios factores (K) en dos niveles (“alto” y “bajo”), el diseño más simple es el diseño  $2^k$ . (Montgomery, 2004, p.218)

### 5.6.1. El diseño $2^2$

Este el diseño más simple, usa dos factores A y B, cada uno en dos niveles. Requiere la realización de 4 experimentos que pueden replicarse “n” veces, tomando las sumas de los resultados de las réplicas se tiene:



Los efectos de interés en el diseño son los efectos principales de A y B y los efectos de la interacción AB, se denominan **contrastos** calculados como sigue:

$$A = \frac{a + ab}{2n} - \frac{b + (1)}{2n} = \frac{1}{2n} (a + ab - b - (1))$$

$$B = \frac{b + ab}{2n} - \frac{a + (1)}{2n} = \frac{1}{2n} (b + ab - a - (1))$$

$$AB = \frac{ab + (1)}{2n} - \frac{a + b}{2n} = \frac{1}{2n} (ab + (1) - a - b)$$

Las cantidades entre paréntesis se denominan **contrastos**, aquí los coeficientes siempre son +1 o -1. También se pueden determinar usando una tabla de signos como sigue:

Corrida	I	A	B	AB		
1	(1)	+	-	-	+	ANOVA
2	a	+	+	-	-	$Fo = SS_T / SS_E$
3	b	+	-	+	-	$gl. SS_T = 4n - 1$
4	ab	+	+	+	+	$gl. SSS_E = 4(n - 1)$

Para obtener la suma de cuadrados de A, B, y AB se usa:

$$SS = \frac{(\text{contraste})^2}{n \sum (\text{coeficientes de contrastes})^2}$$

Por tanto, la suma de cuadrados para A, B y AB es:

$$\left. \begin{aligned} SSA &= \frac{1}{4n} (a + ab - b - (1))^2 \\ SSB &= \frac{1}{4n} (b + ab - a - (1))^2 \\ SSAB &= \frac{1}{4n} (ab + (1) + -a - b)^2 \end{aligned} \right\}$$

### 5.6.2. Una sola réplica del diseño $2^k$

Conforme el número de factores en un diseño factorial se incrementa, también se incrementa el número de efectos que pueden ser estimados. Por ejemplo, un diseño experimental  $2^4$  tiene 4 efectos principales, 6 interacciones de dos factores, 4 interacciones de tres factores y 1 interacción de cuatro factores.

En muchas situaciones se aplica el principio de escasez de efectos donde el sistema es dominado por los efectos principales y las interacciones de bajo orden, por tanto, las interacciones de tercer o mayor, son normalmente insignificantes. Por lo anterior, cuando el número de factores es moderadamente grande ( $k \geq 4$ ), es práctica común correr sólo una réplica del diseño y juntar o combinar las interacciones de alto orden como una estimación del error.

### 5.6.3. Adición de puntos centrales al diseño $2^k$

Una preocupación al usar estos modelos, es que se asume que los efectos de los factores son casi lineales, sin embargo, si la curvatura es mucha, debe complementarse el modelo con términos de segundo orden. (Montgomery, 2004, p.271)

Para el caso de  $k = 2$ , un modelo que incluye efectos de segundo orden es el siguiente:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$$

La ecuación anterior se denomina **modelo de superficie de respuesta de segundo orden**, y los coeficientes  $\beta_{11}, \beta_{22}$  miden los efectos cuadráticos puros. Para ajustar el modelo se requieren tres niveles en todos los factores, siendo importante determinar si los términos cuadráticos puros son requeridos.

Existe un método de agregar un punto central al diseño factorial  $2^k$  para determinar si son necesarios los términos cuadráticos. Además, si este punto es replicado, se puede obtener una estimación independiente del error experimental. Los puntos centrales consisten de " $n_c$ " réplicas de experimentos realizados en  $x_i = 0, (i = 1, 2, \dots, k)$  Al agregar puntos centrales, no se tiene impacto en la estimación de los efectos de los factores. Se asume que los  $k$  factores son cuantitativos, de otra forma el punto central no existiría.

Para el caso de un diseño de 2 factores en 2 niveles y " $n_c$ " observaciones en los puntos centrales (0,0), si  $\bar{y}_F$  es el promedio de los cuatro experimentos en los cuatro puntos factoriales y  $\bar{y}_C$  el promedio de los " $n_c$ " experimentos con los puntos centrales. Si su diferencia es pequeña, entonces los puntos se encuentran cerca del plano que pasa por los puntos factoriales y no hay curvatura. Si la diferencia es grande, entonces se tiene una curvatura. La suma de cuadrados de la curvatura cuadrática pura con un grado de libertad es la siguiente:



$$SS_{\text{cuadrático puro}} = \frac{n_F n_C (\bar{y}_F - \bar{y}_C)^2}{n_F + n_C}$$

Donde en general " $n_f$ " es el número de puntos factoriales del diseño. Esta cantidad puede ser comparada a la media cuadrática del error para probar la curvatura.

## **CAPÍTULO 6: APLICACIONES DEL CAPITULO 4 Y CAPITULO 5**

**Ejemplo 1:**” Efecto antidepresivo y toxicidad del veneno de sapo sabanero (*Rhinella marina*) en ratones de laboratorio”

### **Directrices para el diseño del experimento**

Para ejemplificar las directrices de un diseño experimental, se considerará la tesis “Efecto antidepresivo y toxicidad del veneno de sapo sabanero “*Rhinella marina* (R.marina)” en ratones de laboratorio”.

Se considera necesario que el investigador tenga de antemano una idea clara de lo qué es exactamente lo que se va a estudiar; es decir, como se van a recopilar los datos, y como se van a analizar.

El resumen del estudio presentado es el siguiente: “Algunos estudios determinan que por lo menos cinco de cada diez personas sufren episodios depresivos por distintas razones. Sin embargo, la mayoría de medicamentos utilizados en el tratamiento de depresión causan efectos secundarios, entre estos están el aumento o pérdida de peso, disfunciones sexuales, náuseas, vómitos, diarreas, retención urinaria, confusión, sedación, debilidad, fatiga, delirio, convulsiones tónico-clónicas, hipomanía, entre otros.

Es por ello la importancia de buscar medicamentos de origen natural para evitar complicaciones en los tratamientos contra la depresión. En la presente investigación se buscó determinar el efecto antidepresivo del veneno de sapo sabanero que contiene la Bufotenina, sustancia química que por su parecido molecular a la serotonina tenga la posibilidad que presente los mismos efectos neurofarmacológicos. Las pruebas se realizaron en ratones de laboratorio de la cepa NIH (National Institute of Health), procedentes del laboratorio de Experimentación Animal del Centro de Investigación y Desarrollo en Salud (CENSALUD) de la Universidad de El Salvador (UES).

La toxicidad del veneno de sapo sabanero o sapo común (*Rhinella marina*) se reporta a una dosis letal media (DL50) de 400 mg/kg en ratones. Para evaluar el posible efecto antidepresivo del veneno se utilizó el modelo de Estrés Leve Crónico (ELC), para inducir a los animales de experimentación a un estado depresivo. Se utilizó Citalopram como

fármaco antidepresivo de control a una concentración de 20 mg/kg. Mientras que la concentración utilizada del veneno como antidepresivo fue de 50, 100 y 200 mg/kg.

Las evaluaciones para validar el modelo de Estrés Leve Crónico, así como para evaluar el efecto antidepresivo de la sustancia de estudio se realizaron con el “Test de Consumo de Sacarosa (TCS)” a una concentración de 1.5% y con el “Test de Nado Forzado (TNF)”. Los resultados de la prueba arrojan que el modelo de Estrés Leve Crónico utilizado es efectivo para la inducción a depresión, sin embargo, para la actividad biológica del veneno de “*Rhinella marina* (R. marina)”, se descarta su posible acción antidepresiva.”

### **Primero: Reconocimiento y Planeamiento del problema.**

Los autores medicas consideran la falta de atención que se le brindan a las enfermedades mentales en El Salvador, especialmente la depresión, que al no ser tratados a tiempo se tratan de corregir el comportamiento a través de fármacos que producen efectos secundarios en los pacientes, así es como surge la idea de este experimento, para analizar si es viable el tratamiento de bufotenina como un antidepresivo natural, y medir su nivel de toxicidad y sus efectos, para completar su investigación, se rigen bajo el objetivo general “evaluar el efecto antidepresivo y el efecto toxico del veneno de sapo sabanero (*Rhinella marina*) en ratones de laboratorio” y para poder lograrlo siguen los siguientes objetivos específicos:

Identificar los signos de la toxicidad del veneno de sapo sabanero (*Rhinella marina*) tras su administración en ratones de laboratorio.

Estandarizar el método de estrés leve crónico tras la aplicación de tres estresores (hacinamiento, inmovilización, e inclinación de la caja a 45°) a utilizar en ratones de laboratorio.

Comparar el efecto antidepresivo del Citalopram (fármaco antidepresivo) con el efecto del veneno de sapo sabanero en ratones de laboratorio.

Determinar la posible actividad antidepresiva del veneno de sapo sabanero *Rhinella marina*, por medio de la prueba de consumo de sacarosa.

Demostrar el efecto del veneno de sapo sabanero como antidepresivo en ratones de laboratorio, por medio de la prueba del Test de Nado Forzado

**Segundo: Elección de Factores y Niveles.**

El factor que se analizara es el de Estrés leve crónico (ELC), para ello se utilizaron tres estresores de depresión:

El primero: fue teniendo a los roedores en hacinamiento.

El segundo fue la inmovilización de los roedores por una hora.

El tercer estresor consistió en mantener las jaulas inclinadas a 45° durante una hora posterior al estresor de inmovilización.

Esto se realizó para 6 grupos de 5 ratones, los cuales se sometieron a prueba 2 veces al día, de 7:00-8:00am y de 2:00-3:00pm; por un periodo de 5 semanas.

El grupo 1 es tratado con agua destilada sin sometimiento a estrés, el grupo 2, tratado con agua destilada y sometido a estrés, el grupo 3 fue sometido a estrés con Citrapolam, 4, 5 y 6 fueron sometidos a estrés más las diferentes dosis de veneno de R. marina.

Estos factores se controlaron a través del test de nado forzado, y a través del test del consumo de sacarosa. El test de nado forzado consiste en colocar a un ratón en un recipiente cilíndrico lleno de agua durante 15 minutos y observar su conducta, luego se registra el tiempo de latencia (tiempo de nado del animal desde que se coloca en la superficie, hasta que queda inmóvil) entre menos tiempo dure el nado de los animales, más deprimido se considera el animal.

El test de consumo de sacarosa consiste en ponerle al roedor 2 botellas para que beba, una llena de agua, y la otra de solución sacarosa, intercambiando de puesto las botellas, y luego se pesan para conocer las preferencias del roedor, la preferencia de agua, en lugar de la solución sacarosa, indica características depresivas.

Estos test se les aplicaron semanalmente a los roedores para evaluar la actividad del posible efecto antidepresivo.

**Tercero: Selección de la Variable Respuesta.**

Las respuestas que servirán para el análisis de los resultados son el peso de las botellas (con agua, con solución sacarosa), y el peso de los roedores.

**Cuarto: Elección del Diseño Experimental.**

Se utiliza un diseño de experimentos por factores.

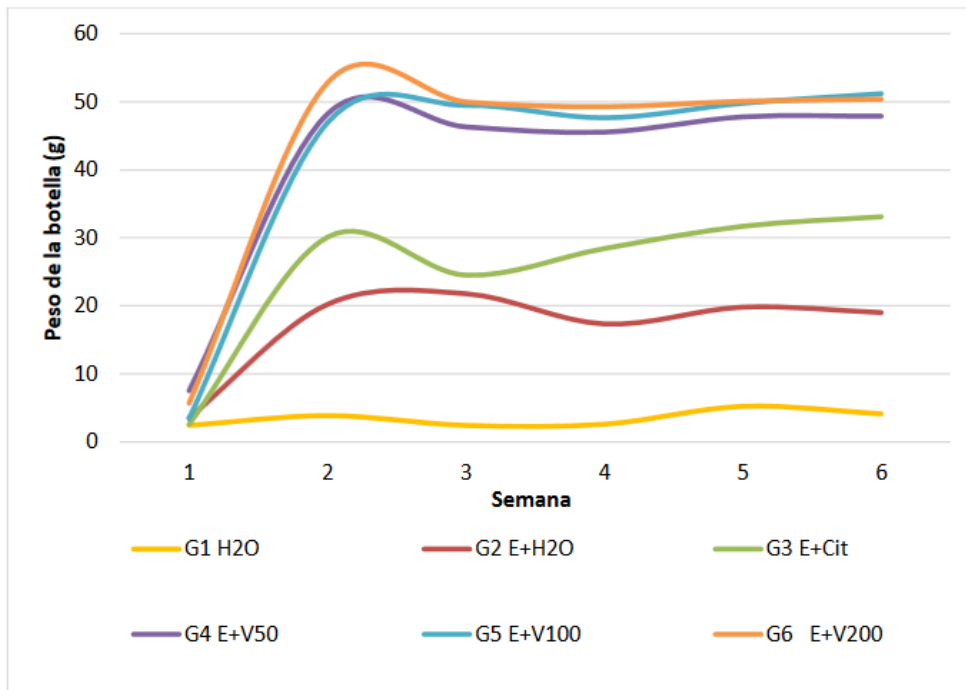
**Quinto: Ejecución del Experimento.**

Se seleccionaron 30 ratones de ambos sexos, con un peso corporal entre 20 y 25 g, con aproximadamente 5 semanas de nacidos, y se dividieron en 6 grupos de 5, el grupo 1 es tratado con agua destilada sin sometimiento a estrés, el grupo 2, tratado con agua destilada y sometido a estrés, el grupo 3 fue sometido a estrés con Citrapolam (antidepresivo), 4, 5 y 6 fueron sometidos a estrés más las diferentes dosis de veneno de R. marina.

Se tomaron los datos iniciales de consumo de sacarosa para cada uno de los animales de experimentación, y la duración del nado forzado, y se tomaron de nuevo al finalizar el experimento, así como el peso de los roedores al finalizar las 5 semanas. Después de la primera semana se les aplicó los 3 estresores de depresión, hasta finalizar las 5 semanas.

## Sexto: Análisis de las observaciones obtenidas.

**Gráfico 5:** peso de las botellas con agua y solución sacarosa por cada semana.



**Descripción del gráfico:** Consumo semanal de sacarosa de los grupos G1 H2O= grupo blanco sin estrés (más agua); G2 E+H2O= grupo control negativo (estrés más agua); G3 E+Cit= grupo control positivo (estrés más Citalopram); G4 E+V50= grupo tratamiento estrés más dosis baja (50 mg/kg) de veneno; G5 E+100= grupo tratamiento estrés más dosis media (100 mg/kg) de veneno; G6 E+200= grupo tratamiento estrés más dosis alta (200 mg/kg) de veneno, en la evaluación de actividad biológica del veneno de R. marina en ratones hembra.

Se pudo observar que el grupo de control que no fue sometido a factores de estrés leve crónico, consumieron la mayor cantidad de solución sacarosa (se puede encontrar menor peso en las botellas), se puede decir que los animales de este grupo se encontraron sin indicios de conducta depresiva, puesto que a mayor consumo de sacarosa, menores características depresivas, en cambio, en los demás grupos después de unas semanas comenzaron a perder el interés por la solución azucarada, con lo que se lograron individuos con características depresivas.

**Tabla 9:** Promedio y relación entre pesos corporales (g) de ratones hembra al inicio y final de la prueba.

<b>Grupo</b>	<b>Peso Inicial</b>	<b>Peso Final</b>	<b>Aumento %</b>	<b>p valor</b>
<b>G1 H2O</b>	18.5	25.26	36.54	0.000040*
<b>G2 E+H2O</b>	19.28	21.04	9.13	0.027888*
<b>G3 E+Cit</b>	19.3	16.72	-13.37	0.566518
<b>G4 E+V50</b>	20	21.5	7.50	0.058347
<b>G5 E+V100</b>	19.28	20.5	6.33	0.230691
<b>G6 E+V200</b>	19.6	21.78	11.12	0.055156

Los valores se expresan, peso inicial y peso final en aumento porcentual y \*p valor ( $p < 0.05$ )

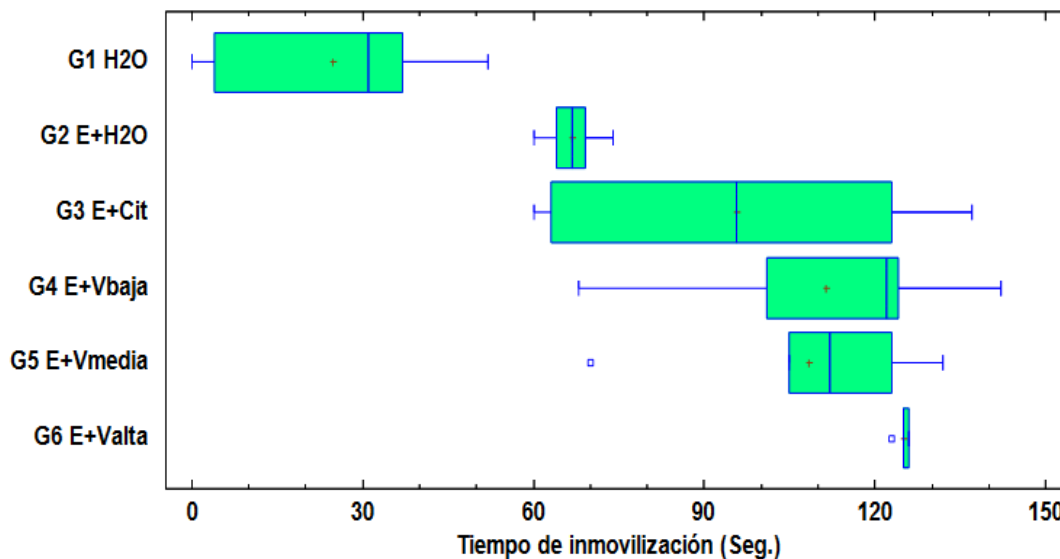
En el comportamiento de los pesos corporales de los ratones, se pudo observar que en los grupos donde los sometieron al factores estresantes (2-6) los animales no lograron alcanzar su peso óptimo, en el grupo 1 se puede observar que logro aumentar de peso normalmente, y los grupos en los cuales se aplicó el veneno no muestran diferencia significativa con respecto al grupo 2 en el cual solo se aplicó agua más factores estresantes, el grupo 3 en el que se aplicó el citalopram es el que muestra una pérdida de peso en los ratones.

**Tabla 10:** Tiempos promedios (seg.) de los grupos de ratones hembra sometidos a evaluación de efecto antidepresivo del veneno de R. marina.

<b>Grupos</b>	<b>Media ± D.S</b>		
<b>G1 H2O</b>	31	±	22.22
<b>G2 E+H2O</b>	66.75	±	5.26
<b>G3 E+Cit</b>	95.75	±	34.62
<b>G4 E+V50</b>	122	±	28.28
<b>G5 E+V100</b>	112	±	23.82
<b>G6 E+V200</b>	125	±	1.22

Los valores se expresan como la media ± la desviación estándar

**Gráfico 6:** Tiempo de inmovilización ratones hembra (segundos) al finalizar las pruebas de estrés y Test de Nado Forzado.



**Descripción del gráfico:** por grupo: G1 H2O= grupo control blanco (agua sin estrés); G2 E+ H2O= grupo control negativo (estrés más agua); G3 E+Cit= grupo control tratamiento (Citalopram mas estrés); G4 E+V50= grupo tratamiento (estrés más 50mg/kg de veneno) G5 E+V100= grupo tratamiento (estrés más 100mg/kg de veneno); G6 E+V200= grupo tratamiento (estrés más 200mg/kg de veneno) en ratones hembras.

La tabla y la prueba de Dunnett para comparaciones múltiples presenta el tiempo de inmovilidad en el test de nado forzado, se puede observar que tanto en los grupos de control donde se aplicó el veneno y en donde se aplicó el fármaco citalopram, presentan tiempos de letargo, de los que se pudo concluir que los roedores mostraban síntomas de depresión, aún mayores que en el grupo donde se aplicó agua y factores estresantes.

**Séptimo: Conclusiones y Recomendaciones.**

**Conclusión planteada por los investigadores:**

El modelo para inducción a depresión fue un éxito, ya que, al evaluarlos con el Test de Consumo de Sacarosa, mostrando una inhibición en la conducta, con características depresivas en los individuos sometidos al protocolo de Estrés Leve Crónico. Los animales



mostraron un comportamiento irregular en relación al peso corporal durante la prueba con el veneno de esta especie, debido a alteraciones provenientes del estrés crónico aplicado, lo que indica que los individuos poseían un grado depresivo, que era el necesario para el desarrollo de la evaluación del veneno como antidepresivo.

Por otro lado, la determinación de la dosis del fármaco (Citalopram) bajo el modelo presentado, resulto adecuado, pero luego mostró un comportamiento no esperado en el ensayo principal, por esta razón se concluye que el Citalopram posee una limitada acción como tal, dado a que pudo observarse que no se comportó como un verdadero antidepresivo en los animales evaluados. Sin embargo, se retoma lo ocurrido en el Test de Consumo de Sacarosa, donde si actuó como se esperaba lo cual abre un debate con relación a este hecho.

En conclusión, el veneno de R. marina no posee efecto antidepresivo, dado a que en la evaluación de la actividad antidepresiva del veneno con el Test de Consumo de Sacarosa los animales mostraron pérdida del interés de la sustancia que le produce placer es decir el veneno, no logró reducir la anhedonia en los animales. Mientras que en el Test de Nado Forzado los ratones perdieron la voluntad de seguir en movimiento mientras nadaban en el recipiente, lo que se traduce como la pérdida en la voluntad de vivir.

### **Conclusión propuesta por las creadoras del manual:**

En esta investigación se realizó una prueba previa con una cantidad de ratones en distintas circunstancias y de distintos sexos; para conocer cuál era, el nivel de toxicidad del veneno R. marina; y así poder realizar la investigación con la cantidad de dosis necesaria para realizar dicho experimento y que los ratones no murieran en el proceso.

Se realizó la separación de los ratones en seis diferentes grupos, Aplicándoles previamente estrés leve; se pudo dar a conocer, que reacción tenía la aplicación del veneno y del fármaco. Las pruebas realizadas fueron de mucha utilidad para dar a conocer si el veneno R. marina era de utilidad o no, para formar parte del medicamento antidepresivo.

Se concluye, después de realizar las diferentes pruebas las cuales fueron el consumo de sacarosa y el test de nadado forzado, a los grupos que se les aplico el veneno R.

marina, que no se considera un antidepresivo ya que en la prueba del consumo de sacarosa no mostro ningún interés por realizar dicho consumo de la sustancia y en la prueba del test de nadado forzado se mantenían inmóviles y eso se puede describir como: perdida en la voluntad de vivir. Es por eso que este veneno no se puede considerar como un antidepresivo.

### **Recomendaciones planteadas por los investigadores para los laboratorios:**

Como recomendación se deben realizar la mayor cantidad de repeticiones en los ensayos, utilizando animales de ambos sexos y aumentando el número en cada grupo dentro del modelo experimental que se elabore, esto debido a que las pruebas conductuales son más difíciles de evaluar en la experimentación animal, se recomienda aislar lo más posible el compuesto o principio activo que se desea evaluar.

En algunas investigaciones, es importante evaluar la química completa de la sustancia. Sería importante la creación de un extracto del veneno como tratamiento previo de la sustancia.

Deben seguirse realizando este tipo de investigaciones con énfasis en problemáticas de salud que impactan a una sociedad, como en este caso la depresión, es decir, que se continúe investigando sustancias de origen natural que aporten a la medicina en general y controlar de mejor manera enfermedades de cualquier índole, pero en este caso, más a padecimientos de la salud mental.

Para la utilización de fármacos de referencia ya existentes en el mercado es importante que se seleccione el más idóneo, según las características descritas para cada uno, para evitar una reacción no esperada del medicamento que se utilice.

Deben realizarse la mayor cantidad de repeticiones en los ensayos, utilizando animales de ambos sexos y aumentando el número en cada grupo dentro del modelo experimental que se elabore, esto debido a que las pruebas conductuales son más difíciles de evaluar en la experimentación animal, teniendo en cuenta el principio de las 3Rs.

### **Recomendaciones propuestas por las creadoras del manual para los laboratorios:**

Se recomienda que cuando se realicen estas investigaciones se pueda tomar en cuenta, si en caso la persona pareciera alguna otra enfermedad congénita la cual pudiera alterar la investigación de este veneno o de algún fármaco a probar.

La repetición de esta investigación sería de mucha importancia, ya que el ambiente cambia día con día y sería de mucho interés la comparación de esta investigación con una actual

Al realizar nuevamente la investigación sería de mucha utilidad que utilizaran los sapos de la misma posición geográfica, para poder realizar una investigación puntual.

En el aspecto estadístico para realizar comparaciones sería recomendable utilizar otras pruebas con los datos recolectados.

En una posterior investigación se pueden comparar otros fármacos antidepresivos para poder dar a conocer la reacción secundaria que estos presentan.

## **Ejemplo 2:** “Estudio sobre la roya en El Salvador período 2016-2017”

### **Introducción:**

El cultivo del café en El Salvador, se desarrolla en tres estratos altitudinales, con diferentes características climáticas en cada uno de ellos. En este contexto se presenta la roya del cafeto, enfermedad causada por el hongo, *Hemileia vastatrix*, la cual está causando serios daños a la caficultura salvadoreña; debido a que su efecto repercute directamente en la formación del grano y consecuentemente, reduce severamente la producción; influyendo directamente en los ingresos y rentabilidad del caficultor.

Además, el cultivo del café, desde el punto de vista social, representa generación de empleo e ingresos, para una gran cantidad de familias del campo y de la ciudad. Así mismo es fuente de divisas para nuestro país, sin quitar méritos al consumo nacional del café, así como su importancia ambiental.

La roya en el cafeto es una enfermedad que fue detectada por primera vez, a finales de 1979, en la finca California, jurisdicción de Santiago de María Usulután, siendo una enfermedad cuarentenaria, policíclica y de carácter bianual, conociéndose a su vez, sus efectos; tiene como ya se describió anteriormente, una relación con los niveles de producción, motivo por el cual es de vital importancia estudiarla, para un mejor conocimiento de su comportamiento, en las condiciones ambientales de nuestro país, y así poder aplicar los resultados obtenidos, en el manejo integrado del cultivo, la roya del café (*Hemileia vastatrix*).

Tanto en las épocas secas y lluviosas, y su relación con los factores ambientales de mayor incidencia en los tres estratos altitudinales de la zona cafetalera del país, durante el ciclo productivo, comprendido de mayo de 2011 hasta abril de 2012. Durante este periodo se determinó, que la enfermedad, no tiene el mismo comportamiento, en cada estrato altitudinal; sino que responde a un factor en específico. Encontrándose que al menos un factor influyó en el desarrollo de la enfermedad (incidencia y severidad); sea éste, la precipitación, la temperatura y humedad relativa; según la época del año.

reflejándose, en los resultados de la prueba estadística y la curva de progreso de la enfermedad.

Las epidemias de roya, están sujetas a variables medio ambientales, tales como: la temperatura, el viento, la precipitación de agua; la radiación solar los tiempos de sequía y tiempos de lluvia, e incluso a circunstancias indirectas de cambios en las variables ambientales, como el apareamiento de otras enfermedades, insectos y microorganismos en un área específica y que podrían ejercer influencia en las epidemias de roya, ya sea positiva o negativamente para la agricultura.

Los efectos del clima en el comportamiento de la vida en general, ejercen mayor grado de influencia. En el nacimiento, crecimiento y reproducción de los seres vivos se dan una serie de requerimientos específicos en cuanto al ambiente, los cuales si no se cumplen la vida no se presenta, se extingue o en otros casos sufre procesos en los cuales se acomoda al medio ambiente y a los cambios de éste por medio de un fenómeno denominado “adaptación ambiental”. Este caso no es ajeno a la roya del café, hongo fitopatógeno, cuyos requerimientos climáticos varían según la cepa de roya que se esté tratando.

## **Metodología**

### **Método de campo**

Se utilizó la metodología propuesta por Avelino (1991). Se seleccionaron tres fincas, una en cada estrato altitudinal. En cada finca se delimito al centro de la de la parcela, tres surcos de cuatro cafetos cada uno, para un total de 12. Los cafetos se numeraron del uno al doce y se muestrearon una vez por mes durante 12 meses.

### **Forma de muestreo de Roya**

Cada uno de los doce cafetos se estratifico en tres niveles altitudinales: parte alta, media y baja. En cada estrato se tomaron, cuatro bandolas, y en cada una se hizo un conteo cada 30 días de: a) hojas viejas, b) hojas nuevas, c) cantidad de hojas enfermas por árbol (incidencia) por roya, d) número de pústulas de roya por árbol (severidad), en el caso del conteo de hojas nuevas, se tomaron en cuenta las que surgieron hasta el mes de noviembre ya que las siguientes corresponden al siguiente ciclo de producción. Las visitas para toma

de datos, se realizaron mensualmente en la finca en las fechas comprendidas, entre los días 10 y 15 de cada mes.

Además, se hizo un conteo de frutos, y nudos fructíferos en el mes junio para medir inicialmente la producción de frutos y en el mes de noviembre se verifico, cual fue la cantidad de frutos de la cosecha.

### **Característica de las fincas**

#### **Finca Normandía (Bajío)**

Está ubicada en Cantón Santa Rosa, jurisdicción del municipio de San Juan Opico, departamento de La Libertad, a una altura de 538 msnm y coordenadas 240956 y 1533552 UTM (Sistema de Coordenadas Universal Transversal de Mercator, en metros).

Los promedios anuales de temperatura de 24°C, humedad relativa de 76% y precipitación de 1688 mm. La parcela seleccionada, era de topografía semiplana, con un suelo de textura franco arenoso, de color pardo café, con buena cobertura de materia orgánica. Los niveles de sombra del cafetal fueron de un 50%, los arboles de sombra predominantes eran del genero Inga y leguminosas papilionáceas. La variedad sembrada, de café es Pacas en un 100%, y la edad de la plantación, aproximadamente 30 años. El manejo de la finca, se enfocaba principalmente, al control de malezas dos veces al año, raleo de sombra y fertilización química foliar.

#### **La Finca San Antonio (Media Altura)**

Está ubicada en Las Aradas, jurisdicción del municipio y departamento de Santa Ana, a una altura de 831 msnm y coordenadas 0225922 Y 1541502 UTM (Sistema de Coordenadas Universal Transversal de Mercator, en metros). Con promedios anuales de temperatura de 22°C, humedad relativa de 71% y precipitación de 1780mm.

La topografía es semiplana; el suelo es de textura franco limoso, de color café oscuro, con moderado contenido de materia orgánica. Los niveles de sombra fueron del 60%, y las especies de árboles de sombra son de los géneros Manguífera, Inga y Cedrela. La variedad de café era Pacas en un 100%, la edad promedio de la plantación es de 20 años.

El manejo de la finca, se orienta principalmente al control de malezas dos veces al año; manejo de los niveles de sombra, fertilización química foliar, control de la broca del café y aplicación de cal dolomita, después de la cosecha del grano.

**c) Finca El Olvido (Estricta Altura).**

Está ubicada en el cantón El Boquerón, jurisdicción del municipio de Santa Tecla, del departamento de La Libertad; a una altura de 1554 msnm y coordenadas 0254968 y 1520396 UTM (Sistema de Coordenadas Universal Transversal de Mercator, en metros).

Con promedios anuales de temperatura de 20°C, humedad relativa del 79% y precipitación de 1860mm. La topografía es moderadamente inclinada a muy inclinada; el suelo es de textura franco arenosa, de color café a negro, con moderado contenido de materia orgánica. El nivel de sombra fue de 25%, las especies de árboles de sombra son del género, Inga. La variedad de café es Bourbon en un 100%, la edad promedio de los árboles de café promedia los 35 años de siembra del cafetal. El manejo de la finca se encamina principalmente al control de malezas dos veces al año, fertilizaciones al suelo y foliar, aplicación de cal dolomita después de la cosecha del grano y poda de árboles de café de forma rotativa cada dos o tres años.

**Métodos estadísticos:**

Se utilizó el método de estadística descriptiva y el modelo de análisis de regresión lineal simple, se relacionó la influencia de los factores climáticos; precipitación, humedad relativa y temperatura, con los niveles de incidencia y severidad, relativos de las hojas enfermas y pústulas por árbol, obtenidos de cada finca. Además, se determinó la relación entre, incidencia, severidad y caída de frutos.

Los datos mensuales obtenidos, tanto por los muestreos, como los datos de precipitación, temperatura y humedad relativa respectivamente; se promediaron en una matriz de 12 datos, luego se procesaron y analizaron por medio del modelo de regresión lineal simple ( $y = a \pm bx$ ) en dónde se evaluó cada caso, los niveles de incidencia y severidad, versus las variables meteorológicas de; temperatura, humedad relativa y precipitación respectivamente; relacionándolas con 60-90 días posteriores a los niveles de

incidencia y severidad encontrados. Los parámetros epifitológicos como tasa de infección aparente ( $r$ ) y la curva de progreso de la enfermedad ( $cpe$ ), se trabajaron por medio de un modelo logarítmico de desarrollo de enfermedad (modelo logístico:  $y = \ln [x / (1-x)]$ ) y por el modelo de área bajo la curva de progreso de la enfermedad ( $abcpe = \Sigma(y_i + y_{i+1})/2 (t_{i+1} - t_i)$ ), lo que permitió, conocer las unidades de enfermedad alcanzadas por el cultivo, durante el tiempo de estudio.

Esto permitió conocer las fases de desarrollo de la enfermedad, en cada finca y poder generar estrategias de manejo, de acuerdo a las curvas de desarrollo y las variables climáticas. Además del conteo de frutos se hizo un análisis estadístico descriptivo paramétrico y el coeficiente de determinación ( $r^2$ ), donde se precisó la caída o pérdida de frutos de café.

## Resultados

### Finca Normandía (Bajío)

#### Niveles de incidencia y severidad

**Tabla 11:** Nivel de incidencia de la enfermedad en hojas viejas y nuevas en los meses de mayo de 2016 hasta abril de 2017 en finca de bajío.

Ubicación de bandola en el árbol	Bandola																	
	Cantidad de hojas enfermas por árbol																	
	Hojas viejas						Hojas nuevas											
	May	Jun	Jul	Ago	Sep	Oct	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Ene	Feb	Mar	Abr
Alta	1.0	0.2	0.3	0	0	0	1.3	0.2	0.3	0	0.3	1.2	3.6	10.7	41.5	18.3	16.6	9.4
Media	0.6	0.3	0.6	0.6	0	0	0.9	0.4	0.7	0.7	0.4	1.5	4.3	15.4	53.3	26.5	21.4	12.4
Baja	0.6	0.9	0.7	0	0.1	0	0.8	1	0.8	0	0.3	0.9	5.1	14.7	45.4	22.4	20.3	16.1
Total	2.2	1.4	1.6	0.6	0.1	0	3.0	1.6	1.8	0.7	1.0	3.6	13.0	40.8	140.2	67.2	58.3	37.9

Los niveles de incidencia en hojas viejas y nuevas; determinándose que, en el mes de mayo de 2016, la presencia de roya en hojas viejas estaba con un promedio, de 2.20 hojas enfermas, decreciendo en los meses subsiguientes, hasta que desapareció, prácticamente en septiembre del mismo año, como resultado de la caída de las hojas. Con esto se demuestra, que el inóculo inicial de la enfermedad se encuentra en las hojas viejas; de las cuales se dispersa el inóculo, infestando las hojas nuevas.



Estos resultados coinciden con lo que describen, Avelino (1991) y Rivillas Osorio, et al (2011), que mencionan que el inoculo inicial, se encuentra en las hojas viejas del año anterior, las cuales infestan e infectan a las nuevas.

En el caso de las hojas nuevas, se encontró en el mes de mayo de 2016, un promedio de 3.00 hojas enfermas y se mantuvo en un nivel bajo, hasta el mes de septiembre del mismo año; pero comenzó a crecer en octubre del mismo año y alcanzo un máximo de 140.20 hojas enfermas, para el mes de enero del año 2017 y a partir de ahí, comenzó a descender.

Datos similares a lo que describen, MAG e ISIC (sf), aunque con un desfase de un mes. Avelino, et al (1999), describen este fenómeno, donde explican que la fase de desarrollo lento, se presenta de mayo a julio, en seguida la de fase de crecimiento acelerado, hasta el mes de noviembre y la cúspide, en el mes de diciembre del mismo año, y posteriormente, la fase de descenso.

En cuanto a la ubicación de las bandolas en el árbol, el comportamiento de la incidencia, fue similar en los niveles; baja, media y alta de los cafetos, contrario a lo que afirmaban estos autores, anteriormente mencionados, que la incidencia y severidad de la enfermedad, se presentaba en mayor proporción, de la parte baja hacia la superior.

**Tabla 12:** Nivel de la severidad de la enfermedad en hojas viejas y nuevas durante los meses de mayo de 2011 a abril de 2012 en la finca de Bajío.

Ubicación de bandola en el árbol	Bandola																	
	Cantidad de pústula por árbol																	
	Hojas viejas						Hojas nuevas											
	May	Jun	Jul	Ago	Sep	Oct	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Ene	Feb	Mar	Abr
Alta	1.3	0.2	0.3	0	0	0	1.5	0.2	0.3	0	0.4	5.3	10.6	51.5	262	120	53.8	24.9
Media	0.8	0.3	0.7	0.7	0	0	1.1	0.4	0.7	1.5	0.6	5	14.3	47.8	310	177.7	67.8	51.2
Baja	0.8	1.1	0.8	0	0.1	0	0.9	1.1	0.8	0	0.4	1.2	13.1	42.8	274	121.8	63.6	46.3
Total	2.9	1.6	1.8	0.7	0.1	0	3.5	1.7	1.8	1.5	1.4	11.5	38.0	142.1	846	419.5	185.2	122.4

La severidad en las hojas viejas y nuevas, muestra un promedio de 2.90 pústulas en hojas viejas por árbol, y descendiendo paulatinamente hasta desaparecer en el mes de octubre; como producto de la caída de las hojas. Es de mencionar que, en este tiempo, es suficiente para que se infesten e infecten las hojas nuevas de los árboles. En el caso de las hojas nuevas, se encontró en el mes de mayo de 2016 un promedio de 3.5 pústulas por árbol y se mantuvo bajo hasta el mes de septiembre del mismo año, sin embargo, a partir

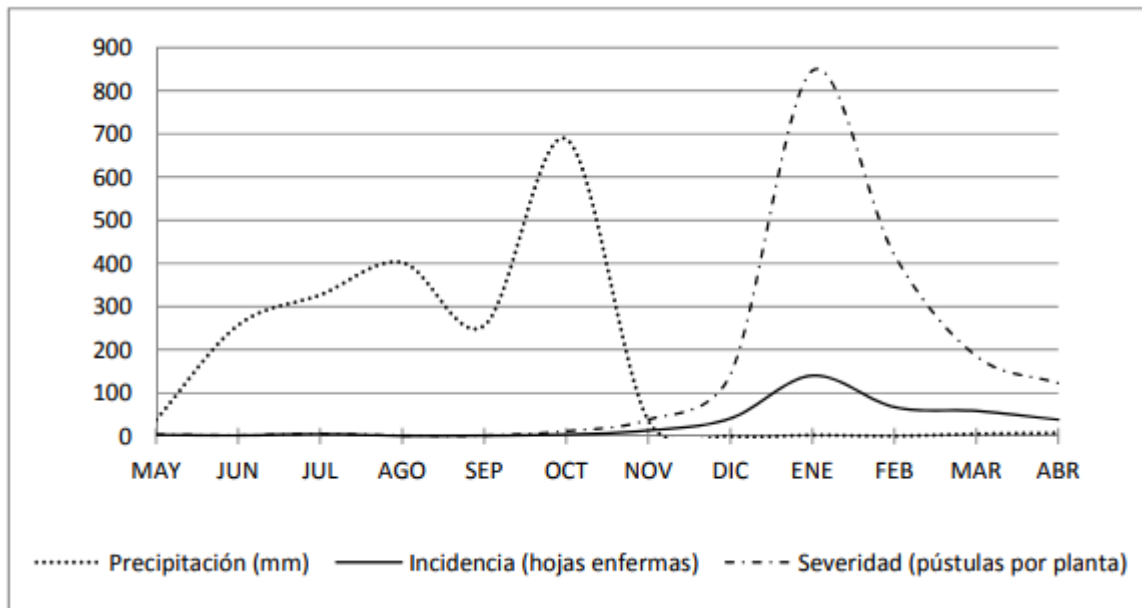
de octubre y noviembre mostró, un crecimiento acelerado, encontrándose un promedio de 38 pústulas en noviembre, alcanzando un máximo de 846 pústulas en enero del 2017, y comenzó a descender hasta que terminó con 122.4 en abril; situación que concuerda con los resultados de Avelino et al (1999).

Además, se muestra un alto valor de inoculo residual, para el siguiente ciclo de producción; evidenciado que la epifítia puede tener mayor repercusión. En relación al comportamiento de la severidad, en la ubicación de las bandolas en el árbol, fue similar en los tres estratos del árbol, contrario a que se decía que se presentaba en proporciones significativamente mayores de las secciones bajas y que disminuía gradualmente en las partes más altas del árbol, según Avelino y Muller (1991).

A continuación, se describe cada uno de los análisis de regresión lineal simple en cada finca, según nivel altitudinal de cada finca; relacionando, los niveles de incidencia y severidad, con los factores climáticos de: precipitación, humedad relativa y temperatura. Determinándose si existe regresión, y si el factor influye directamente o no, con los niveles de la enfermedad; según la variación de cada factor y coeficiente de determinación, para medir el porcentaje de influencia, en la evolución de la incidencia y severidad de la enfermedad en el periodo comprendido de mayo de 2016 hasta abril de 2017, en bajío, media y estricta altura.

### Relación incidencia, severidad con la precipitación.

**Gráfico 7:** Curvas de incidencia y severidad de roya y su relación con la precipitación en la finca Normandía (bajío).



Fuente: creación propia

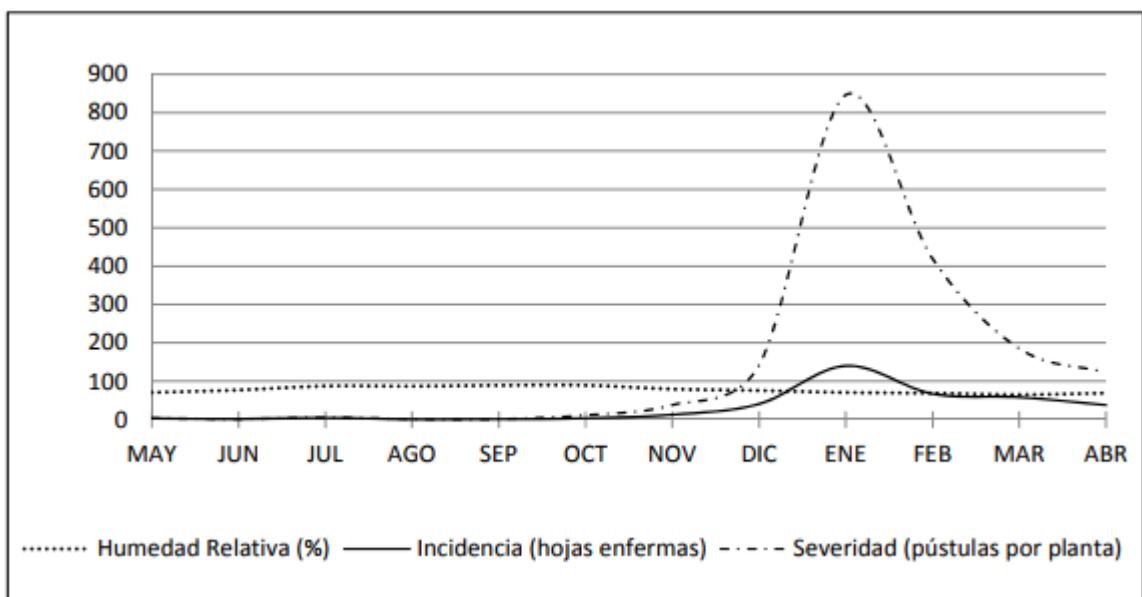
Los resultados de la variable precipitación y sus niveles, no presentaron diferencias estadísticas significativas, en los niveles de incidencia de roya, mientras que en la severidad de la enfermedad mostró mayor valor; lo que evidenció que la precipitación influye en mayor medida, en la dispersión de esporas; lo que genera nuevas pústulas, derivadas de las ya existentes. Este resultado coincide con Avelino et al (1999), quien describe, que la lluvia incide en la dispersión de esporas; fenómeno que aumenta el número de pústulas en una misma unidad foliar y hojas adyacentes.

Por otra parte, el análisis de correlación, mostró que la precipitación influyó en 28% en la incidencia y en 34% en la severidad de la enfermedad; lo que demuestra que la precipitación influyó considerablemente, en la epifitología de la roya del café; sin embargo es de recalcar, que existen otros factores, que interactuaron con el factor lluvia, en el desarrollo de la enfermedad; situación similar a lo que afirma Rayner (1972), que al menos un factor ambiental influye en el desarrollo de la enfermedad.

Es de hacer notar que la proporción restante, puede ser atribuida también a la interacción, con la humedad relativa, temperatura y otros factores como el manejo y hospedero.

**Relación incidencia y severidad con la humedad relativa.**

**Gráfico 8:** Curvas de incidencia y severidad de roya y su relación con la humedad relativa en la finca Normandía (bajío).

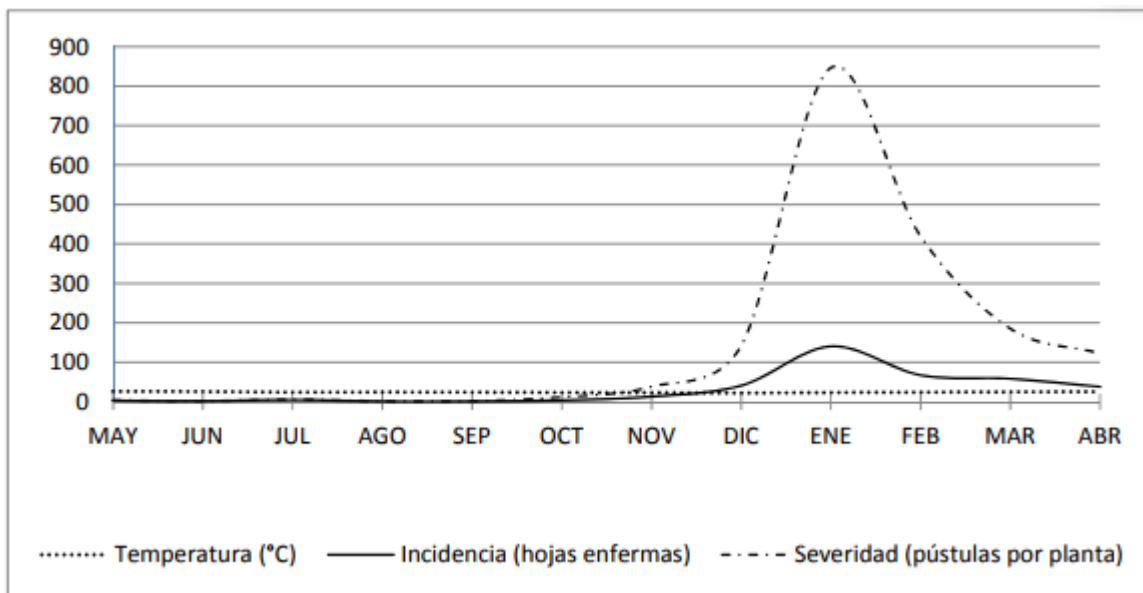


Fuente: creación propia

La relación de la humedad relativa y comportamiento epifitiológico de la roya del cafeto, en las curvas de incidencia y severidad, en el periodo de mayo de 2016 hasta abril de 2017, no presento diferencias significativas en la incidencia y la severidad; sin embargo, en el análisis de regresión, utilizando el coeficiente de determinación ( $r^2$ ) se encontró, que la humedad relativa influyó en un 24%, en la incidencia y en un 23%, en la severidad de la enfermedad. Además, como se muestra en la grafico 8, en el periodo de julio a octubre del año 2016, se presentaron cifras 87% a 88% de humedad relativa, cifra que favorece la germinación de esporas.

### Relación incidencia y severidad con la temperatura.

**Gráfico 9:** Curvas de incidencia y severidad de roya y su relación con la temperatura en la finca Normandía (bajío).



Fuente: creación propia

Los resultados de relación y comportamiento epifitológico de la roya del cafeto, en la incidencia y severidad, con respecto a la temperatura registrada, en el periodo de mayo de 2016 a abril de 2017, reflejan que la temperatura, estadísticamente influyo significativamente en la incidencia de la enfermedad, no así en la severidad.

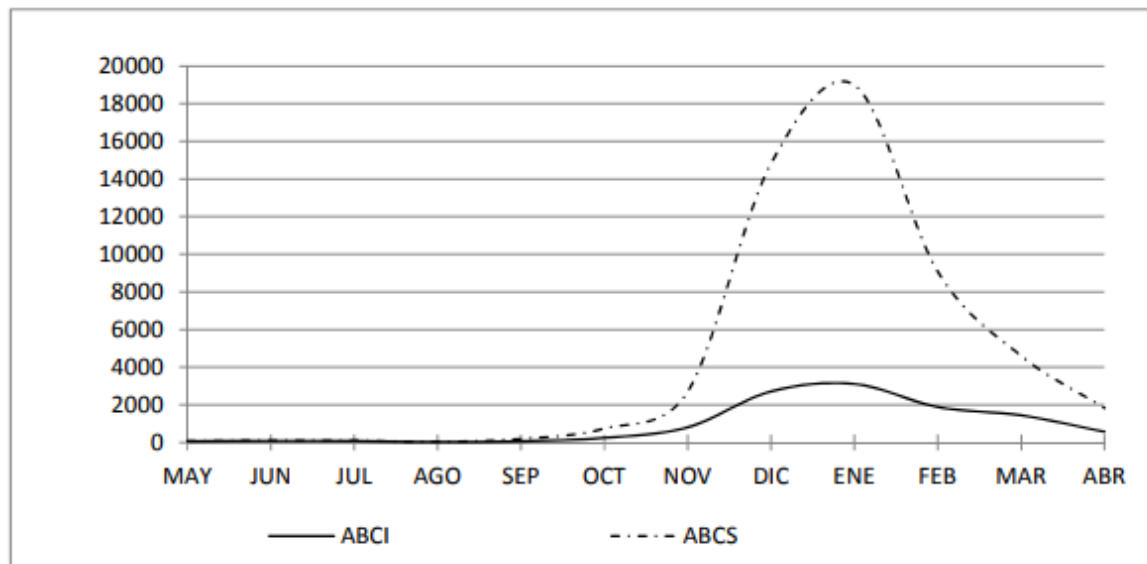
Sin embargo en el análisis de regresión, utilizando el coeficiente de determinación ( $r^2$ ) muestra que la temperatura, influyo en un 42%, en la incidencia y en un 29%, en la severidad; lo que demuestra que la temperatura, además de otros factores, influye proporcionalmente en desarrollo de la enfermedad. Además, durante la fase de crecimiento lento de la enfermedad, la temperatura rondó los 25°C, cifra que favoreció la germinación de esporas del hongo.

### Relación incidencia con la severidad.

En este análisis se demuestra estadísticamente, que la severidad de la roya del café (*Hemileia vastatrix*), dependió significativamente de la incidencia de la misma, como se demuestra en el análisis de la precipitación, que ésta, influye en mayor medida en la severidad 34%, provocada por la dispersión por salpicadura producida por la lluvia. En términos numéricos, se muestra que la severidad, dependió en un 95% de la incidencia, ya que esta última, es fuente de inoculo que infesta e infecta áreas y hojas adyacentes; estos resultados son similares a los descritos por Avelino (1991) y Rivillas Osorio, et al (2011).

### Tasa de infección aparente (r)

**Gráfico 10:** Área Bajo la Curva de Progreso de la Incidencia y Severidad (ABCPI Y ABCPS) de la roya del café en la finca Normandía (bajío).



Fuente: creación propia

La velocidad de infección promedio, utilizando la tasa de infección aparente, tuvo un crecimiento proporcional al inoculo inicial en la incidencia de 4.63, y severidad de 4.68, de la epidemia de la roya del café. En la severidad, es levemente superior la tasa de infección de la enfermedad, debido a los factores que favorecen más a la severidad dado;

el caso que ésta última depende de la incidencia, y ésta procede como fuente de inoculo inicial, para que la lluvia disperse las esporas, y estas unidades infectivas, aumentan el número de pústulas por unidad foliar; además también favorecida por la temperatura, en esta categoría de finca.

En el grafico 10, se muestra el comportamiento de la enfermedad: crecimiento lento y acelerado, la cúspide y el descenso de la incidencia y severidad, en las hojas infectadas, lo cual reveló que ambos parámetros coinciden proporcionalmente en las épocas de crecimiento, no así en la intensidad, para cada parámetro epifitológico.

Por lo tanto, la enfermedad mostró, la etapa lenta de mayo a octubre de 2016, mientras que la fase de desarrollo acelerado entre los meses de noviembre y diciembre del mismo año y la cúspide, se produjo en enero de 2017 y la fase de descenso posteriormente. Por otra parte, como se observa en el gráfico, el Área Bajo la Curva de Progreso de la Severidad (ABCPS), es muy superior a la de la incidencia.

La lluvia principalmente favorece más a la severidad, aun cuando, naturalmente existen otros factores como, el viento y el manejo agronómico que influyen en cierta medida también.

### **Perdida de frutos**

Según los resultados obtenidos, en el conteo de frutos; se obtuvo una leve disminución de éstos, previo a la cosecha. El conteo se realizó, cuando éstos estaban en formación (junio) y en su fase de madurez antes de la cosecha (noviembre).

En este caso se registró poca perdida, porque probablemente en esa fecha, la enfermedad su etapa de crecimiento acelerado y no produjo mayor efecto negativo. Sin embargo, utilizando el coeficiente de determinación, el resultado indica, que la disminución, si dependió directamente un 99%, de la enfermedad. Al respecto, Avelino y Muller (1991), indican que la enfermedad causa efectos negativos en el llenado y cuajado de los frutos, disminuyendo la producción.

## Finca San Antonio (media altura)

### Niveles de incidencia y severidad

**Tabla 13:** Nivel de incidencia de la enfermedad en hojas viejas y nuevas durante los meses de mayo de 2016 a abril de 2017 en la finca de media altura.

Ubicación de bandola en el árbol	Bandola																	
	Cantidad de hojas enfermas por árbol																	
	Hojas viejas						Hojas nuevas											
	May	Jun	Jul	Ago	Sep	Oct	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Ene	Feb	Mar	Abr
Alta	1.4	0.4	0.6	0	0.1	0	2.3	1.7	1.9	5	7.8	16.1	19.5	18.8	18.8	11	8.3	3.7
Media	0.8	0.7	0.4	0.6	0	0	2.6	1	2.4	7.3	14.2	26.4	31.6	26.4	18.7	16.1	9.3	5
Baja	0.9	0.6	0.9	0.1	0.1	0	1.6	1.7	3.4	9.8	19.9	27.5	34.6	26.8	14.2	15.8	12	3.8
Total	3.1	1.7	1.9	0.7	0.2	0	6.5	4.4	7.7	22.1	41.0	70.0	85.7	72.0	51.7	42.9	29.6	12.5

Se muestra el nivel de infección encontrada en hojas viejas y nuevas; en las hojas viejas, se encontró un nivel de 3.10 hojas infectadas en promedio y con decrecimiento consecutivo; desapareciendo en el mes de octubre de 2016, debido a la caída total de hojas. De esta manera se confirmó que el inóculo inicial se encuentra en las hojas viejas, tal como lo que descrito por Avelino y Muller (1991).

Sin embargo, en el caso de hojas nuevas, los niveles de incidencia promedio (mayo y meses posteriores) encontrados al inicio fueron superiores en relación a las hojas viejas; ya que en mayo fue de 6.5 hojas enfermas; manteniendo un crecimiento continuo y progresivo, a partir de julio, alcanzando un nivel máximo en noviembre, de 85.70 hojas enfermas; decreciendo a partir de diciembre.

En cuanto a los valores encontrados en cada estrato altitudinal del árbol, fueron levemente mayores en las partes baja y media del árbol. Situación similar a lo descrito por Avelino (1991), que la enfermedad inicialmente, se presenta en mayor intensidad en las partes bajas del árbol y asciende a las partes más altas posteriormente.



**Tabla 14:** Niveles de severidad de la enfermedad en hojas viejas y nuevas durante los meses de mayo de 2016 a abril de 2017 en la finca de media altura.

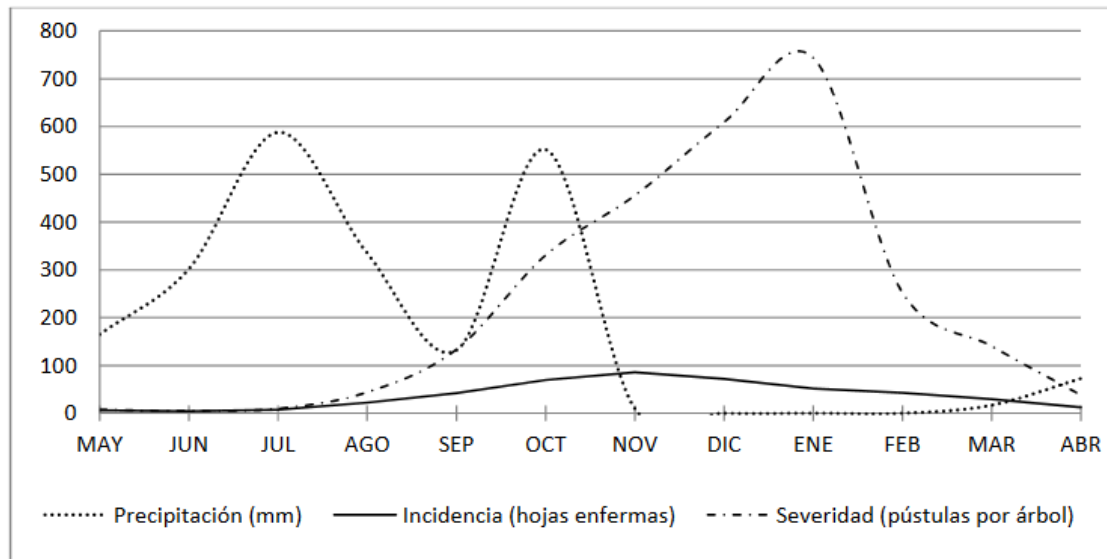
Ubicación de bandola en el árbol	Bandola																		
	Cantidad de pústula por árbol																		
	Hojas viejas						Hojas nuevas												
	May	Jun	Jul	Ago	Sep	Oct	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Ene	Feb	Mar	Abr	
Alta	2	0.3	0.5	0	0	0	2.9	1.9	3.5	13	20.8	71.3	88.7	153.2	255.5	61.3	33.3	8.9	
Media	1.2	0.5	1.2	1.1	0	0	2.9	1.1	2.9	14.2	47.7	141.2	164	227.8	287.3	99.9	41.1	12.5	
Baja	1.2	1.7	1.3	0	0.2	0	1.8	1.8	3.4	16.9	65.3	118.4	204	228.3	200.9	91.4	66.1	16.8	
Total	4.4	2.5	3.0	1.1	0.2	0	7.6	4.8	9.8	44.1	133.8	330.9	456.7	609.3	743.7	252.6	140.5	38.2	

El nivel promedio de severidad inicial fue de 4.40 pústulas por árbol, disminuyendo, hasta que las hojas viejas se cayeron por completo en el mes de octubre (tabla 12). Donde se confirma de igual manera que, la enfermedad se encuentra en hojas viejas, la cual disminuye paulatinamente, según se produce la defoliación o caída de éstas en mayor medida en agosto y septiembre. Resultado, coincide con lo que mencionado por Avelino y Muller (1991); pues el inoculo encontrado en las hojas viejas, infesta e infecta a las hojas nuevas.

En las hojas nuevas, se encontró una severidad promedio, de 7.60 pústulas por árbol, en el mes de mayo; y se observó un leve crecimiento hasta el mes de julio, continuando con un crecimiento acelerado, a partir de agosto de 2016, hasta alcanzar un máximo en enero de 2017, luego comenzó a decrecer rápidamente en febrero, debido a que las hojas más infectadas son las primeras en caer, el descenso fue menos acelerado hasta abril porque la tasa de infección y crecimiento de la enfermedad en este periodo fue menor, este resultado coincide en cierta medida con lo que relata Avelino y Muller (1991).

### Relación incidencia, severidad con la precipitación.

**Gráfico 11:** Curvas de incidencia y severidad de la enfermedad y su relación con la precipitación en la finca San Antonio (media altura).



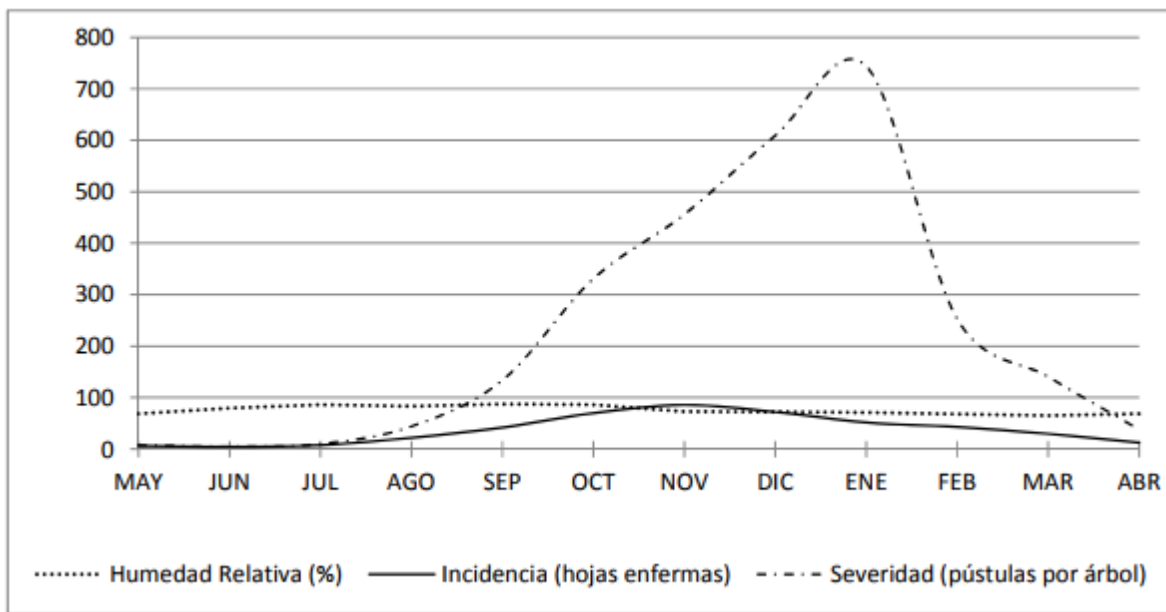
Fuente: creación propia

Al analizar los niveles de incidencia y severidad, se encontró que la precipitación, presento diferencias significativas en ambas variables. También la prueba de regresión, utilizando el coeficiente de determinación; muestra que la precipitación influyo en un 44%, en la incidencia y un 42%, en la severidad. Por lo que se demostró que la lluvia, influye en el desarrollo de la enfermedad.

Sin embargo, hay otros factores que interactúan con la precipitación, para que el fenómeno haya ocurrido de la manera en que se presentó. Este resultado, coincide con lo descrito por Avelino y Muller (1991), que expresan que la lluvia dispersa las esporas, con el salpique de las gotas sobre la superficie foliar y por escurrimiento hacia el envés de las hojas, infestando e infectando áreas y hojas vecinas a la pústula inicial.

### Relación incidencia y severidad con la humedad relativa.

**Gráfico 12:** Curvas de incidencia y severidad de roya del café y su relación con la humedad relativa en la finca San Antonio (media altura).



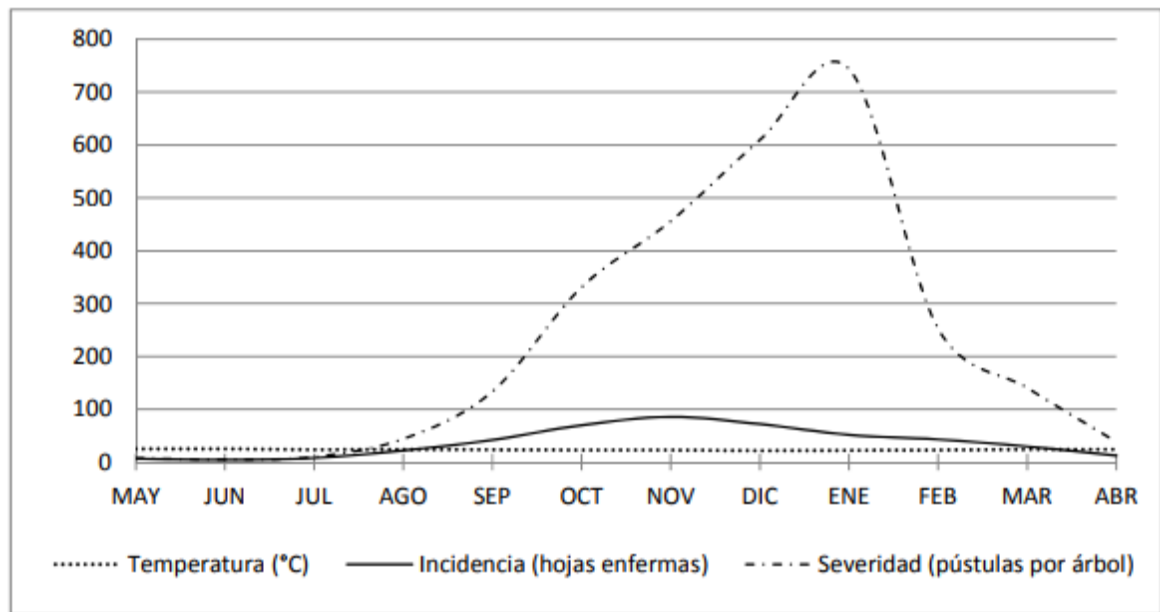
Fuente: creación propia

En la relación incidencia y severidad, con la humedad relativa, se encontraron diferencias altamente significativas, por lo que los niveles de humedad relativa encontrados principalmente en los meses de junio a octubre, se presentaron los mayores niveles promedio de 84.5%; fueron los que favorecieron, la propagación del hongo en las plantas.

De acuerdo a investigaciones realizadas por Moreno de Alas citado por MAG, e ISIC et al (sf), las esporas germinan en la fase de crecimiento lento de la enfermedad, con humedades relativas mayores al 80%. Los valores promedios de humedad relativa, descritos al inicio de este inciso, fueron mayores a los descritos por Moreno de Alas; lo que demuestra que estos niveles de humedad relativa presentados influyeron, en gran medida en el desarrollo epifitológico de la enfermedad. Además, con la prueba de regresión y tomando en cuenta el coeficiente de determinación, demuestra que la humedad relativa, influyo en un 82%, en la incidencia y 79%, en la severidad.

## Relación incidencia y severidad con temperatura

**Gráfico 13:** Curvas de incidencia y severidad de roya del café y su relación con la temperatura en la finca San Antonio (media altura).



Fuente: creación propia.

Esta relación no se encontraron diferencias estadísticas significativas, por lo que las variaciones de temperatura registrados, no influyeron en los niveles de incidencia y severidad de la epifitología, no así los niveles suscitados ya que los promedios que se registraron fueron óptimos, ya que estos fueron mayores a 22 °C, y Rivillas Osorio et al (2011), describe que con presencia de humedad en la hojas y temperaturas de entre 16 y 28°C, el hongo puede germinar, sin embargo en la prueba de regresión y verificando el coeficiente de determinación muestra que la influencia de la temperatura en la incidencia fue de 0.10% y en la severidad 4.0%.

El gráfico, muestra el comportamiento epifitológico de la roya del café en la incidencia y severidad; y su relación con la temperatura. Donde se registró un promedio de temperatura en los meses de mayo a septiembre mayor a 23°C y menor a 25°C (24.2 en promedio); temperaturas que favorecen, la germinación de esporas. De acuerdo a Rivillas Osorio et al (2011), las esporas germinan dentro de un rango de 16 a 28 °C. Pero

sin embargo el análisis de varianza y el coeficiente de determinación, indican, que no existe influencia directa, de este factor ambiental. Sin embargo, cada factor tiene alguna contribución en los procesos biológicos en etapas específicas.

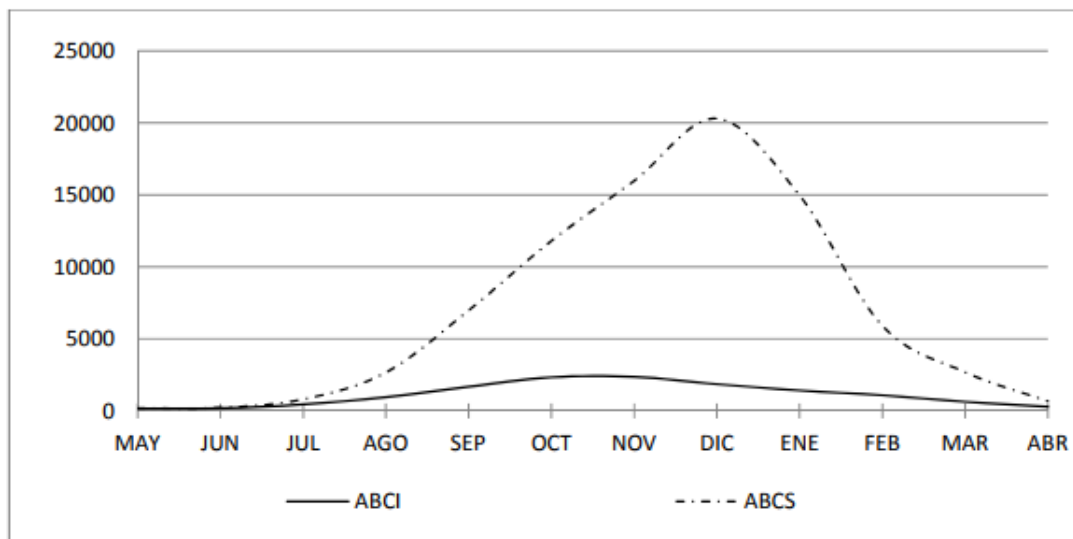
### **Relación incidencia con severidad.**

La incidencia la enfermedad, influyó en el aumento de la severidad de roya del café, producida en sus diferentes niveles de la enfermedad.

En el análisis de regresión y el coeficiente de determinación, reflejó que la severidad depende en un 66%, de los niveles de incidencia de la enfermedad presentada, donde una sola pústula de roya, se convierte en fuente de inóculo, y producto de la lluvia (salpique y escurrimiento), infeste e infecte áreas y hojas adyacentes. Es de recalcar que esta afirmación, tiene validez considerando que el inóculo encontrado, infeste e infecte a otras aéreas adyacentes, producto de otros factores, como la precipitación, viento, manejo agronómico, etc.; siempre y cuando las condiciones de germinación sean óptimas.

### **Tasa de infección aparente (r).**

**Gráfico 14:** Área Bajo la Curva de Progreso de la Incidencia y Severidad (ABCPI y ABCPS), en la finca San Antonio (media altura).



Fuente: creación propia

La tasa de infección promedio, utilizando la tasa de infección aparente, tiene un crecimiento desigual, en la incidencia 3.37 y severidad 5.10 de la enfermedad.

La severidad se dio en mayor grado, debido a los factores que influyeron en mayor medida; ya que ésta depende de la incidencia como fuente de inóculo, de la lluvia, porque el salpique dispersa las esporas; además, la humedad relativa que se registró en el periodo de crecimiento lento de la enfermedad, generó condiciones óptimas para la germinación; situación que aumentó el número de pústulas por unidad foliar infectada, en este estrato altitudinal, en el que se ubica la finca.

Los resultados obtenidos, son similares a los encontrados, en los estudios realizados por Avelino y Müller (1991) y de acuerdo a Rivillas Osorio (2011), la lluvia contribuye en el desarrollo de epifitología de roya y la infección se produce en hojas o áreas vecinas a la infestación inicial; por lo que la severidad tiende a crecer en mayor medida.

### **Perdida de frutos.**

En este caso se encontró un resultado significativo, que indicó, que la caída de frutos se debió en cierta medida, a los efectos que la enfermedad produce en las plantas; debido a que la incidencia y severidad, se desarrollan aceleradamente, mucho antes que la maduración de frutos; empezando en el mes de julio y finalizando en noviembre, reflejando un efecto más devastador. En el análisis estadístico, utilizando el coeficiente de determinación, se demuestra que la caída de frutos, se debió a la incidencia y severidad en un 99%, lo cual se atribuye a ambas variables de la enfermedad.

Tal resultado, es similar a lo descrito por Avelino, Muller (1991) y Moreno de Alas, que altas infecciones de roya, producen severas defoliaciones en las plantas de café.

## Finca El Olvido (estricta altura)

### Niveles de incidencia y severidad

**Tabla 15:** Nivel de incidencia de la enfermedad en hojas viejas y nuevas en los meses de mayo de 2016 a abril de 2017 en finca de estricta altura.

Ubicación de bandola en el árbol	Bandola																	
	Cantidad de hojas enfermas por árbol																	
	Hojas viejas						Hojas nuevas											
	May	Jun	Jul	Ago	Sep	Oct	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Ene	Feb	Mar	Abr
Alta	0.9	0.7	0.5	0.4	1	0	1.5	1.1	0.8	0.7	2.2	1.1	0.5	1.9	3.7	0.7	0.2	0.2
Media	1.3	1.2	0.2	0.3	1.1	0	2.1	1.9	0.4	0.5	3.6	1.7	0.4	4.5	5.9	4.9	3.3	0.8
Baja	1.5	0.8	0.2	0.1	0.1	0	2.4	1.3	0.3	0.2	0.5	0.42	0.8	1.4	5.1	1.5	2.2	0.5
Total	3.7	2.7	0.9	0.8	2.2	0	6.0	4.3	1.5	1.4	6.3	3.22	1.7	7.8	14.7	7.1	5.7	1.5

La tabla 15, muestra que, en hojas viejas, se encontró, una infección inicial en el mes mayo de 3.70 hojas enfermas; con una progresiva disminución, hasta que desaparecieron en el mes de octubre; debido a la caída de hojas en su totalidad. En el caso de hojas nuevas, la infección, en el mes de mayo; se encontró un promedio de 6 hojas infectadas; pero como producto del azar se observa una baja presencia hasta los meses de octubre y noviembre, con un ligero repunte en el mes de diciembre y enero, luego continuó con su descenso hasta abril.

En cuanto a los valores encontrados en cada estrato del árbol, fueron similares en las tres secciones, contrario a lo que describe Avelino (1991) y Rivillas Osorio, et al (2011), que la roya del café tiene mayor presencia en las partes inferiores del árbol.

Es de hacer notar, que para los tres casos (bajío, media y estricta altura) el tamaño de las pústulas que se encontraron en las hojas viejas fueron de mayor tamaño, que las encontradas en las hojas nuevas, que fueron pústulas muy pequeñas y crecieron hasta igualarse a las encontradas en las hojas viejas, y en muchos casos se unieron con otras pústulas vecinas formando una sola lesión ocasionando en su mayoría la muerte de la hoja en su totalidad. Resultados similares a lo descrito por Avelino y Muller (1991), que mencionan que altas infectaciones en una misma área foliar, las pústulas tienden a unirse con otras, lo que produce que áreas adyacentes aún no afectadas directamente tienda a morir, afectando a la hoja entera

**Tabla 16:** Nivel de severidad de la enfermedad en hojas viejas y nuevas, durante los meses de mayo de 2016 a abril de 2017 en la finca de astricta altura.

Ubicación de bandola en el árbol	Bandola																	
	Cantidad de pústula por árbol																	
	Hojas viejas						Hojas nuevas											
	May	Jun	Jul	Ago	Sep	Oct	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Ene	Feb	Mar	Abr
Alta	1.5	1.1	0.8	0.67	1.7	0	1.3	1	0.8	0.7	1.2	0.6	0.2	0.8	1.1	0.5	0.2	0.2
Media	2.1	1.9	0.3	0.5	1.8	0	1.8	1.1	0.4	0.5	2	0.6	0.4	2.7	3.6	2.8	1.5	0.6
Baja	2.4	1.3	0.3	0.17	0.1	0	2.1	1.3	0.3	0.2	0.5	0.4	0.8	1	2.8	0.5	1.7	0.4
Total	6.0	4.3	1.4	1.34	3.6	0	5.2	3.4	1.5	1.4	3.7	1.6	1.4	4.5	7.5	3.8	3.4	1.2

La tabla 16, se describen, los niveles de severidad encontrados, en las hojas viejas y nuevas; de los cuales en las hojas viejas de 6.0 pústulas por árbol en promedio, descendiendo hasta desaparecer en el mes de octubre, como resultado de la caída total de las hojas.

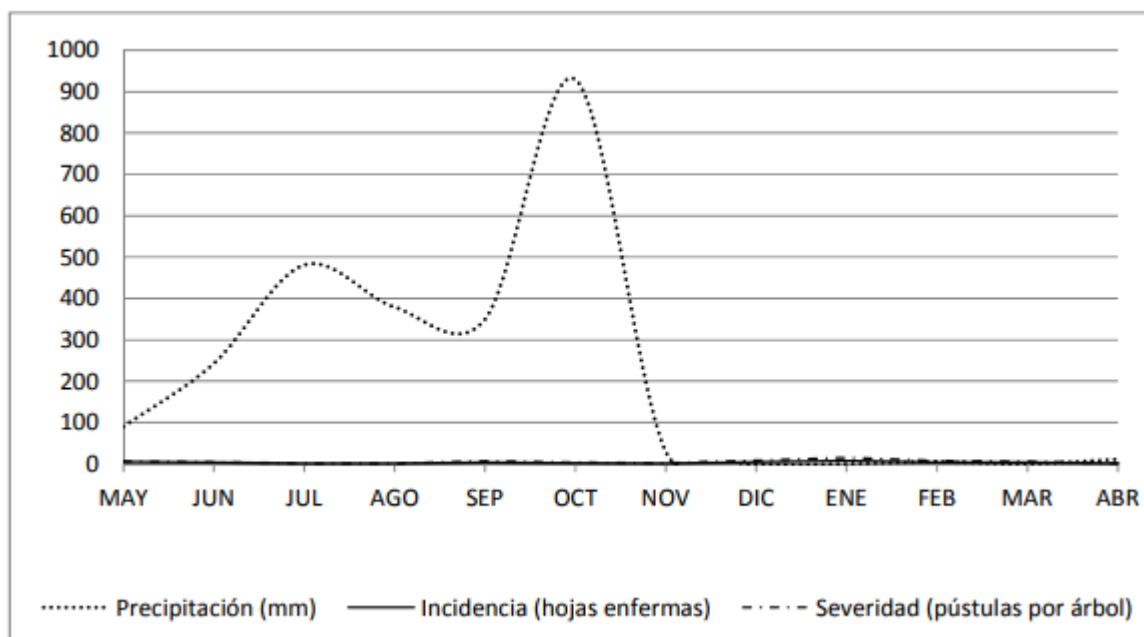
Tanto en la incidencia como en la severidad de la enfermedad, el inóculo residual se encontró en las hojas viejas de los árboles; que sucesivamente infestan e infectan las hojas nuevas, producto de la policiclicidad del patógeno, las hojas continúan infectándose, aunque en este caso los niveles fueron bajos en relación a lo registrado en bajo y media altura.

En las hojas nuevas, los niveles encontrados fueron mínimos, en todo el periodo de estudio; encontrando valores promedios de 5.2 pústula por árbol en el mes de mayo, y con un leve repunte en los meses de diciembre y enero, luego comienza un descenso hasta el mes de abril. Es muy importante resaltar, que la tendencia de la curva, es similar en los tres estratos altitudinales donde se presenta la fase de desarrollo lento (periodo de infestación e infección inicial), crecimiento acelerado, el punto máximo (cúspide de la curva de la enfermedad) y fase de descenso, aunque en este caso los niveles de infección fueron bajos.



### Relación incidencia y severidad con la precipitación.

**Gráfico 15:** Curvas de incidencia y severidad y su relación con la precipitación en la finca El Olvido (estricta altura).



Fuente: creación propia

En el gráfico 15, se indica la relación en los niveles de la precipitación, y según el análisis estadístico, el factor lluvia no influyó en la incidencia de la roya del café (*Hemileia vastatrix*); mientras que, para la severidad, sí influyó significativamente en su desarrollo, en el mismo periodo comprendido de mayo de 2016 hasta abril de 2017.

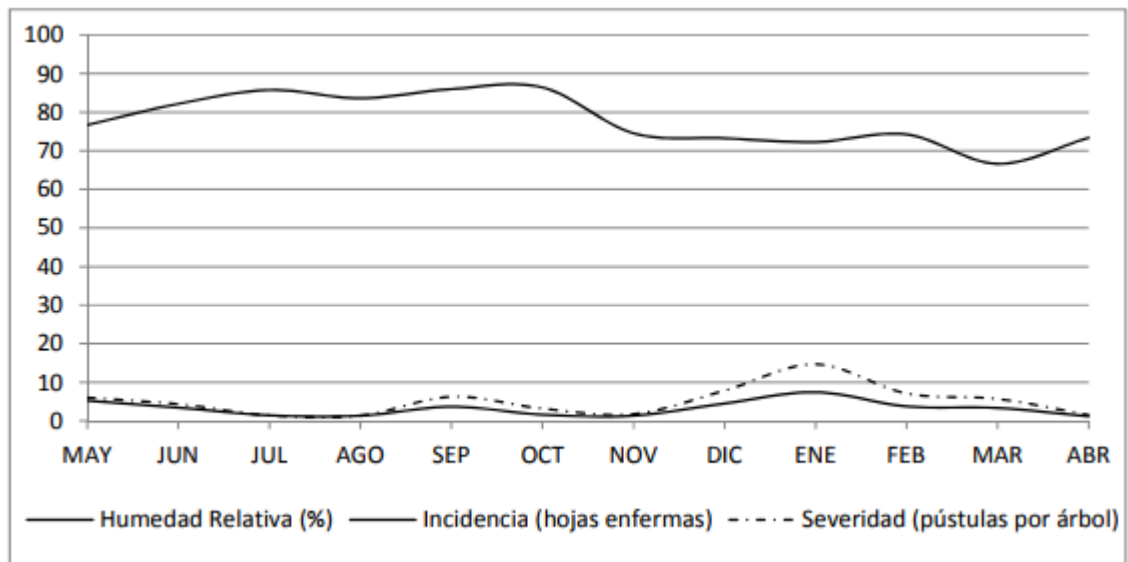
Aunque es este caso los niveles fueron muy bajos, en relación a las fincas ubicadas en bajo y media altura. Sin embargo, al aplicar la prueba de regresión y el coeficiente de determinación indica que la precipitación influyó en un 23% en la incidencia y un 40% en la severidad de la enfermedad.

Así se comprueba que la lluvia, influyó en el desarrollo de la epifitología de la roya del café, similar a lo que describe Rayner (1972), que dicen que, al menos un factor climático influye en mayor medida en el comportamiento de la enfermedad. La severidad, se ve incrementada, debido a que, de una pústula, la lluvia dispersa las esporas de roya,

infestando e infectando las áreas adyacentes, lo que aumentó, el número de pústulas en áreas y hojas vecinas.

### Relación incidencia y severidad con la humedad relativa

**Gráfico 16:** Curvas de incidencia y severidad y su relación con la humedad relativa en la finca El Olvido (estricta altura).

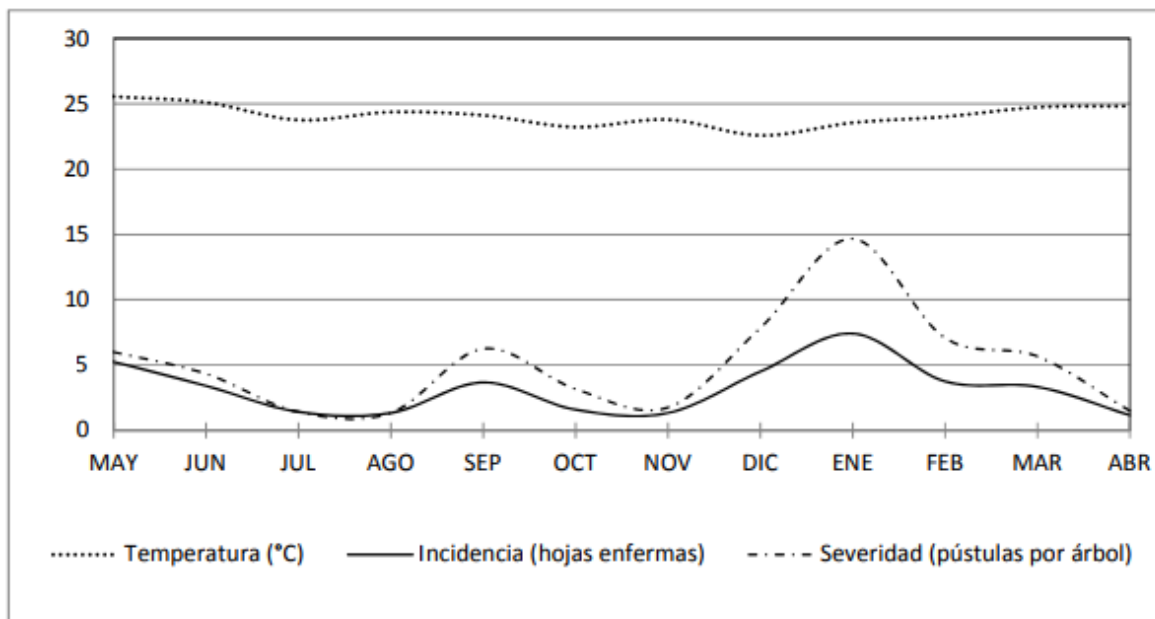


Fuente: creación propia

En el gráfico 16, se muestra la variación en los niveles de la humedad relativa y su relación con la incidencia y severidad de la roya del café, (*Hemileia vastatrix*), observándose que la humedad relativa, no influyó en la curva epifitológica de la enfermedad en estricta altura. Además, la prueba de regresión y el coeficiente de determinación muestra que la humedad relativa influyó en un 6%, en la incidencia y en un 16% en la severidad. Sin embargo, la tendencia que se observa en la curva, muestra asimetría en su comportamiento en ambas variables (incidencia y severidad), con relación a la humedad relativa, lo cual es un resultado, que no se puede ignorar.

## **Incidencia y severidad con la temperatura**

**Gráfico 17:** Curvas de incidencia y severidad y su relación con la humedad relativa en la finca El Olvido (estricta altura).



Fuente: creación propia

En el gráfico 17, se puede observar, la variación en la temperatura y su relación con la incidencia y severidad de la roya del café (*Hemileia vastatrix*), notándose en ambos casos, que ésta no tuvo influencia estadística significativa, en el periodo de mayo de 2016 hasta abril de 2017. Sin embargo, en la prueba de regresión; utilizando el coeficiente de determinación, se muestra que la temperatura registrada, influyó en un 13%, en la incidencia y un 20%, en la severidad de la enfermedad; lo cual es un resultado, que no se puede ignorar.

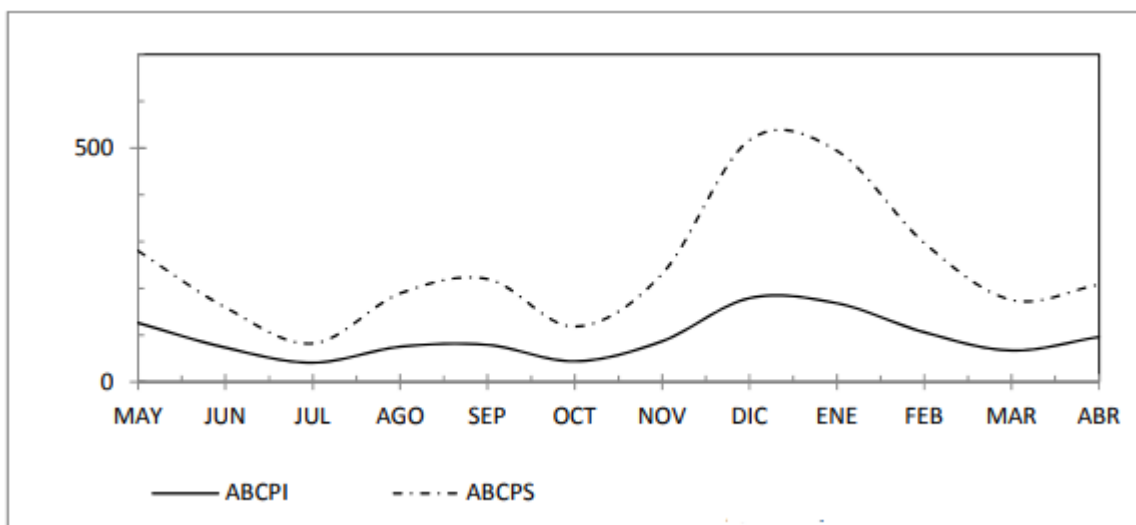
### **Relación incidencia con la severidad**

Los niveles de severidad de la roya del café (*Hemileia vastatrix*), dependieron de los niveles de incidencia, como resultado de la dispersión por salpicadura provocada por la lluvia y otros factores. Resultado similar a lo descrito por Avelino (1991) y Rivillas Osorio et al (2011). Además, con la prueba de regresión utilizando el coeficiente de

determinación se reflejan que, el aumento en la severidad, dependió en un 90% del nivel de incidencia encontrado.

### Tasa de infección aparente (r).

**Gráfico 18:** Área Bajo la Curva de la Incidencia y Severidad (ABCPI y ABCPS) de la roya del café en la finca El Olvido (estricta altura).



Fuente: creación propia.

La velocidad de infección promedio, tuvo un crecimiento desigual, utilizando la tasa de infección aparente, en la incidencia se tiene 0.79 y severidad 1.22. En la última es mayor, debido a un factor (lluvia) en especial, que favoreció más a la severidad; dado que esta dependió de la incidencia, como fuente de inóculo.

La lluvia dispersó y aumentó la cantidad de pústulas, generaron mayor velocidad, en el crecimiento de la severidad; en la misma unidad foliar, en la finca de estricta altura; aun cuando los niveles encontrados, en ambos casos fueron insignificantes. El gráfico 18, muestra las etapas y la intensidad de la incidencia y severidad de la enfermedad; determinándose que ambos parámetros, coinciden en su comportamiento; no así, en los niveles alcanzados, aunque no se encontraron niveles alarmantes, si mostró, que el comportamiento es similar.

Sin embargo, el Área Bajo la Curva de Progreso de la severidad (ABCPS), es siempre mayor que la incidencia; donde se observa etapa de crecimiento lento de mayo a octubre, una leve aceleración en noviembre, un máximo en diciembre y levemente descendió en los meses siguientes enero y febrero. Además, se observa proporcionalmente un comportamiento simétrico en las dos variables.

### **Perdida de frutos.**

En estricta altura, los resultados encontrados, sobre los niveles de incidencia y severidad analizados estadísticamente son insignificantes, salvo en el caso de la precipitación, que influyó moderadamente en la severidad. Así mismo, se determinó; que la intensidad de la enfermedad, no tuvo incidencia, en la cantidad de frutos encontrados; debido a que tanto la producción como los niveles de la enfermedad fueron bajos.

Es de resaltar, que en las fincas de bajo (Normandía) y media altura (San Antonio), los niveles de la enfermedad fueron altos, así como la producción, contrario a lo que ocurrió en estricta altura (El Olvido), donde la producción y los niveles de roya fueron bajos. Este resultado coincide, con lo descrito por Avelino y Muller (1991) y Rivillas Osorio et al (2011), que expresan, que altos niveles de roya están asociados, a ciclos de alta producción de café

### **Conclusiones y recomendaciones.**

#### **Conclusiones:**

La enfermedad no desaparece de un ciclo de producción a otro solo mantiene gradualmente ciertos niveles, en las diferentes épocas, según condiciones de ambiente y de manejo, manifestándose de forma policíclica y bianual, afectando la caficultura en pisos altitudinales que antes no se desarrollaba y debe existir un manejo adecuado de la enfermedad, para que no se presenten los mismos problemas, observados en media altura.

En este estudio se ha presentado que los factores ambientales como lo son: lluvia, temperatura y humedad, influyen para la incidencia y severidad de la roya del cafeto en las distintas zonas altitudinales, pero de diferentes maneras en donde se cosecha el café.

La incidencia de la roya, en baja altura se debió a la lluvia y la temperatura ya que son los factores que presentan un mayor porcentaje de influencia.

La incidencia de la roya, en media altura, se debió a la lluvia y humedad relativa ya que se obtuvo un porcentaje de influencia muy elevado.

La incidencia de la roya, en estricta altura, se debió a los factores ambientales lluvia y temperatura, pero en muy bajo porcentaje de influencia.

La severidad de la enfermedad de la roya, se debe a la aceleración de la incidencia.

La mayor tasa de infección aparente se tiene en la zona de baja altura y la mayor tasa de infección aparente se tiene en la zona de mediana altura, es decir que en la zona de baja altura la roya en el cafeto se propaga más rápido.

La pérdida de los frutos en las tres zonas, es causada por la incidencia y severidad de la enfermedad de la roya en el cafeto.

### **Recomendaciones para los caficultores:**

El control de la roya del café (*hemileia vastatrix*), debe iniciarse en los meses donde se produce la infestación e infección inicial, eliminando de esta manera el inoculo residual, que da origen a toda infestación.

Realizar estudios más precisos y detallados de la roya del café, en estricta altura, ya que este estrato altitudinal, no existía evidencia de que esta enfermedad se presentase.

El manejo integrado del cultivo, debe basarse en el comportamiento de la enfermedad, según el nivel altitudinal donde se ubique la finca.

En el establecimiento y renovación de cafetales, sembrar variedades resistentes y tolerantes, y seleccionar suelos de buena consistencia textural y nutricional, para asegurar un buen desarrollo de las plantas y de esta manera mejorar la resistencia a la enfermedad.

Realizar vigilancia constante, para evitar que la enfermedad sobrepase los niveles de daño económico; relacionándolos a diferentes prácticas de manejo tradicionales y

convencionales, y dosis de productos químicos, para establecer un control efectivo y eficaz de la enfermedad.

Realizar este estudio con más frecuencia y con las mismas zonas que se tomaron en esta investigación para poder conocer que tanto ha cambiado la enfermedad al pasar del tiempo.

Incluir otros factores ambientales para verificar, que tanto influyen estos factores en la propagación de la roya.

Estudiar a que factores se debe la incidencia de la roya a estricta altura ya que los factores ambientales mencionados tienen muy baja la influencia.

Realizar un estudio con los distintos tipos de fertilizantes para combatir la roya, y verificar cual es el mejor para cada zona altitudinal.

## **CAPÍTULO 7: CONCLUSIONES Y RECOMENDACIONES DEL MANUAL**

### **7.1 CONCLUSIONES**

La elaboración de este manual es de suma importancia para los estudiantes de la Universidad de El Salvador Facultad Multidisciplinaria de Occidente de la Carrera de Biología, ya que les ayuda en la realización de análisis estadístico tanto descriptivo como inferencial para los trabajos de materia o de grado.

En este manual se ha querido presentar de una forma más sencilla como poder realizar un análisis estadístico descriptivo e inferencial, cuenta con conceptos básicos y también con un capítulo donde se ponen en práctica el contenido de Inferencia estadística y diseño de experimentos, para que el lector puede conocer el tratamiento que se les da a los distintos datos y las interpretaciones que se les da a los resultados de cada uno de los procesos en la estadística.

Se tomaron en cuenta en este manual diseño de experimentos para contestar algunos objetivos o realizar pruebas de comparación, ya que en el área de Biología se realizan muchos experimentos en donde se aplican distintos tratamientos de los cuales se quiere conocer cuál debiera ser el más conveniente de utilizar para poder tomar una buena decisión en los experimentos.

El manual presenta en el capítulo dos un apartado, en donde se explica el procedimiento que se lleva a cabo para la recolección de datos.

Se proporcionan algunos ejemplos para que los estudiantes, conozcan de qué forma se ponen en práctica cada uno de los temas presentados ya sea, para materia o para trabajo de grado.

En la parte de aplicaciones se presentan las conclusiones planteadas en esa tesis y también las conclusiones propuestas; de igual manera, se presentan las recomendaciones.

La estadística descriptiva es de mucha utilidad en el área de Biología, ya que la mayoría de las investigaciones hacen uso de ella.



Diseño de experimentos ayuda a las investigaciones en el área de biología a proporcionar información pertinente al problema que se está planteando en una investigación y ajustar el diseño para que sea más simple y efectivo; para poder ahorrar dinero, tiempo, personal y material experimental.

La estadística descriptiva y estadística inferencial no es solo presentar un estudio de manera organizada y precisa, sino que es un estudio que ayuda a la Biología para dar respuesta a las interrogantes que se presentan en cada una de las investigaciones y así poder solucionar o atacar el problema planteado.

La utilización del software estadístico SPSS en las investigaciones es de mucha ayuda para organizar los datos observados en los experimentos y así poder ahorrar tiempo y dinero.

## **7.2 RECOMENDACIONES**

### Para el Departamento de matemáticas:

La elaboración de un manual para dar a conocer cómo se pueden utilizar los distintos programas informáticos en el área de estadística para darle tratamiento a los datos recolectados en cada investigación.

Los alumnos de la Licenciatura en Estadística, puedan hacer su servicio social asesorando a los egresados que estén realizando su trabajo de grado; ya sea de la Licenciatura en Biología o de otras carreras.

### Para el Departamento de Biología:

Agregar una materia en la cual se pueda ampliar la forma de poder realizar los Análisis Estadísticos y el uso de software especializado.

Es necesario que los alumnos de Licenciatura en Biología cuenten con computadoras dentro de los laboratorios donde realizan las investigaciones para que guarden cada uno de los datos del experimento y así ahorrar tiempo de digitación para realizar análisis estadísticos.

Los docentes que imparten la materia de bioestadística, es recomendable que los ejemplos y ejercicios que se le presenten al alumno, del contenido impartido sean relacionados con la Biología para una mejor comprensión de los temas que desarrollan.

Se recomienda para los docentes que imparten la materia de bioestadística, que puedan dar horas de laboratorio, prácticas o discusiones para que los alumnos puedan plantear sus dudas del área de estadística

## **BIBLIOGRAFIA**

- Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2008). *Estadística para administración y economía, 10a. edición*. México,DF: Cengage Learning Editores,1060p.
- Arias, F. G. (2006). *El Proyecto de Investigación, 6a edición*. Venezuela: Editorial Episteme,145p.
- Chao, L. L. (1993). *Estadística para las Ciencias Administrativas* . México: McGraw-Hill,458p.
- gorgas, J., Gardiel, N., & Zamorano, J. (2011). *Estadística básica para estudiantes de ciencia, 1a edición* . Madrid, España: Universidad complutense de madrid,260p.
- Gujarati, D. N., & Porter, D. C. (2010). *Econometría, 5a edición*. México: McGraw-Hill,923p.
- Jonhson, R., & Kuby, P. (2008). *Estadística Elemental, 10a edición*. México: Cengage Learning Editores,732p.
- Lastra, P., Ramos, R., & Ramos, L. F. (1995). *Mujeres deportadas por las autoridades Norteamericanas*. México: UNAEM,150p.
- Levin, R. L., & Rubin, D. S. (1996). *Estadística para Administración y Economía, 2a edición*. Mexico: Pearson Educación,958p.
- Lind, D. A., Mason, R. D., & Marchal, W. G. (2001). *Estadística para administración y economía, 3a edición*. México: McGraw-Hill,576p.
- Montgomery, D. C. (2004). *Diseño y Análisis de Experimentos, 2a edición*. México, DF: Editorial Limusa,688p.
- Montgomery, D. C., Peck, E. A., & Vining, G. (2006). *Introducción al Análisis de Regresión Lineal, 3a edición*. México: Continental,590p.

- Peña, D. (2001). *Fundamentos de Estadística 1a edición*. Madrid, España: Alianza Editorial,683p.
- Sampieri, R. H. (2010). *Metodología de la Investigación 5a edición*. México: McGraw-Hill,615p.
- Sampieri, R. H. (2014). *Metodología De La Investigación, 6a edición*. México: MacGraw-Hill,636p.
- Spiegel, M. R., & Stephens, L. J. (2009). *Estadística 4a edición*. México, DF: McGrawHill,577p.
- Walpole, R. E. (1999). *Probabilidad y Estadística Para Ingenieros, 6a edición*. México: Prentice-Hall. Inc,739p.
- Walpole, R. E., Myers, R. H., & Myers, S. L. (2012). *Probabilidad y Estadística Para Ingeniería y Ciencias, 9a edición*. México: Pearson Education, Inc,792p.
- Webster, A. L. (1996). *Estadística aplicada a los negocios y la economía, 3a edición*. México: McGraw-Hill,652p.
- Wilhelmi, M. R. (2004). *combinatoria y probabilidad, 1a edición*. España: GEEUG,576p.

# ANEXO

## ANEXOS

### Anexo 1: Ejemplos prácticos del programa SPSS

El software SPSS es de mucha utilidad, en el área de estadística para poder realizar de una forma más rápida tablas y gráficos, para luego solo realizar el análisis de los resultados obtenidos, que nos ayudaran a contestar nuestros objetivos.

El software nos ayuda para realizar cálculos tanto en la estadística descriptiva como en la estadística inferencial.

#### Estadística descriptiva

Para explicar cómo se pueden realizar los cálculos para la estadística descriptiva en este software tomaremos los siguientes datos.

Se ha realizado una investigación para conocer el tiempo de vida de 100 moscas.

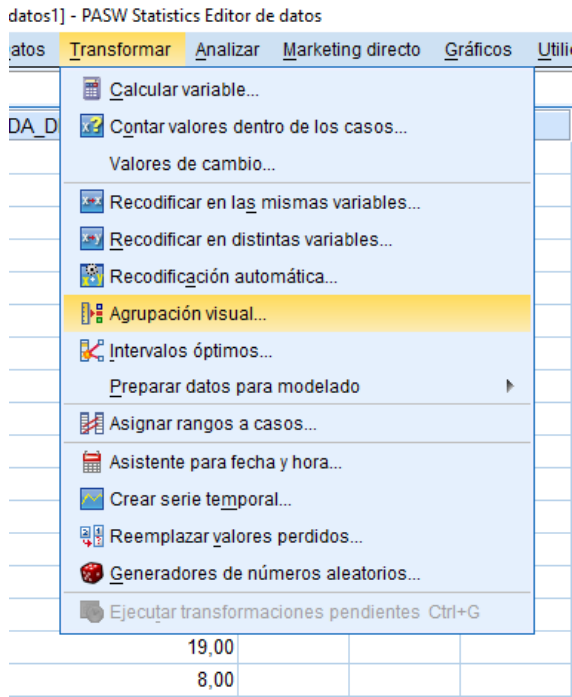
#### Anexo 1.1: Distribuciones de frecuencia

Los datos tienen que estar en el área de trabajo del programa SPSS.

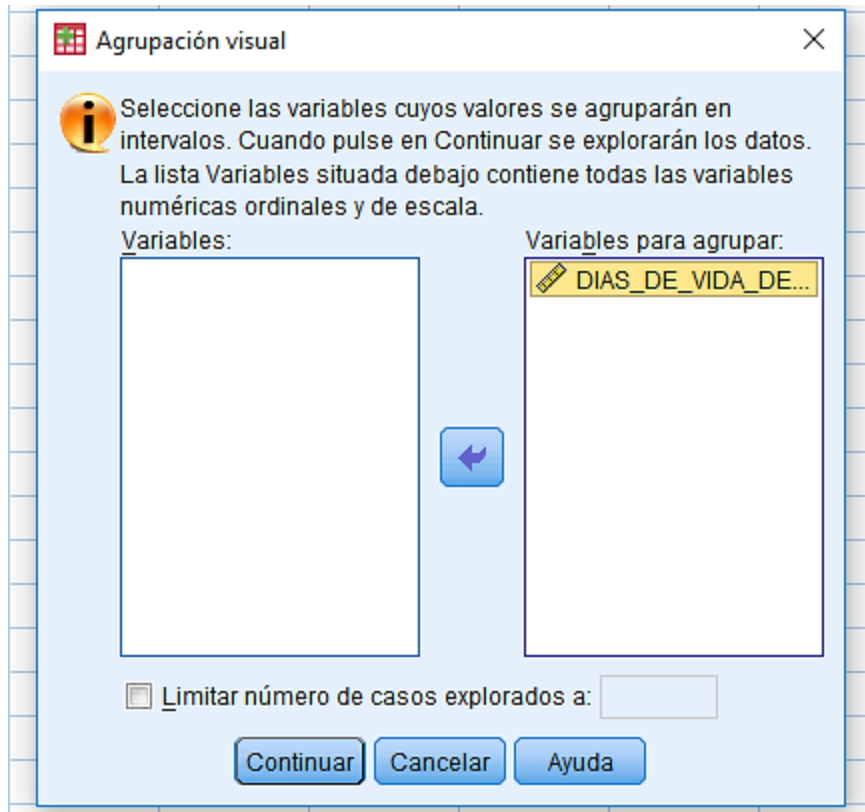
	DIAS_DE_VIDA_DE_UN_100_MOSCA	var
1	26,00	
2	15,00	
3	14,00	
4	7,00	
5	5,00	
6	,0	
7	17,00	
8	4,00	
9	19,00	
10	13,00	
11	19,00	
12	7,00	
13	19,00	
14	12,00	
15	23,00	
16	19,00	
17	8,00	
18	18,00	
19	10,00	
20	19,00	
21	,0	
22	17,00	
23	5,00	
24	11,00	
25	23,00	

Vista de datos Vista de variables

Luego se calcula el número de clases y la amplitud del intervalo



Al seleccionar: agrupación visual se abre la siguiente pantalla



Donde se selecciona la variable deseada y luego continuar.





En el cuadro posición del primer punto de corte, se colocara el valor mínimo más la amplitud, que en este caso sería (  $0 + 4 = 4$  ) y número de puntos y anchura se pueden probar hasta tener el valor deseado en posición del último punto de corte.

Recomendación: calcular el número de clases que este representaría el número de puntos de corte y la amplitud del intervalo sería la anchura, de forma manual para obtener datos más puntuales.

Crear puntos de corte

Intervalos de igual amplitud

Intervalos: rellene al menos dos campos

Posición del primer punto de corte: 4,00

Número de puntos de corte: 8

Anchura: 4

Posición del último punto de corte: 32,00

Percentiles iguales basados en los casos explorados

Intervalos - rellene cualquiera de los dos campos

Número de puntos de corte:


% de casos:

Puntos de corte en media y desviaciones típicas seleccionadas, basadas en casos explorados

+/- 1 Desv. Desviación

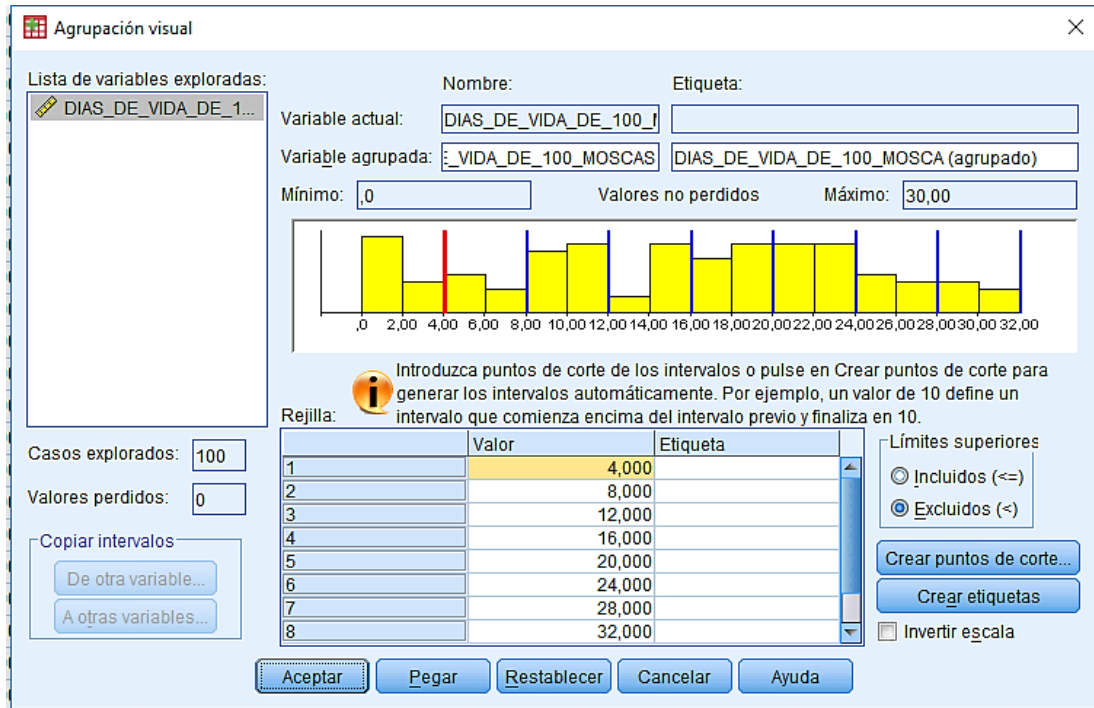
+/- 2 Desv. Desviación

+/- 3 Desv. Desviación

 Aplicar reemplazará las definiciones de los puntos de corte actuales con esta especificación.  
Un intervalo final incluirá todos los valores restantes: N puntos de corte generan N+1 intervalos.

Aplicar Cancelar Ayuda

Luego de dar aplicar, regresa a la pantalla de inicio para verificar si el usuario está de acuerdo con los resultados y en límites superiores se selecciona excluidos, para que tome el valor menor que.

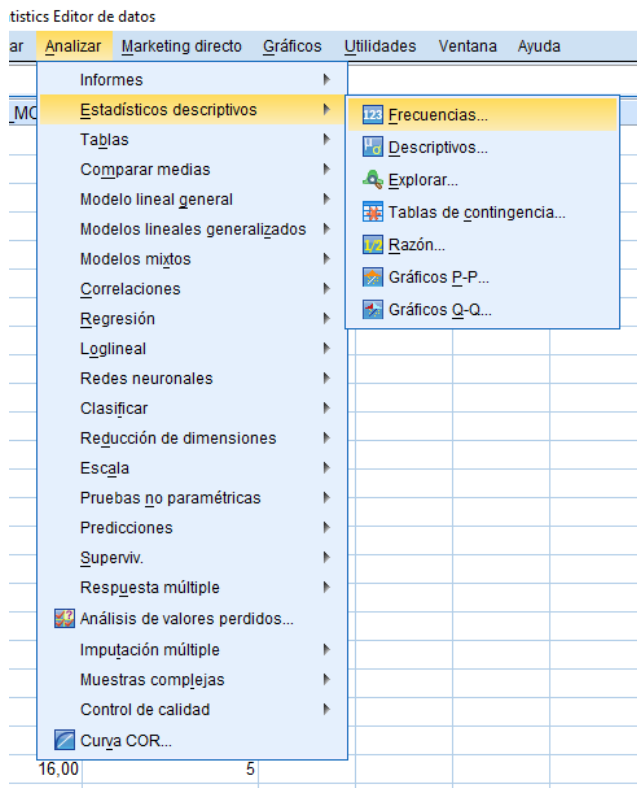


Se verifica en la parte inferior donde dice rejilla y si al usuario le parece se selecciona aceptar y en el área de trabajo aparecerán los resultados de la siguiente manera.

	DIAS_DE_VIDA_DE_100_MOSCA	DIAS_DE_VIDA_DE_100_MOSCAS
1	26,00	7
2	15,00	4
3	14,00	4
4	21,00	6
5	5,00	2
6	,0	1
7	17,00	5
8	4,00	2
9	19,00	5
10	13,00	4
11	19,00	5
12	7,00	2
13	19,00	5
14	12,00	4
15	23,00	6
16	19,00	5
17	8,00	3
18	18,00	5
19	10,00	3
20	19,00	5
21	,0	1
22	17,00	5
23	5,00	2
24	11,00	3
25	23,00	6

Vista de datos    Vista de variables

Luego para conocer la tabla de distribución de frecuencia se realizaran los siguientes pasos, solo que en este caso será con la nueva variable.



Luego se selecciona la variable creada para trabajar con datos agrupados



Se selecciona mostrar tablas de frecuencia y luego aceptar; el resultado sería el siguiente.

Resultados:

**Estadísticos**

DIAS\_DE\_VIDA\_DE\_100\_  
MOSCA (agrupado)

N	Válidos	100
	Perdidos	0

**DIAS\_DE\_VIDA\_DE\_100\_MOSCA (agrupado)**

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos 1	14	14,0	14,0	14,0
2	8	8,0	8,0	22,0
3	17	17,0	17,0	39,0
4	11	11,0	11,0	50,0
5	16	16,0	16,0	66,0
6	18	18,0	18,0	84,0
7	9	9,0	9,0	93,0
8	7	7,0	7,0	100,0
Total	100	100,0	100,0	

En la parte superior a la tabla nos dice estadísticos, eso nos sirve para conocer cuántos datos están válidos en nuestro estudio, ya que hay casos en los que se pueden perder datos o se vuelven datos atípicos (datos diferentes a los datos del estudio).

En la tabla llamada tipo de orquídeas se puede ver que nos aparece frecuencia esa frecuencia es la frecuencia absoluta simple, porcentaje es la frecuencia relativa simple, porcentaje válidos es el porcentaje de cada uno de los tipos que fueron válidos, porcentaje acumulado representa la frecuencia relativa acumulada.

Para la interpretación de los resultados se hace de igual manera como se explicó anteriormente.

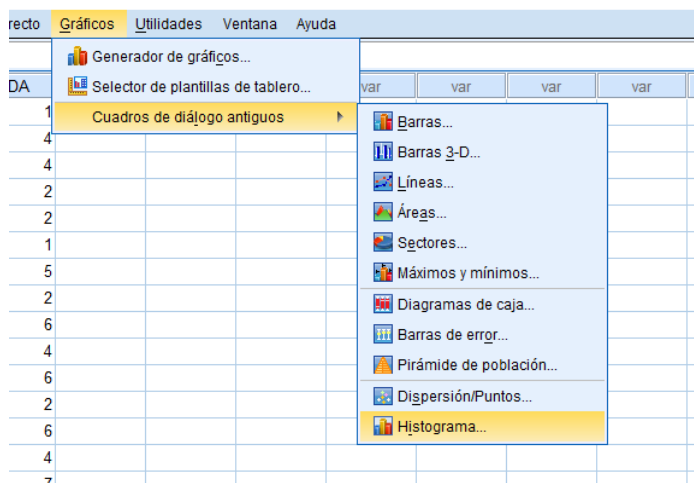
## Anexo 1. 2: Creación de gráficos

Teniendo los datos en el área de trabajo.

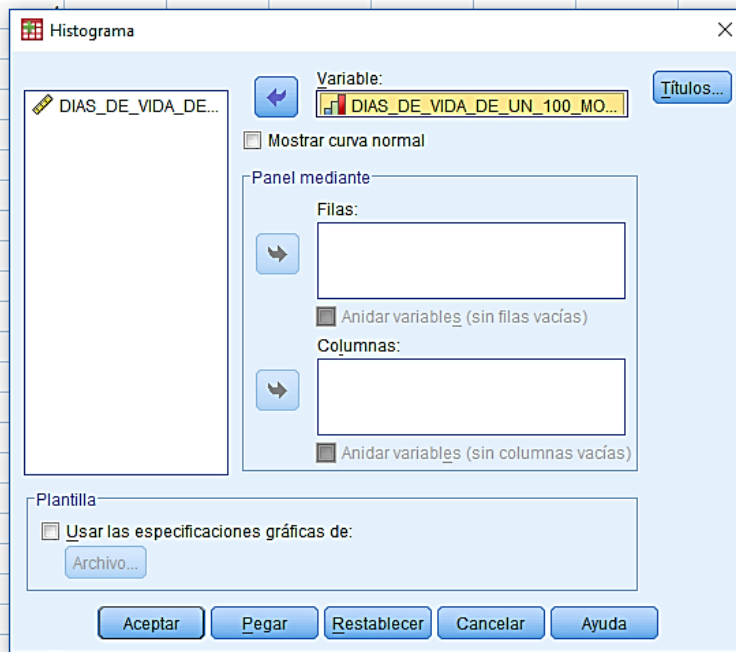
	DIAS_DE_VIDA_DE_100_MOSCA	DIAS_DE_VIDA_DE_100_MOSCAS
1	26,00	7
2	15,00	4
3	14,00	4
4	21,00	6
5	5,00	2
6	,0	1
7	17,00	5
8	4,00	2
9	19,00	5
10	13,00	4
11	19,00	5
12	7,00	2
13	19,00	5
14	12,00	4
15	23,00	6
16	19,00	5
17	8,00	3
18	18,00	5
19	10,00	3
20	19,00	5
21	,0	1
22	17,00	5
23	5,00	2
24	11,00	3
25	23,00	6

Vista de datos Vista de variables

Se realizaran los siguientes pasos: graficar, cuadro de dialogo antiguos e histograma. Si no es un histograma el que se desea se puede seleccionar cualquier gráfico.



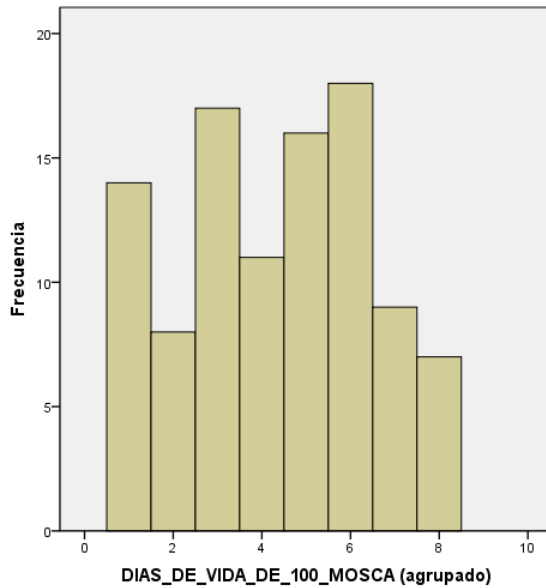
Luego aparecera la siguiente pantalla, en donde se utilizara la variable que se creo anteriormente con los datos agrupados y se colocara en en el cuadro donde dice variable.



Se selecciona aceptar y los resultados son los siguientes

## Resultados:

Grafico : Histograma dias de vida de 100 moscas



En este grafico en la parte inferior se tiene escala del 0 al 10; ya que cuando se agruparon los datos las opciones quedaron de 1 a 8 el numero de intervalos.

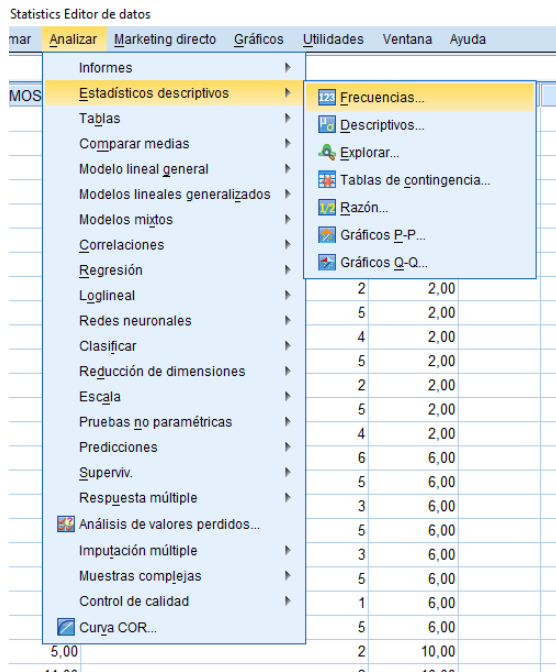
Descripcion: Este grafico muestra que las moscas mueren con más frecuencia en el intervalo 6 es cual representa entre 20 y 24 días, y con menor frecuencia mueren en el intervalo 5 el cual se encuentra entre 28 y 32 días.

### **Anexo 1.3: Análisis descriptivo**

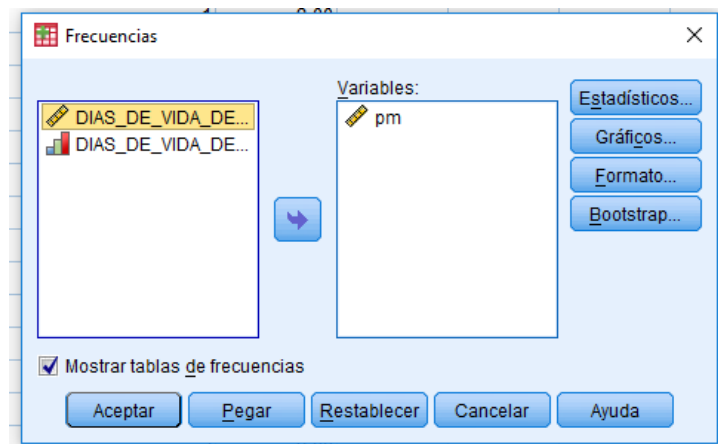
Con los siguientes datos que se han formado podemos realizar los distintos cálculos para el análisis descriptivo. Pero en este caso trabajaremos con los puntos medios de cada intervalo.



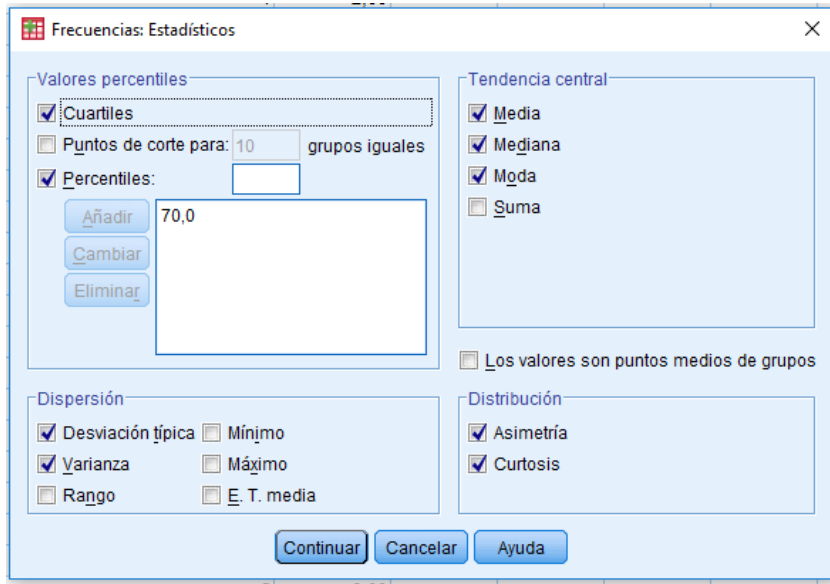
Se selecciona: analizar, estadísticos descriptivos, frecuencia.



Se seleccionará la variable pm que serán los puntos medios



Se selecciona estadísticos y todos los datos que se desean conocer y también se puede seleccionar mostrar tabla de frecuencia como se realizó anteriormente



Luego se continúa y después aceptar y los resultados son los siguientes.

Resultados:

**Estadísticos**

pm		
N	Válidos	100
	Perdidos	0
Media		15,2800
Mediana		16,0000
Moda		22,00
Desv. típ.		8,48776
Varianza		72,042
Asimetría		-,042
Error típ. de asimetría		,241
Curtosis		-1,040
Error típ. de curtosis		,478
Percentiles	25	10,0000
	50	16,0000
	70	22,0000
	75	22,0000

Interpretación:

**Media:** el promedio de vida de una mosca es de 15.28 días

**Mediana:** El 50% de los zancudos viven menos de 16 días y el otro 50% vive más de 50%.

**Moda:** los días en que más moscas mueren es a los 21.44 día.

**Desviación típica:** en promedio, la cantidad de días para que muera cada una de las moscas de estudio se separa de la media, 8.4878 (días).

**Varianza:** el promedio de las distancias a cuadrado del dato a la media es de 72.0420 días al cuadrado

**Asimetría:** la asimetría es negativa es decir tiende a la izquierda, lo que significa que las moscas tardan varios días en morir.

**Curtosis:** es negativa por lo tanto, la distribución es platicúrtica o por debajo de la normal. Ya que hay un intervalo en el que las moscas mueren con más frecuencia.

#### Anexo 1.4: Tablas de contingencia

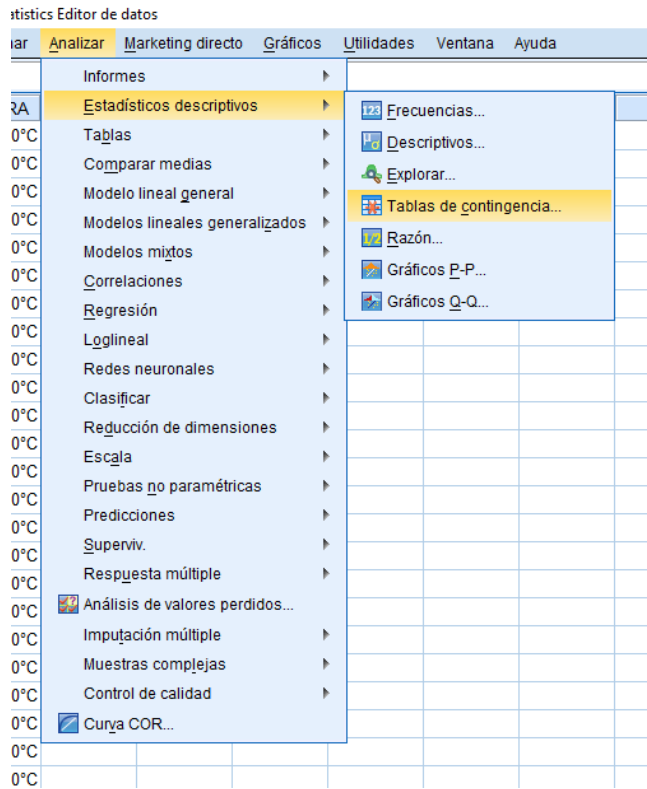
Se ha realizado una investigación para conocer a que temperatura en grados centígrados, mueren 100 bacterias de alimentos de distintos tipos. Temiendo los datos en el área de trabajo.

\*Sin título2 [Conjunto\_de\_datos2] - PASW Statistics Edi

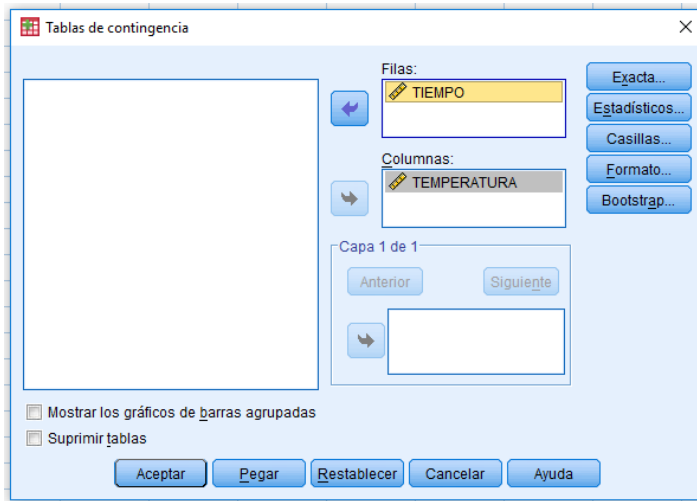
	TIEMPO	TEMPERATURA
1	1 HORA	40°C
2	2 HORAS	20°C
3	3 HORAS	30°C
4	1 HORA	30°C
5	2 HORAS	40°C
6	2 HORAS	40°C
7	1 HORA	30°C
8	1 HORA	40°C
9	2 HORAS	20°C
10	1 HORA	30°C
11	1 HORA	30°C
12	3 HORAS	40°C
13	3 HORAS	40°C
14	2 HORAS	40°C
15	1 HORA	20°C
16	2 HORAS	40°C
17	3 HORAS	40°C
18	2 HORAS	40°C
19	2 HORAS	20°C
20	3 HORAS	20°C
21	3 HORAS	30°C
22	1 HORA	30°C
23	3 HORAS	20°C
24	3 HORAS	40°C
25	2 HORAS	30°C

Vista de datos    Vista de variables

Para realizar una tabla de contingencia se realizan los siguientes pasos, se selecciona: analizar, estadísticos descriptivos y luego tablas de contingencia



Aparecerá el siguiente cuadro, en donde la filas será (x) y columnas (Y).



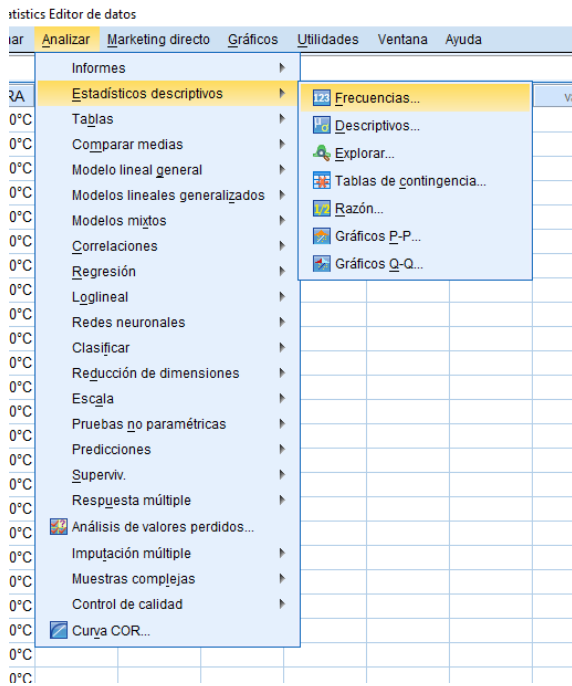
Resultados:

**Tabla de contingencia TIEMPO \* TEMPERATURA**

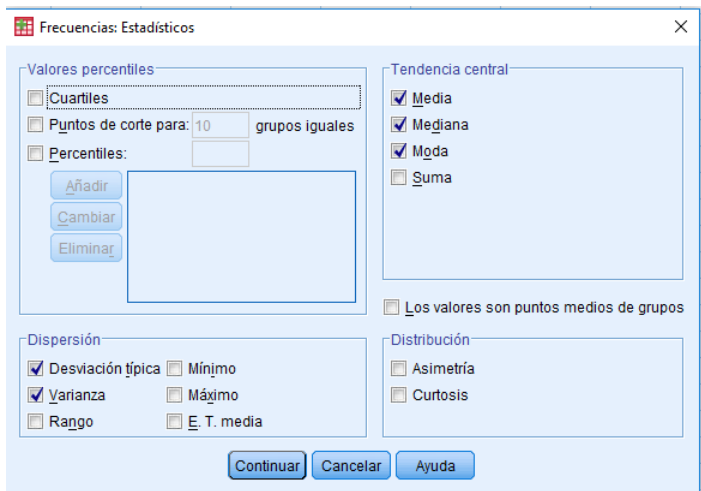
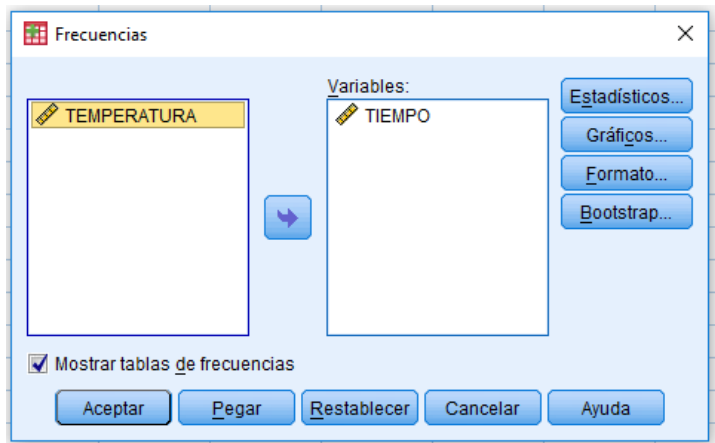
Recuento

		TEMPERATURA			Total
		20°C	30°C	40°C	
TIEMPO	1 HORA	1	5	2	8
	2 HORAS	3	3	6	12
	3 HORAS	3	2	5	10
Total		7	10	13	30

Para conocer los las medidas descriptiva se realiza por separado, en analizar



Y luego se selecciona la variable y luego todos los estadísticos que se deseen conocer.



Luego se selecciona continuar y despues aceptar y los resultados seran los siguientes.

Resultados:

**Estadísticos**

TIEMPO

N	Válidos	30
	Perdidos	0
Media		2,0667
Mediana		2,0000
Moda		2,00
Desv. típ.		,78492
Varianza		,616

**TIEMPO**

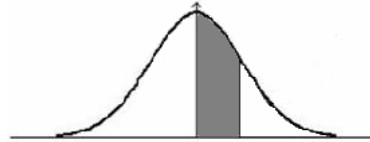
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 HORA	8	26,7	26,7	26,7
	2 HORAS	12	40,0	40,0	66,7
	3 HORAS	10	33,3	33,3	100,0
	Total	30	100,0	100,0	

Presenta los estadísticos (medidas descriptivas) y distribución de frecuencia.

**Anexo 2: Distribución normal estandarizada,  $N(0,1)$ .**

Distribución Normal

En los ejes están los valores de  $z$  y la tabla muestra el área del eje central a la derecha.



<b>Z</b>	<b>0.00</b>	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>	<b>0.04</b>	<b>0.05</b>	<b>0.06</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>
<b>0.0</b>	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
<b>0.1</b>	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
<b>0.2</b>	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
<b>0.3</b>	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
<b>0.4</b>	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
<b>0.5</b>	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
<b>0.6</b>	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
<b>0.7</b>	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
<b>0.8</b>	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
<b>0.9</b>	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
<b>1.0</b>	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
<b>1.1</b>	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
<b>1.2</b>	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
<b>1.3</b>	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
<b>1.4</b>	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
<b>1.5</b>	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
<b>1.6</b>	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
<b>1.7</b>	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633



<b>1.8</b>	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
<b>1.9</b>	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
<b>2.0</b>	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
<b>2.1</b>	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
<b>2.2</b>	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
<b>2.3</b>	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
<b>2.4</b>	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
<b>2.5</b>	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
<b>2.6</b>	0.495	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
<b>2.7</b>	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
<b>2.8</b>	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
<b>2.9</b>	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
<b>3.0</b>	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

### Anexo 3: Tabla de la distribución T de student

En los renglones se ubican los grados de libertad En las columnas son las áreas más usadas.

La tabla son los valores del eje  $x$  ( $t_{\text{área}}$ )



g.l.	$t_{0.1}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$
1	3.0777	6.3138	12.7062	31.8205	63.6567
2	1.8856	2.9200	4.3027	6.9646	9.9248
3	1.6377	2.3534	3.1824	4.5407	5.8409
4	1.5332	2.1318	2.7764	3.7469	4.6041
5	1.4759	2.0150	2.5706	3.3649	4.0321
6	1.4398	1.9432	2.4469	3.1427	3.7074
7	1.4149	1.8946	2.3646	2.9980	3.4995
8	1.3968	1.8595	2.3060	2.8965	3.3554
9	1.3830	1.8331	2.2622	2.8214	3.2498
10	1.3722	1.8125	2.2281	2.7638	3.1693
11	1.3634	1.7959	2.2010	2.7181	3.1058
12	1.3562	1.7823	2.1788	2.6810	3.0545
13	1.3502	1.7709	2.1604	2.6503	3.0123
14	1.3450	1.7613	2.1448	2.6245	2.9768
15	1.3406	1.7531	2.1314	2.6025	2.9467
16	1.3368	1.7459	2.1199	2.5835	2.9208
17	1.3334	1.7396	2.1098	2.5669	2.8982
18	1.3304	1.7341	2.1009	2.5524	2.8784
19	1.3277	1.7291	2.0930	2.5395	2.8609
20	1.3253	1.7247	2.0860	2.5280	2.8453
21	1.3232	1.7207	2.0796	2.5176	2.8314
22	1.3212	1.7171	2.0739	2.5083	2.8188
23	1.3195	1.7139	2.0687	2.4999	2.8073
24	1.3178	1.7109	2.0639	2.4922	2.7969
25	1.3163	1.7081	2.0595	2.4851	2.7874
26	1.3150	1.7056	2.0555	2.4786	2.7787
27	1.3137	1.7033	2.0518	2.4727	2.7707
28	1.3125	1.7011	2.0484	2.4671	2.7633
29	1.3114	1.6991	2.0452	2.4620	2.7564

Anexo 4: tabla de distribucion F de fisher (0.01)

Distribución F 0.01

En las columnas se encuentran los valores F que corresponden al área 0.05 a la derecha  
 En las columnas se encuentran los grados de libertad del numerador  
 En los renglones se encuentran los grados de libertad del denominador.

	1	2	3	4	5	6	7	8	9	10	11	12	15	20	24	30	40	60	120
1	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6083	6106	6157	6209	6235	6261	6287	6313	6339
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.41	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.13	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.45	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.96	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.24	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	3.06	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96	2.82	2.66	2.58	2.50	2.42	2.33	2.23
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.99	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.93	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53

**Anexo 5: tabla de valores críticos de la prueba de tukey  $q_\alpha(V_1, V_2)$**

$V_2$ 9	$\alpha$ 9	$V_1$									
		2	3	4	5	6	7	8	9	10	11
1	0.05	18.00	29.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07	50.59
	0.01	90.03	135.0	164.3	185.6	202.2	215.8	227.2	237.0	245.6	253.2
2	0.05	6.10	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99	14.39
	0.01	14.04	19.02	22.29	24.72	26.63	28.20	29.53	30.68	31.69	32.59
3	0.05	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	9.72
	0.01	8.26	10.62	12.17	13.33	14.24	15.00	15.64	16.20	16.69	17.13
4	0.05	3.93	5.04	5.76	6.29	6.71	7.05	7.34	7.60	7.83	8.03
	0.01	6.51	8.12	9.17	9.96	10.58	11.10	11.55	11.93	12.27	12.57
5	0.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17
	0.01	5.70	6.97	7.80	8.42	8.91	9.32	9.67	9.97	10.24	10.48
6	0.05	3.46	4.34	4.90	5.31	5.63	5.89	6.12	6.32	6.49	6.65
	0.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30
7	0.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30
	0.01	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55
8	0.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05
	0.01	4.74	5.63	6.20	6.63	6.96	7.24	7.47	7.68	7.87	8.03
9	0.05	3.20	3.95	4.42	4.76	5.02	5.24	5.43	5.60	5.74	5.87
	0.01	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.32	7.49	7.65
10	0.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72
	0.01	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36
11	0.05	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61
	0.01	4.39	5.14	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13
12	0.05	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.40	5.51
	0.01	4.32	5.04	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94
13	0.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43
	0.01	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79
14	0.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36
	0.01	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66
15	0.05	3.01	3.67	4.08	4.37	4.60	4.78	4.94	5.08	5.20	5.31
	0.01	4.17	4.83	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55
16	0.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26
	0.01	4.13	4.78	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46

v <sub>2</sub> 9	α 9	v <sub>1</sub>								
		12	13	14	15	16	17	18	19	20
1	0.05 0.01	51.96 260.0	53.20 266.2	54.33 271.8	55.36 277.0	56.32 281.8	57.22 286.3	58.04 290.4	58.83 294.3	59.56 298.0
2	0.05 0.01	14.75 33.40	15.08 34.13	15.38 34.81	15.65 35.43	15.91 36.00	16.14 36.53	16.37 37.03	16.57 37.50	16.77 37.95
3	0.05 0.01	9.95 17.53	10.15 17.89	10.35 18.22	10.53 18.52	10.69 18.81	10.84 19.07	10.98 19.32	11.11 19.55	11.24 19.77
4	0.05 0.01	8.21 12.84	8.37 13.09	8.52 13.32	8.66 13.53	8.79 13.73	8.91 13.91	9.03 14.08	9.13 14.24	9.23 14.40
5	0.05 0.01	7.32 10.70	7.47 10.89	7.60 11.08	7.72 11.24	7.83 11.40	7.93 11.55	8.03 11.68	8.12 11.81	8.21 11.93
6	0.05 0.01	6.79 9.49	6.92 9.65	7.03 9.81	7.14 9.95	7.24 10.08	7.34 10.21	7.43 10.32	7.51 10.43	7.59 10.54
7	0.05 0.01	6.43 8.71	6.55 8.86	6.66 9.00	6.76 9.12	6.85 9.24	6.94 9.35	7.02 9.46	7.09 9.55	7.17 9.65
8	0.05 0.01	6.18 8.18	6.29 8.31	6.39 8.44	6.48 8.55	6.57 8.66	6.65 8.76	6.73 8.85	6.80 8.94	6.87 9.03
9	0.05 0.01	5.98 7.78	6.09 7.91	6.19 8.03	6.28 8.13	6.36 8.23	6.44 8.32	6.51 8.41	6.58 8.49	6.64 8.57
10	0.05 0.01	5.83 7.48	5.93 7.60	6.03 7.71	6.11 7.81	6.20 7.91	6.27 7.99	6.34 8.07	6.40 8.15	6.47 8.22
11	0.05 0.01	5.71 7.25	5.81 7.36	5.90 7.46	5.99 7.56	6.06 7.65	6.14 7.73	6.20 7.81	6.26 7.88	6.33 7.95
12	0.05 0.01	5.62 7.06	5.71 7.17	5.80 7.26	5.88 7.36	5.95 7.44	6.03 7.52	6.09 7.59	6.15 7.66	6.21 7.73
13	0.05 0.01	5.53 6.90	5.63 7.01	5.71 7.10	5.79 7.19	5.86 7.27	5.93 7.34	6.00 7.42	6.05 7.48	6.11 7.55
14	0.05 0.01	5.46 6.77	5.55 6.87	5.64 6.96	5.72 7.05	5.79 7.12	5.85 7.20	5.92 7.27	5.97 7.33	6.03 7.39
15	0.05 0.01	5.40 6.66	5.49 6.76	5.58 6.84	5.65 6.93	5.72 7.00	5.79 7.07	5.85 7.14	5.90 7.20	5.96 7.26
16	0.05 0.01	5.35 6.56	5.44 6.66	5.52 6.74	5.59 6.82	5.66 6.90	5.72 6.97	5.79 7.03	5.84 7.09	5.90 7.15

v <sub>2</sub> 9	α 9	v <sub>1</sub>									
		2	3	4	5	6	7	8	9	10	11
17	0.05	2.98	3.63	4.02	4.30	4.52	4.71	4.86	4.99	5.11	5.21
	0.01	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38
18	0.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17
	0.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31
19	0.05	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14
	0.01	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25
20	0.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11
	0.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19
24	0.05	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01
	0.01	3.96	4.54	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02
30	0.05	2.89	3.49	3.84	4.10	4.30	4.46	4.60	4.72	4.83	4.92
	0.01	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.85
40	0.05	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.74	4.82
	0.01	3.82	4.37	4.70	4.93	5.11	5.27	5.39	5.50	5.60	5.69
60	0.05	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73
	0.01	3.76	4.28	4.60	4.82	4.99	5.13	5.25	5.36	5.45	5.53
120	0.05	2.80	3.36	3.69	3.92	4.10	4.24	4.36	4.48	4.56	4.64
	0.01	3.70	4.10	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.38
4	0.05	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55
	0.01	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23

Anexo 6: tabla de valores críticos de la prueba de Duncan  $U_\alpha(V_1, V_2)$

$v_2$	$\alpha$	$v_1$													
		2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	0.05	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0
	0.01	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
2	0.05	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
	0.01	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	0.05	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
	0.01	8.26	8.5	8.6	8.7	8.8	8.9	8.9	9.0	9.0	9.0	9.1	9.2	9.3	9.3
4	0.05	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
	0.01	6.51	6.8	6.9	7.0	7.1	7.1	7.2	7.2	7.3	7.3	7.4	7.4	7.5	7.5
5	0.05	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
	0.01	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.5	6.6	6.6	6.7	6.7	6.8
6	0.05	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
	0.01	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.0	6.0	6.1	6.2	6.2	6.3	6.3
7	0.05	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61
	0.01	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.8	5.8	5.9	5.9	6.0	6.0
8	0.05	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56
	0.01	4.74	5.0	5.14	5.23	5.32	5.40	5.47	5.51	5.5	5.6	5.7	5.7	5.8	5.8
9	0.05	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52
	0.01	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.4	5.5	5.5	5.6	5.7	5.7
10	0.05	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.48
	0.01	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.36	5.42	5.48	5.54	5.55
11	0.05	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	3.46	3.46	3.46	3.46	3.47	3.48
	0.01	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.24	5.28	5.34	5.38	5.39
12	0.05	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.46	3.46	3.46	3.47	3.48
	0.01	4.32	4.55	4.68	4.76	4.84	4.92	4.96	5.02	5.07	5.13	5.17	5.22	5.24	5.26
13	0.05	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.46	3.46	3.47	3.47	3.47
	0.01	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.04	5.08	5.13	5.14	5.15
14	0.05	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.45	3.46	3.46	3.47	3.47
	0.01	4.21	4.42	4.55	4.63	4.70	4.78	4.83	4.87	4.91	4.96	5.00	5.04	5.06	5.07
15	0.05	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.44	3.45	3.46	3.47	3.47
	0.01	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	4.90	4.94	4.97	4.99	5.00
16	0.05	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.44	3.45	3.46	3.47	3.47
	0.01	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79	4.84	4.88	4.91	4.93	4.94

V2 9	$\alpha$ 9	V1													
		2	3	4	5	6	7	8	9	10	12	14	16	18	20
17	0.05	2.90	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.44	3.45	3.46	3.47	3.47
	0.01	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.72	4.75	4.80	4.83	4.86	4.88	4.89
18	0.05	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.43	3.45	3.46	3.47	3.47
	0.01	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71	4.76	4.79	4.82	4.84	4.85
19	0.05	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.43	3.44	3.46	3.47	3.47
	0.01	4.05	4.24	4.25	4.43	4.50	4.56	4.61	4.64	4.67	4.72	4.76	4.79	4.81	4.82
20	0.05	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.43	3.44	3.46	3.46	3.47
	0.01	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.69	4.73	4.76	4.78	4.79
22	0.05	2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39	3.42	3.44	3.45	3.46	3.47
	0.01	3.99	4.17	4.28	4.36	4.42	4.48	4.53	4.57	4.60	4.65	4.68	4.71	4.74	4.75
24	0.05	2.92	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38	3.41	3.44	3.45	3.46	3.47
	0.01	3.96	4.14	4.24	4.33	4.39	4.44	4.49	4.53	4.57	4.62	4.64	4.67	4.70	4.72
26	0.05	2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38	3.41	3.43	3.45	3.46	3.47
	0.01	3.93	4.11	4.21	4.30	4.36	4.41	4.46	4.50	4.53	4.58	4.62	4.65	4.67	4.69
28	0.05	2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37	3.40	3.43	3.45	3.46	3.47
	0.01	3.91	4.08	4.18	4.28	4.34	4.39	4.43	4.47	4.51	4.56	4.60	4.62	4.65	4.67
30	0.05	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.40	3.43	3.44	3.46	3.47
	0.01	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.54	4.58	4.61	4.63	4.65
40	0.05	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.39	3.42	3.44	3.46	3.47
	0.01	3.82	3.99	4.10	4.17	4.24	4.30	4.34	4.37	4.41	4.46	4.51	4.54	4.57	4.59
60	0.05	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.37	3.40	3.43	3.45	3.47
	0.01	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.39	4.44	4.47	4.50	4.53
100	0.05	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.36	3.40	3.42	3.45	3.47
	0.01	3.71	3.86	3.98	4.06	4.11	4.17	4.21	4.25	4.29	4.35	4.38	4.42	4.45	4.48
4	0.05	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.34	3.38	3.41	3.44	3.47
	0.01	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.26	4.31	4.34	4.38	4.41



**Anexo 7: tabla de valores críticos de Dunnett**

		<b><math>\alpha= 0.05.</math> Prueba de dos colas</b>													
<b>v</b>	<b>p=</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>16</b>	<b>21</b>
5		2.57	3.03	3.29	3.48	3.62	3.73	3.82	3.90	3.97	4.03	4.09	4.14	4.26	4.42
6		2.45	2.86	3.10	3.26	3.39	3.49	3.57	3.64	3.71	3.76	3.81	3.86	3.97	4.11
7		2.36	2.75	2.97	3.12	3.24	3.33	3.41	3.47	3.53	3.58	3.63	3.67	3.78	3.91
8		2.31	2.67	2.88	3.02	3.13	3.22	3.29	3.35	3.41	3.46	3.50	3.54	3.64	3.76
9		2.26	2.61	2.81	2.95	3.05	3.14	3.20	3.26	3.32	3.36	3.40	3.44	3.53	3.65
10		2.23	2.57	2.76	2.89	2.99	3.07	3.14	3.19	3.24	3.29	3.33	3.36	3.45	3.57
11		2.20	2.53	2.72	2.84	2.94	3.02	3.08	3.14	3.19	3.23	3.27	3.30	3.39	3.50
12		2.18	2.50	2.68	2.81	2.90	2.98	3.04	3.09	3.14	3.18	3.22	3.25	3.34	3.45
13		2.16	2.48	2.65	2.78	2.87	2.94	3.00	3.06	3.10	3.14	3.18	3.21	3.29	3.40
14		2.14	2.46	2.63	2.75	2.84	2.91	2.97	3.02	3.07	3.11	3.14	3.18	3.26	3.36
15		2.13	2.44	2.61	2.73	2.82	2.89	2.95	3.00	3.04	3.08	3.12	3.15	3.23	3.33
16		2.12	2.42	2.59	2.71	2.80	2.87	2.92	2.97	3.02	3.06	3.09	3.12	3.20	3.30
17		2.11	2.41	2.58	2.69	2.78	2.85	2.90	2.95	3.00	3.03	3.07	3.10	3.18	3.27
18		2.10	2.40	2.56	2.68	2.76	2.83	2.89	2.94	2.98	3.01	3.05	3.08	3.16	3.25
19		2.09	2.39	2.55	2.66	2.75	2.81	2.87	2.92	2.96	3.00	3.03	3.06	3.14	3.23
20		2.09	2.38	2.54	2.65	2.73	2.80	2.86	2.90	2.95	2.98	3.02	3.05	3.12	3.22
24		2.06	2.35	2.51	2.61	2.70	2.76	2.81	2.86	2.90	2.94	2.97	3.00	3.07	3.16
30		2.04	2.32	2.47	2.58	2.66	2.72	2.77	2.82	2.86	2.89	2.92	2.95	3.02	3.11
40		2.02	2.29	2.44	2.54	2.62	2.68	2.73	2.77	2.81	2.85	2.87	2.90	2.97	3.06
60		2.00	2.27	2.41	2.51	2.58	2.64	2.69	2.73	2.77	2.80	2.83	2.86	2.92	3.00
120		1.98	2.24	2.38	2.47	2.55	2.60	2.65	2.69	2.73	2.76	2.79	2.81	2.87	2.95
□		1.96	2.21	2.35	2.44	2.51	2.57	2.61	2.65	2.69	2.72	2.74	2.77	2.83	2.91

		<b><math>\alpha= 0.01.</math> Prueba de dos colas</b>													
<b>v</b>	<b>p=</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>16</b>	<b>21</b>
5		4.03	4.63	4.98	5.22	5.41	5.56	5.69	5.80	5.89	5.98	6.05	6.12	6.30	6.52
6		3.71	4.21	4.51	4.71	4.87	5.00	5.10	5.20	5.28	5.35	5.41	5.47	5.62	5.81
7		3.50	3.95	4.21	4.39	4.53	4.64	4.74	4.82	4.89	4.95	5.01	5.06	5.19	5.36
8		3.36	3.77	4.00	4.17	4.29	4.40	4.48	4.56	4.62	4.68	4.73	4.78	4.90	5.05
9		3.25	3.63	3.85	4.01	4.12	4.22	4.30	4.37	4.43	4.48	4.53	4.57	4.68	4.82
10		3.17	3.53	3.74	3.88	3.99	4.08	4.16	4.22	4.28	4.33	4.37	4.42	4.52	4.65
11		3.11	3.45	3.65	3.79	3.89	3.98	4.05	4.11	4.16	4.21	4.25	4.29	4.38	4.52
12		3.05	3.39	3.58	3.71	3.81	3.89	3.96	4.02	4.07	4.12	4.16	4.19	4.29	4.41
13		3.01	3.33	3.52	3.65	3.74	3.82	3.89	3.94	3.99	4.04	4.08	4.11	4.20	4.32
14		2.98	3.29	3.47	3.59	3.69	3.76	3.83	3.88	3.93	3.97	4.01	4.05	4.13	4.24
15		2.95	3.25	3.43	3.55	3.64	3.71	3.78	3.83	3.88	3.92	3.95	3.99	4.07	4.18
16		2.92	3.22	3.39	3.51	3.60	3.67	3.73	3.78	3.83	3.87	3.91	3.94	4.02	4.13
17		2.90	3.19	3.36	3.47	3.56	3.63	3.69	3.74	3.79	3.83	3.86	3.90	3.98	4.08
18		2.88	3.17	3.33	3.44	3.53	3.60	3.66	3.71	3.75	3.79	3.83	3.86	3.94	4.04
19		2.86	3.15	3.31	3.42	3.50	3.57	3.63	3.68	3.72	3.76	3.79	3.83	3.90	4.00
20		2.85	3.13	3.29	3.40	3.48	3.55	3.60	3.65	3.69	3.73	3.77	3.80	3.87	3.97
24		2.80	3.07	3.22	3.32	3.40	3.47	3.52	3.57	3.61	3.64	3.68	3.70	3.78	3.87
30		2.75	3.01	3.15	3.25	3.33	3.39	3.44	3.49	3.52	3.56	3.59	3.62	3.69	3.78
40		2.70	2.95	3.09	3.19	3.26	3.32	3.37	3.41	3.44	3.48	3.51	3.53	3.60	3.68
60		2.66	2.90	3.03	3.12	3.19	3.25	3.29	3.33	3.37	3.40	3.42	3.45	3.51	3.59
120		2.62	2.85	2.97	3.06	3.12	3.18	3.22	3.26	3.29	3.32	3.35	3.37	3.43	3.51
□		2.58	2.79	2.92	3.00	3.06	3.11	3.15	3.19	3.22	3.25	3.27	3.29	3.35	3.42

<b><math>\alpha=0.05</math>. Prueba de una cola</b>										
<b>V</b>	<b>p=</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
5		2.02	2.44	2.68	2.85	2.98	3.08	3.16	3.24	3.30
6		1.94	2.34	2.56	2.71	2.83	2.92	3.00	3.07	3.12
7		1.89	2.27	2.48	2.62	2.73	2.82	2.89	2.95	3.01
8		1.86	2.22	2.42	2.55	2.66	2.74	2.81	2.87	2.92
9		1.83	2.18	2.37	2.50	2.60	2.68	2.75	2.81	2.86
10		1.81	2.15	2.34	2.47	2.56	2.64	2.70	2.76	2.81
11		1.80	2.13	2.31	2.44	2.53	2.60	2.67	2.72	2.77
12		1.78	2.11	2.29	2.41	2.50	2.58	2.64	2.69	2.74
13		1.77	2.09	2.27	2.39	2.48	2.55	2.61	2.66	2.71
14		1.76	2.08	2.25	2.37	2.46	2.53	2.59	2.64	2.69
15		1.75	2.07	2.24	2.36	2.44	2.51	2.57	2.62	2.67
16		1.75	2.06	2.23	2.34	2.43	2.50	2.56	2.61	2.65
17		1.74	2.05	2.22	2.33	2.42	2.49	2.54	2.59	2.64
18		1.73	2.04	2.21	2.32	2.41	2.48	2.53	2.58	2.62
19		1.73	2.03	2.20	2.31	2.40	2.47	2.52	2.57	2.61
20		1.72	2.03	2.19	2.30	2.39	2.46	2.51	2.56	2.60
24		1.71	2.01	2.17	2.28	2.36	2.43	2.48	2.53	2.57
30		1.70	1.99	2.15	2.25	2.33	2.40	2.45	2.50	2.54
40		1.68	1.97	2.13	2.23	2.31	2.37	2.42	2.47	2.51
60		1.67	1.95	2.10	2.21	2.28	2.35	2.39	2.44	2.48
120		1.66	1.93	2.08	2.18	2.26	2.32	2.37	2.41	2.45
∞		1.64	1.92	2.06	2.16	2.23	2.29	2.34	2.38	2.42

<b><math>\alpha=0.01</math>. Prueba de una cola</b>										
<b>V</b>	<b>p=</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
5		3.37	3.90	4.21	4.43	4.60	4.73	4.85	4.94	5.03
6		3.14	3.61	3.88	4.07	4.21	4.33	4.43	4.51	4.59
7		3.00	3.42	3.66	3.83	3.96	4.07	4.15	4.23	4.30
8		2.90	3.29	3.51	3.67	3.79	3.88	3.96	4.03	4.09
9		2.82	3.19	3.40	3.55	3.66	3.75	3.82	3.89	3.94
10		2.76	3.11	3.31	3.45	3.56	3.64	3.71	3.78	3.83
11		2.72	3.06	3.25	3.38	3.48	3.56	3.63	3.69	3.74
12		2.68	3.01	3.19	3.32	3.42	3.50	3.56	3.62	3.67
13		2.65	2.97	3.15	3.27	3.37	3.44	3.51	3.56	3.61
14		2.62	2.94	3.11	3.23	3.32	3.40	3.46	3.51	3.56
15		2.60	2.91	3.08	3.20	3.29	3.36	3.42	3.47	3.52
16		2.58	2.88	3.05	3.17	3.26	3.33	3.39	3.44	3.48
17		2.57	2.86	3.03	3.14	3.23	3.30	3.36	3.41	3.45
18		2.55	2.84	3.01	3.12	3.21	3.27	3.33	3.38	3.42
19		2.54	2.83	2.99	3.10	3.18	3.25	3.31	3.36	3.40
20		2.53	2.81	2.97	3.08	3.17	3.23	3.29	3.34	3.38
24		2.49	2.77	2.92	3.03	3.11	3.17	3.22	3.27	3.31
30		2.46	2.72	2.87	2.97	3.05	3.11	3.16	3.21	3.24
40		2.42	2.68	2.82	2.92	2.99	3.05	3.10	3.14	3.18

<b>60</b>		2.39	2.64	2.78	2.87	2.94	3.00	3.04	3.08	3.12
<b>120</b>		2.36	2.60	2.73	2.82	2.89	2.94	2.99	3.03	3.06
	□	2.33	2.56	2.68	2.77	2.84	2.89	2.93	2.97	3.00