UNIVERSIDAD DE EL SALVADOR

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

ESCUELA DE MATEMÁTICA

TESIS:

APLICACIÓN DEL ANÁLISIS DISCRIMINANTE PARA LA DETECCIÓN DE FACTORES DE RIESGO EN PACIENTES CON DIABETES MELLITUS EN LA REGIÓN DEL BAJO LEMPA DE EL SALVADOR

POR:

JAVIER ELISEO CHÁVEZ MANCIA CM05028

MARIO ISRAEL SANTOS NOLASCO SN01004

PARA OPTAR AL GRADO DE:

LICENCIADO EN ESTADÍSTICA

Ciudad Universitaria, Diciembre de 2015

UNIVERSIDAD DE EL SALVADOR

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

ESCUELA DE MATEMÁTICA

TESIS:

APLICACIÓN DEL ANÁLISIS DISCRIMINANTE PARA LA DETECCIÓN DE FACTORES DE RIESGO EN PACIENTES CON DIABETES MELLITUS EN LA REGIÓN DEL BAJO LEMPA DE EL SALVADOR

POR:

JAVIER ELISEO CHÁVEZ MANCIA CM05028

MARIO ISRAEL SANTOS NOLASCO SN01004

ASESORES:

DR. JOSÉ NERYS FUNES

LIC. MOISÉS DÍAZ

Ciudad Universitaria, Diciembre de 2015

AUTORIDADES

RECTOR INTERINO:

LIC. JOSÉ LUIS ARGUETA ANTILLÓN

SECRETARIA GENERAL:

DRA. ANA LETICIA ZAVALETA DE AMAYA

FISCAL GENERAL:

LIC. NORA BEATRIZ MELÉNDEZ

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

DECANO:

LIC. MAURICIO HERNÁN LOVO CÓRDOVA

SECRETARIA:

LIC. DAMARIS MELANY HERRERA TURCIOS

ESCUELA DE MATEMÁTICA

DIRECTOR:

DR. JOSÉ NERYS FUNES TORRES

SECRETARIA:

MSC. ALBA IDALIA CÓRDOVA CUÉLLAR

Ciudad Universitaria, Diciembre de 2015

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA

_____

DR. JOSÉ NERYS FUNES
ASESOR

_____

LIC. MOISÉS DÍAZ
ASESOR EXTERNO
MINISTERIO DE SALUD
INSTITUTO NACIONAL DE SALUD

Ciudad Universitaria, Diciembre de 2015

Agradecimientos

Gracias a Dios por habernos permitido lograr una meta más en nuestra vida, por la fuerza y sabiduría dada a cada uno de nosotros, de seguir adelante a pesar de los obstáculos de la vida.

Gracias a nuestras familias por su valioso apoyo moral y económico, que estuvieron presente en la evolución y en el desarrollo total de nuestra carrera.

Gracias a todos nuestros amigos y amigas por su valioso apoya que estuvieron en las buenas y en las malas, a los docentes que han sido parte fundamental de nuestro desarrollo académico profesional.

Gracias a nuestros asesores Dr. Nerys Funes y Lic. Moisés Díaz por habernos ayudado en la última etapa de nuestra carrera, por sus aportes en el desarrollo del trabajo de graduación.

Mario Santos y Javier Chávez

Contenido

Introducción.

El presente trabajo de investigación está referido a la enfermedad diabetes mellitus (DM) que actualmente es una de las enfermedades de mayor incidencia a nivel mundial. La Organización Mundial de la Salud (OMS) reconoce tres formas de diabetes mellitus: tipo 1, tipo 2 y diabetes gestacional (ocurre durante el embarazo), cada una con diferentes causas y con distinta incidencia. Para el año 2013, se estimó que alrededor de 382 millones de personas eran diabéticas en el mundo y que llegarán a 592 millones en 2035.

La diabetes es una de las enfermedades no transmisibles más comunes. Es la cuarta o quinta causa de muerte en la mayoría de los países de ingresos altos, y hay pruebas sustanciales de que es una epidemia en muchos países en vías de desarrollo.

La diabetes mellitus es una de las enfermedades con mayor incidencia en América Latina, y en particular El Salvador. Toda América Latina se encuentra en una etapa de transición epidemiológica, demográfica y nutricional, debido a que en su presentación intervienen múltiples factores de riesgos, destacándose entre ellos: factores hereditarios, obesidad, hipertensión arterial, colesterol elevado y la mala alimentación.

En El Salvador en 1993 un estudio reveló que la prevalencia de diabetes era casi del 8%, en correspondencia con esta información, en un estudio en el año 2007 en una población mayor de 20 años, se reportó un 9.7% de prevalencia de diabetes mellitus, es decir, una de cada 10 personas padecen de diabetes mellitus, reflejando un incremento en la prevalencia en los últimos años.

Esta investigación se desarrollará haciendo uso de una base de datos proporcionada por el Instituto Nacional de Salud (INS). Dichos datos están comprendidos en el periodo de Agosto hasta Diciembre del 2009; tomando como universo a la población mayor o igual a 18 años de edad del área del Bajo Lempa en el municipio de Jiquilisco, El Salvador.

Para el desarrollo de esta investigación fue necesaria la siguiente organización: En el Capítulo I, se describe el contenido teórico de la enfermedad diabetes mellitus se agrega información general: definiciones, factores de riesgo que la generan, tipos de DM, clasificación y características distintivas. En el Capítulo II, se incluye teoría sobre el modelo matemático–estadístico por medio del análisis discriminante. La forma para

estimar y validar el modelo. En el Capítulo III, se presentan las variables consideradas en el estudio, además de exponer la aplicación del modelo matemático del análisis discriminante, la validación del modelo encontrado y sus conclusiones de los principales resultados. Se pretende concluir con la estimación de los parámetros y el aporte que esta puedan tener para la predicción del modelo que discrimine entre las variables que más influyen al padecimiento de la Diabetes Mellitus.

Para poder realizar el análisis que se presenta en esta investigación se ha utilizado el paquete estadístico SPSS.

Antecedentes.

En los últimos años se han generado una serie de estudios acerca del impacto tanto económicos y como de salud, sobre la enfermedad crónica Diabetes Mellitus (DM) en diferentes países, tal es el caso de México, Estados Unidos, Chile, España, Holanda.

Según estudios del grupo Tamayo y Cols en Aragón (1999). En los países europeos la tasa de mortalidad debido a la DM oscila entre 7.9 y 32.2 por cada 100,000 habitantes. En Estados Unidos los pacientes con Diabetes diagnosticada antes de los 15 años tienen una tasa de mortalidad 11 veces superior a la población general. La mortalidad es 2−3 veces superior en pacientes en los que se diagnostica la Diabetes después de los 40 años. En la mayoría de los países desarrollados, la diabetes ocupa del 4° al 8° lugar entre las causas de defunción. En España la DM representa la 3ª causa en mujeres y la 7ª en hombres. La primera causa de muerte entre los pacientes diabéticos es el infarto de miocardio, que causa el 50−60% de las muertes de los diabéticos no insulinodependientes. La principal causa de defunción de los diabéticos insulinodependientes es la insuficiencia renal por nefropatía diabética.

La Diabetes mellitus fue la cuarta causa de muerte en América Latina y el Caribe en 2001, lo cual correspondió al 5% de las muertes totales. En México fue la causa principal de muerte en la población total en el 2002, causante del 12.8% de las muertes (causa principal entre las mujeres con 15.7% y la segunda entre los hombres, con 10.5%). La mayor tasa de mortalidad por diabetes le corresponde a México y en el Caribe-no Latino con 60 y 75 por 100,000 habitantes, respectivamente.

En un estudio realizado en el 2003 se calculó el costo de la alta prevalencia de la diabetes, representando para los países de América Latina y el Caribe una pérdida de 757,096 años de vida productiva en las personas menores de 65 años (>$ 3 billones). La incapacidad permanente secundaria a esta enfermedad causa una pérdida de 12, 699,087 años y más de $ 50 billones, y las incapacidades temporales representan una pérdida de 136,701 años de la población trabajadora y más de $ 763 millones. En cuanto a los costos relacionados con el tratamiento, la insulina y los medicamentos orales representan $ 4,720 millones, las hospitalizaciones $ 1,012 millones, las consultas

$ 2,508 millones y el cuidado de las complicaciones $ 2,480 millones. Se estimó que el costo anual asociado a la diabetes en América Latina y el Caribe es de $ 65,216 millones (directo $ 10,720; indirecto $ 54,496).

Un estudio realizado por la Federación Internacional de Diabetes estima las mayores prevalencias actuales y futuras: a partir de los datos correspondientes a 215 países, en el año 2007 habría 246 millones de personas con diabetes, lo cual superaría la predicción efectuada en 1994, que preveía 239 millones de diabéticos para el año 2010. En el mundo desarrollado la prevalencia rondaría el 6% de la población total y superaría el 7% de la población adulta. Los estudios de prevalencia en América Latina han sido esporádicos, difiriendo en variables metodológicas importantes (poblaciones estudiadas, edad, métodos de muestreo y criterios diagnósticos). Aun así, se podría concluir que la DM afecta a 6-8% de sus poblaciones adultas urbanas.

En 2011, 366 millones de personas tienen diabetes y hay otros 280 millones que corren un alto riesgo de desarrollarla. De no hacerse nada, el número de personas con diabetes aumentará. La pérdida de productividad laboral y el descenso de los índices de crecimiento económico. En el mundo, los gastos sanitarios por diabetes se han elevado a 465,000 millones de dólares en 2011, lo cual equivale al 11% del gasto sanitario total. Las pérdidas en ingresos nacionales debidas a muertes (en gran parte evitables) por diabetes, enfermedad cardiaca y derrame cerebral son enormes; entre 2005 y 2015, dichas pérdidas se calcula que alcanzarán los 558,000 millones de dólares en China, los 303,000 millones de dólares en Rusia y 237,000 millones de dólares en India.

La carga de la diabetes no sólo se refleja en el creciente número de personas con diabetes, sino también en el creciente número de muertes prematuras debidas a la diabetes. En 2013, aproximadamente la mitad de todas las muertes debidas a la diabetes en adultos fue en personas menores de 60 años, y en las regiones menos desarrolladas como África Subsahariana esa proporción llegó al 75%.

Los diez países con mayor número de personas con Diabetes Mellitus (DM) en el 2013 son China (98.4 millones), India (65.1 millones), Estados Unidos (24.4 millones), Brasil (11.9 millones), Federación de Rusia (10.9 millones), México (8.7 millones), Indonesia (8.5 millones), Alemania (7.6 millones), Egipto (7.5 millones), Japón (7.2 millones).

Tanto en términos humanos como financieros, la carga de la diabetes es enorme. Provoca 5.1 millones de muertes y ha representado unos 548,000 millones de dólares en gastos de salud (11% del gasto total en todo el mundo) en 2013. Dos regiones gastaron más en diabetes que el resto de las regiones juntas: América del Norte y Caribe, con unos 263,000 millones de dólares estimados, el equivalente a casi la mitad de los gastos de salud en diabetes del mundo, y Europa con 147,000 millones de dólares. A pesar de sus crecientes poblaciones con diabetes, el Sudeste Asiático y África dedican menos del 1% de su gasto total sanitario a la enfermedad.

En los últimos años, la prevalencia de diabetes mellitus tipo 2 (DM tipo 2) ha mostrado un rápido incremento en todo el mundo. Este aumento está asociado al desarrollo económico, el envejecimiento de la población, la creciente urbanización, los cambios en la dieta, la poca actividad física y los cambios en otros patrones de estilo de vida.

La diabetes se perfila en la actualidad como uno de los grandes retos para la salud pública, tanto en países desarrollados como en países de ingresos medios y bajos. De acuerdo con la Organización Mundial de la Salud (OMS), la diabetes afecta entre un 10 % y 15 % de la población adulta de América Latina y el Caribe. Numerosos estudios epidemiológicos han demostrado que la DM es uno de los principales factores de riesgo de los llamados modificables o potencialmente modificables para sufrir un infarto cerebral (IC), siendo el riesgo de ictus atribuido a la DM del 18 % en hombres y del 22% en mujeres. Cada 19 segundos, alguien es diagnosticado con diabetes, y la diabetes causa más muertes en un año que el Cáncer y el SIDA combinados.

En El Salvador no se reportan datos exactos sobre prevalencia de estas enfermedades ni de sus factores de riesgo, aunque se cuenta con información circunscrita a ciertos municipios o áreas urbanas y rurales. En el período comprendido entre 1997-2002 el Equipo Técnico Gerencia de Atención Integral al Adulto mayor, realizó un Perfil Epidemiológico de las enfermedades crónicas no transmisibles en El Salvador, en el cual se obtiene un total de casos de Diabetes en personas de 20 a 59 años en ambos sexos; de 7,672 en hombres y 24,674 mujeres. En el año 2003, la prevalencia de la diabetes mellitus en la Ciudad de Santa Tecla (Departamento de La Libertad, El Salvador) era 7.4%.

La prevalencia de diabetes en El Salvador según estudio de 2003. El promedio de edad fue de 39.9 (39.0- 40.8) años. El 57.3% fueron personas menores de 40 años y el 42.7 % fueron mayores de 40 años. En el grupo total, la prevalencia de Diabetes Mellitus fue de 9.7%.

Por último, señalar que a pesar que es relevante y necesario realizar estudios sobre el impacto económico y salud de la diabetes mellitus son escasos y los que se presentan al público salvadoreño sólo reportan la prevalencia de la enfermedad, a nivel gubernamental se le da muy poca importancia a la enfermedad crónica de la diabetes. La base de datos que se trabajará se construirá a partir de los datos ya registrados por el Instituto Nacional de Salud (INS), en la población del Bajo Lempa en el municipio de Jiquilisco, El Salvador.

Justificación.

La realización de este trabajo de investigación surge del interés por conocer los posibles factores de riesgos que contribuyen al padecimiento de la Diabetes Mellitus en la región del Bajo Lempa de El Salvador.

Actualmente las investigaciones realizadas a nivel nacional sobre DM son escasas, (más abajo se presentaran algunos estudios realizados) y los que se presentan al pueblo salvadoreño solo reportan la prevalencia de la enfermedad. Las Unidades de Salud no cuentan con programas para la atención estandarizada de la población Diabética como los tiene para Atención al Niño Sano, Atención Prenatal (APN), Atención Adolescente, ya que en la consulta médica solo se brinda tratamiento farmacológico y educación en salud a criterio del médico tratante. A nivel internacional se cuentan con estudios no sólo orientados a la prevalencia a nivel mundial sino también acerca de los efectos que esta enfermedad está tomando sobre los países en Latinoamérica y el Caribe.

Los estudios realizados a nivel nacional por organizaciones y trabajos de tesis, con respecto a la Diabetes Mellitus están enfocados al análisis descriptivo. Entre los cuales son:

- *Evaluación del conocimiento diabetológico de los médicos y la calidad de atención al paciente diabético en el SIBASI La Libertad 2005.*
- *Valoración del apoyo familiar y del conocimiento sobre la diabetes mellitus y su influencia en el control glicémico en pacientes diabéticos que consultan en las unidades de salud de Lolotique y Chinameca San Miguel, durante los meses de julio y agosto de 2007.*
- *Amputación de miembro inferior por pie diabético en pacientes con diabetes tipo 2, en hospital San Juan de Dios de Santa Ana, periodo 2007 -2009.*
- *Evaluación del proceso educativo brindado a los pacientes hipertensos y diabéticos que acuden a la consulta externa del Hospital Nacional "Santa Gertrudis" de San Vicente, periodo del 4 al 15 julio de 2011.*
- *Realizar seguimiento farmacoterapeutico a pacientes del club de diabéticos del hospital nacional "Dr. Juan José Fernández" Zacamil. Aplicando el Método Dader.*

En El Salvador se desconoce sobre la existencia de estudios sobre la aplicación del Análisis Discriminante para la detección de factores de riesgos en pacientes con Diabetes Mellitus. Debido a ello, se considera importante dicha aplicación, ya que la población está siendo afectada por dicha enfermedad y mediante esta técnica pudiera ser posible la identificación de causas que están provocando el aumento de la incidencia de la enfermedad y así mismo hacer propuesta con el fin de una reducción en las comunidades más afectadas. Los resultados de este estudio servirán en la toma de decisiones y con ello buscar medidas preventivas a tomar en cuenta para evitar el padecimiento de la diabetes.

Planteamiento del Problema.

La diabetes mellitus (DM) es una enfermedad endocrina y metabólica caracterizada por un déficit parcial o absoluto en la secreción de insulina, hormona segregada por las células beta del páncreas. Este déficit tiene múltiples y diversas consecuencias en el organismo, entre las que sobresale la tendencia a mantener los niveles de glucosa en sangre inapropiadamente elevados (hiperglucemia) y que puede estar debida a una resistencia a la acción de la insulina o una deficiente secreción; se asocia a lesiones a largo plazo en diversos órganos (ojos, riñones, nervios, vasos sanguíneos y corazón). La diabetes es un claro ejemplo de enfermedad metabólica cuyo control depende del comportamiento de la persona que la padece.

Históricamente, en el caso de la diabetes mellitus, enfermedad tan antigua como nuestra civilización, los hitos en su historia son numerosos, y muchos de importancia relevante para la ciencia. Las descripciones o investigaciones en torno a esta enfermedad a través del tiempo han sido realizadas en muchos casos por grandes hombres de ciencias que sus nombres han trascendido a la posteridad. En la época contemporánea los estudios sobre los múltiples aspectos de este mal han proporcionado que no pocos científicos alcancen renombre mundial y que incluso hayan merecido varios Premios Nobel.

En el año 2013 el número de personas que padecen de diabetes a nivel mundial es de 382 millones. Y según la encuesta realizada por la Asociación Salvadoreña de Diabéticos (ASADI) en el 2013, existe aproximadamente 800,000 personas en El Salvador con diabetes mellitus, con una prevalencia del 9.69% localizada en San Salvador; 12.5% en San Vicente y el 13.3% en San Francisco Gotera. Hasta la fecha se han producido 5.1 millones de muertes en el 2013, provocando un gasto de $ 548,000 millones en medicina curativa y no preventiva, ocupando el 11% del gasto total de salud en adultos.

La diabetes tipo 2 representa entre el 85% y el 95% del total de la diabetes en los países de ingresos altos, y puede representar un porcentaje aún mayor en los países de ingresos medios y bajos. La diabetes tipo 1, aunque menos común que la diabetes tipo 2, está aumentando cada año en los países ricos y pobres. En la mayoría de los países de ingresos altos, la mayor parte de la diabetes en niños y adolescentes es la diabetes tipo1.

La diabetes gestacional es común y, al igual que la obesidad y la diabetes tipo 2, está aumentando en todo el mundo.

La Diabetes Mellitus se ha asociado a una multiplicidad de condiciones como las ambientales o la edad, entre otras; por ejemplo la prevalencia de esta enfermedad aumenta primordialmente en grupos sociales que han mudado rápidamente del estilo de vida tradicional al moderno.

Región Geográfica del Estudio.

El río Lempa es el más largo en Centroamérica y desagua en el océano pacifico. Este cruza Guatemala, Honduras y El Salvador. En El Salvador, los principales ríos que fluyen a través de las ciudades desaguan en el Lempa, llevando consigo los desechos sólidos y líquidos de las industrias y los asentamientos urbanos y marginales. En el sur de El Salvador, a lo largo de las riberas del Lempa hasta su desembocadura se encuentran distribuidas comunidades pobladas por personas de escasos recursos económicos, que su principal trabajo es la agricultura. Esta región se conoce como el Bajo Lempa.

Los lugares donde se desarrolló la investigación, fueron tres comunidades rurales del área del Bajo Lempa en el municipio de Jiquilisco, El Salvador: Nueva Esperanza, Ciudad Romero y La Canoa. En dicho estudio la muestra que se tomó fue de 1215 personas (534 hombres, 681 mujeres).

Sobre los Datos para el Estudio.

La base de datos del estudio Nefrolempa incluye una gran variedad de variables. Para este estudio se dispone de una Base de Datos, que contiene 7 variables, una variable dependiente y seis variables independientes que son:

Variables Dependiente:

- Y = glucecat1 = Glucemia diagnóstica

Variables Independientes:

- X_1 = colesterl = Valores de Colesterol en la sangre
- X_2 = triglicérido = Triglicéridos

- X_3 = imc = Índice de masa corporal
- X_4 = tadiast = Tensión arterial diastólica
- X_5 = tsistolic = Tensión arterial sistólica
- X_6 = edad = Edad en años del entrevistado

Propósito de este Estudio.

Esta investigación tiene como finalidad aplicar el análisis discriminante para identificar, a partir de una serie de indicadores, si es posible "discriminar" si una observación pertenece a un determinado grupo de entre varios existentes; seleccionar cuál o cuáles de esos indicadores contribuyen más al proceso de discriminación, y adicionalmente permite estimar funciones de clasificación para ubicar nuevos casos. En resume se identificara cuáles de las variables independientes son más influyentes en la Diabetes Mellitus, en la región del Bajo Lempa de El Salvador.

Objetivos.

Objetivo General

Crear un modelo matemático-estadístico que evalúe las relaciones entre los factores de riesgo para la aparición de la Diabetes Mellitus, en la población del Bajo Lempa en el municipio de Jiquilisco, El Salvador.

Objetivos Específicos

- Identificar los factores de riesgo que inciden en el padecimiento de la Diabetes Mellitus.

- Proporcionar al Instituto Nacional de Salud (INS), los factores que inciden en la prevalencia de la DM, validados técnica y científicamente.

- Determinar la prevalencia de Diabetes Mellitus de acuerdo a su distribución de acuerdo a edad, sexo y ocupación en la región del Bajo Lempa, municipio de Jiquilisco.

CAPÍTULO I: DIABETES MELLITUS.

1.1 Definición de Diabetes Mellitus.

La diabetes es una enfermedad crónica que aparece cuando el cuerpo no puede producir suficiente insulina o no puede usar la insulina eficazmente.

La insulina es una hormona producida en el páncreas que permite que la glucosa de los alimentos entre en las células del cuerpo, donde se convierte en la energía necesaria para que funcionen los músculos y los tejidos.

La Diabetes Mellitus comprende un grupo de trastornos metabólicos frecuentes que comparten el fenotipo de la hiperglucemia. Existen varios tipos diferentes de DM resultado de una interacción compleja entre genética y factores ambientales. Se trata de una patología compleja que incluye a varias enfermedades en las cuales coexiste un trastorno global del metabolismo de los hidratos de carbono, grasas y proteínas. Este daño puede conducir a una discapacidad y a complicaciones de salud que pueden llegar a ser mortales.

1.2 Clasificación de la Diabetes Mellitus.

Los criterios de clasificación y diagnóstico de la diabetes mellitus elaborados por el National Diabetes Data Group y recomendados por la OMS, han sido revisados por el Comité de Expertos para el Diagnóstico y Clasificación de la Diabetes Mellitus de la Asociación Americana de Diabetes (ADA) con el objetivo de plantear una nueva clasificación, dejando de lado el criterio terapéutico y teniendo en cuenta la etiología de la enfermedad.

La Organización Mundial de la Salud (OMS) hasta la fecha (2015) clasifica la diabetes mellitus en las siguientes:

- Diabetes Mellitus Tipo 1 (DM-1)
- Diabetes Mellitus Tipo 2 (DM-2)
- Diabetes Mellitus Gestacional

1.2.1 Diabetes Mellitus Tipo 1:

La diabetes tipo 1 (anteriormente denominada diabetes insulinodependiente o juvenil), es causada por una reacción autoinmune, en la que el sistema de defensa del cuerpo ataca las células betas productoras de insulina en el páncreas. Como resultado, el cuerpo ya no puede producir la insulina que necesita.

La enfermedad puede afectar a personas de cualquier edad, pero generalmente se presenta en niños o adultos jóvenes. Las personas con este tipo de diabetes necesitan insulina todos los días para controlar los niveles de glucosa en sangre. Sin insulina, una persona con diabetes tipo 1 muere.

La diabetes tipo 1 (DM-1) suele desarrollarse repentinamente y puede producir síntomas tales como:

- Sed anormal y sequedad de boca
- Micción frecuente
- Falta de energía
- Cansancio extremo
- Hambre constante
- Pérdida repentina de peso
- Heridas de cicatrización lenta
- Infecciones recurrentes
- Visión borrosa

Las causas del desarrollo de la diabetes tipo 1, hay varios factores de riesgo importantes, entre ellos están:

- Herencia genética
- Factores ambientales
- Virus
- Sustancias tóxicas
- Afecciones Inmunológicas
- Resistencia a la insulina

1.2.2. Diabetes Mellitus Tipo 2:

La diabetes tipo 2 (llamada anteriormente diabetes no insulinodependiente o del adulto), es el tipo de diabetes más común. Por lo general ocurre en adultos, pero cada vez más aparece en niños y adolescentes. En la diabetes tipo 2, el cuerpo puede producir insulina, pero o bien esto no es suficiente o bien el cuerpo no puede responder a sus efectos, dando lugar a una acumulación de glucosa en sangre.

Las causas del desarrollo de la diabetes tipo 2, hay varios factores de riesgo importantes, entre ellos están:

- La obesidad
- La mala alimentación
- La inactividad física
- La edad avanzada
- Los antecedentes familiares de diabetes
- El grupo étnico
- La alta glucosa en sangre durante el embarazo que afecta al feto

El número de personas con diabetes tipo 2 está creciendo rápidamente en todo el mundo. Los síntomas pueden ser similares a los de la diabetes de tipo 1, pero a menudo menos intensos. En consecuencia, la enfermedad puede diagnosticarse sólo cuando ya tiene varios años de evolución y han aparecido complicaciones.

1.2.3 Diabetes Mellitus Gestacional:

Las mujeres que desarrollan una resistencia a la insulina y, por tanto, una alta glucosa en sangre durante el embarazo se dice que tienen diabetes gestacional (también conocida como diabetes mellitus gestacional o DMG). La diabetes gestacional tiende a ocurrir tarde en el embarazo, por lo general alrededor de la semana 24-28.

Una glucosa en sangre mal controlada durante el embarazo puede dar lugar a un bebé con un tamaño significativamente superior a la media (una condición conocida como la macrosomía fetal), lo que hace que un parto normal se convierta en difícil y de riesgo. El recién nacido correrá el riesgo de sufrir lesiones en los hombros y problemas respiratorios. Los bebés que nacen de madres con diabetes gestacional también tienen

un mayor riesgo de obesidad y diabetes tipo 2 en la adolescencia o en la edad adulta temprana.

La diabetes gestacional en las mujeres normalmente desaparece después del nacimiento. Sin embargo, las mujeres que han tenido diabetes gestacional tienen un mayor riesgo de desarrollar diabetes gestacional en embarazos posteriores y de desarrollar diabetes tipo 2 más adelante en la vida.

1.3 Criterio para el Diagnóstico de la Diabetes Mellitus.

Hay varias maneras de diagnosticar la diabetes. Por lo general es necesario repetir cada método una segunda vez para diagnosticar la diabetes. Se deben hacer las pruebas en un entorno médico (como el consultorio de su médico o un laboratorio). Si el médico determina que la persona tiene un nivel muy alto de glucosa en la sangre o síntomas clásicos de glucosa alta, además de una prueba positiva, quizá no sea necesario que el médico le haga una segunda prueba para diagnosticar la diabetes.

Los Criterios para el diagnóstico de diabetes son:

- A1C (Hemoglobina glicosilada)
- Glucosa plasmática en ayunas
- Prueba de tolerancia a la glucosa oral
- Prueba aleatoria (o casual) de glucosa plasmática

A1C: La prueba A1C mide su nivel promedio de glucosa en la sangre durante los últimos 2 o 3 meses. Las ventajas de recibir un diagnóstico de esta manera es que no tiene que ayunar ni beber nada.

- Se diagnostica diabetes cuando: A1C $\geq$ 6.5%

Glucosa plasmática en ayunas: Esta prueba generalmente se realiza a primera hora en la mañana, antes del desayuno, y mide su nivel de glucosa en la sangre cuando está en ayunas. Ayunar significa no comer ni beber nada (excepto agua) por lo menos 8 horas antes del examen.

- Se diagnostica diabetes cuando: Glucosa plasmática en ayunas $\geq$ 126 mg/dl.

Prueba de tolerancia a la glucosa oral: Esta es una prueba de dos horas que mide su nivel de glucosa en la sangre antes de beber una bebida dulce especial y 2 horas después de tomarla. Le indica a su médico cómo el cuerpo procesa la glucosa.

- Se diagnostica diabetes cuando: Glucosa en la sangre a las 2 horas ≥ 200 mg/dl.

Prueba aleatoria (o casual) de glucosa plasmática: Esta prueba es un análisis de sangre en cualquier momento del día cuando tiene síntomas de diabetes severa.

- Se diagnostica diabetes cuando: Glucosa en la sangre ≥ 200 mg/dl.

Las personas con niveles altos de glucosa en sangre que no la tienen tan alta como las personas con diabetes, se dice que tienen tolerancia anormal a la glucosa (comúnmente conocida como TAG) o alteración de la glucosa en ayunas (AGA). La TAG se define como altos niveles de glucosa en sangre después de comer; mientras que la AGA se define como alta glucosa en sangre después de un período de ayuno. También se utiliza el término "prediabetes" para describir la condición de estas personas, una "zona gris" entre los niveles normales de glucosa y la diabetes.

La prediabetes es un trastorno en que el nivel de la glucosa en la sangre es mayor de lo normal, pero no lo suficientemente alto como para que sea diabetes. Este trastorno significa que está en peligro de tener diabetes de tipo 2.

Resultados que indican a la persona que es pre-diabética:

- A1C de entre 5.7% – 6.4 %.
- Glucosa en la sangre en ayunas de entre 100 – 125 mg/dl.
- Glucosa en la sangre a las 2 horas de entre 140 mg/dl –199 mg/dl.

1.4 Tratamiento de la Diabetes Mellitus.

El tratamiento de la diabetes consiste en la reducción de la glucemia y de otros factores de riesgo conocidos que dañan los vasos sanguíneos. Para evitar las complicaciones también es importante dejar de fumar.

Los elementos principales del tratamiento de un paciente diabético son:

- Dieta.

- Ejercicio físico.
- Educación.
- Medicamentos administrativos vía oral.
- Administración de insulina.

La absoluta interacción entre estos cinco tipos de medidas hace que no pueda considerarse uno sin los otros.

Dieta: El control del tipo y cantidad de alimentos ingeridos es la base para todos los tratamientos de Diabetes Mellitus. Desde luego que es también indispensable para los no diabéticos, aun cuando comúnmente se ignora. Una gran parte de los pacientes de DM-2, degeneran su habilidad para la producción de insulina y una dieta adecuada puede facilitar la efectividad de la insulina natural. El programa de dieta para los pacientes con DM-1, es menos restrictivo, pero igualmente importante.

Ejercicio físico: El ejercicio físico es una de las maneras que ayudan a controlar y prevenir la diabetes. Mejora el efecto de las otras partes del tratamiento. Es muy importante porque no solamente mejora de manera generalizada la salud, también puede ayudar a reducir los requerimientos de insulina, al hacer ésta más efectiva, probablemente al mejorar el funcionamiento de los receptores para la insulina. La cantidad de actividad física necesaria se designa de manera individual para cada paciente.

Educación: Es apenas hasta hace unos cuantos años que se especifica la utilización de un sistema de información para el diabético establecido como tratamiento fundamental, no como parte de tratamiento inmediato. En este método se ofrecen pláticas para saber qué tipos de dietas utilizar, que y cuánta insulina administrarse, entre otros.

Medicamentos vía oral: Se administran agentes hipoglucémiantes (tabletas ingeridas que disminuyen los niveles de glucosa en la sangre.) Estos agentes pueden estimular la liberación de una mayor cantidad de insulina y ayudan a reducir la resistencia ante la insulina ya disponible. Se utilizan aproximadamente en un 30% - 40% en los pacientes diabéticos de los Estados Unidos. A Pesar de esto, cabe mencionar que no son siempre efectivos para cualquiera que padezca la enfermedad. Son eficaces únicamente cuando por sí, el páncreas no puede generar una buena producción de insulina.

Los medicamentos vía oral son recetados cuando la dieta y el ejercicio no son suficientes para controlar los niveles de azúcar en la sangre.

Existen tres tipos de pastillas:

- Pastillas que retrasan o bloquean la descomposición de almidones y algunos azúcares, por ejemplo: Acarbosa.
- Pastillas que estimulan la producción de mayor cantidad de insulina, por ejemplo: Glibenclamida y Tolbutamida.
- Pastillas que potencian el efecto de la insulina, por ejemplo: Metformina.

Administración de insulina: Si esta sustancia está presente inadecuadamente en el cuerpo (como en la DM-1) o si se necesita de una mayor cantidad de ésta como resultado de una mala alimentación (como en la DM-2), es indispensable administrar insulina. Los diabéticos deben de revisar sus niveles diarios de glucosa, para saber cuándo y cuánta insulina inyectarse. Desde hace varias décadas, se diseñó un aparato especial para la medición precisa de los niveles de glucosa. Éste consiste en una pequeña punción en el dedo pulgar del paciente, que origine el flujo sanguíneo, del cual se toma la muestra. Ésta se coloca en el medidor, que verifica si existe o no necesidad de suministrar insulina.

La Insulina es una hormona que controla el nivel de azúcar en la sangre. Existen distintos tipos de insulina que se diferencian por la rapidez con la que actúan y la duración de su efecto. Las personas con diabetes tipo 1, requieren insulina y algunas personas con diabetes tipo 2 también la necesitan.

Existen cuatro tipos de insulina:

- La insulina de acción rápida.
- La insulina regular o de acción breve.
- La insulina de acción intermedia.
- La insulina de acción prolongada.

La insulina de acción rápida: Comienza a surtir efecto 15 minutos después de la inyección, tiene su máximo efecto al cabo de una hora y es eficaz durante dos a cuatro

horas. Tipos: Insulina glulisina (Apidra), insulina lispro (Humalog) e insulina aspart (NovoLog).

La insulina regular o de acción breve: Generalmente llega al flujo sanguíneo 30 minutos después de la inyección, tiene su máximo efecto de dos a tres horas después de la inyección y es eficaz durante aproximadamente tres a seis horas. Tipos: Humulin R, Novolin R.

La insulina de acción intermedia: Generalmente llega al flujo sanguíneo aproximadamente dos a cuatro horas después de la inyección, tiene su máximo efecto de cuatro a doce horas después de la inyección y es eficaz durante aproximadamente doce a dieciocho horas. Tipos: NPH (Humulin N, Novolin N).

La insulina de acción prolongada: Generalmente llega a la sangre varias horas después de la inyección y tiende a mantener bajo el nivel de glucosa durante un periodo de 24 horas. Tipos: Insulina detemir (Levemir) e insulina glargina (Lantus).

La insulina previamente mezclada puede ser útil para las personas a las que les resulta difícil extraer insulina de dos frascos y leer las indicaciones y dosis correcta. También es útil para quienes tienen problemas de vista o destreza manual, y es conveniente para las personas en las que se ha estabilizado la diabetes con esta combinación.

1.5 Complicaciones de la Diabetes Mellitus.

Las personas con diabetes corren el riesgo de desarrollar una serie de problemas de salud que pueden provocar discapacidad o la muerte. Los constantemente altos niveles de glucosa en sangre pueden conducir a enfermedades graves que afectan al corazón y a los vasos sanguíneos, ojos, riñones y nervios. Las personas con diabetes también tienen un mayor riesgo de desarrollar infecciones. Si es diabético, es cinco veces más probable que tenga una cardiopatía o un accidente cerebro vascular que una persona sin diabetes.

En los hombres con el paso del tiempo, orinar en exceso y tener vasos sanguíneos dañados pueden hacer que los riñones no funcionen eficazmente. La diabetes también puede causar impotencia en el hombre. Sin embargo, esto puede tratarse con medicación.

Los problemas de flujo sanguíneo pueden causar ceguera, cataratas y retinopatía (daños en el fondo del ojo). El médico debe examinarle los ojos regularmente. Aproximadamente 1 de cada 10 personas con diabetes tienen ulceraciones en los pies, que pueden causar infecciones graves. Debe tener las uñas cortas y los pies limpios.

Las embarazadas diabéticas deberán controlarse detenidamente la dosis de azúcar e insulina de la sangre, ya que tienen mayor riesgo de aborto espontáneo o de que el bebé nazca muerto.

Las complicaciones de la diabetes se pueden agrupar en 3 categorías:

- Daño en los nervios (Neuropatía).
- Daño en los grandes vasos sanguíneos (Enfermedad macrovascular).
- Daño en los pequeños vasos sanguíneos (Enfermedad microvascular).

Daño en los nervios (Neuropatía): Las neuropatías diabéticas son un grupo de padecimientos en los nervios que pueden causar entumecimiento y en ocasiones dolor y debilidad en las manos, brazos, pies y piernas. La neuropatía también puede causar problemas en el sistema digestivo, corazón y en los órganos sexuales.

Alrededor del 50% de las personas con diabetes cursan con un grado de daño en los nervios, pero no todos tienen síntomas físicos. La neuropatía es más común en personas que han tenido diabetes por más de 25 años, que además tienen sobrepeso, un mal control de su azúcar en la sangre y presión arterial elevada. La neuropatía más común es la neuropatía periférica, que afecta a los brazos y las piernas, este tipo de daño en los nervios ocasiona adormecimiento y disminución de la sensibilidad en los pies. Lo anterior aumenta la probabilidad de sufrir heridas en los pies que no son tratadas a tiempo llegando a provocar las amputaciones.

Daño en los grandes vasos sanguíneos (Enfermedad macrovascular): Niveles elevados de glucosa en la sangre pueden provocar un endurecimiento de las arterias (aterosclerosis) que puede provocar un ataque cardiaco, infarto o mala circulación en los pies.

Las enfermedades cardiacas es una de las principales causas de muerte en nuestro país. Los adultos con diabetes tienen una probabilidad 4 veces mayor de padecer alguna enfermedad cardiaca que los pacientes que no padecen diabetes. Asimismo, la probabilidad de que se presente un infarto es de 4 veces más en pacientes con diabetes.

Daño en los pequeños vasos sanguíneos (enfermedad microvascular): Los niveles elevados de glucosa en la sangre pueden hacer más gruesas las paredes de los pequeños vasos sanguíneos, hace a la sangre más espesa y puede llegar a romper algún vaso sanguíneo. Todo lo anterior provoca que haya una disminución en la circulación de la sangre en la piel, brazos, piernas y pies. Por otra parte, también puede cambiar la circulación de la sangre en ojos y riñones. La reducción del flujo sanguíneo a las piernas puede ocasionar la aparición de manchas cafés en las piernas.

1.6 Síndrome Metabólico.

El síndrome metabólico es un grupo de cuadros de varias enfermedades o factores de riesgo en un mismo individuo que aumentan su probabilidad de padecer una enfermedad cardiovascular o diabetes mellitus tipo 2, enfermedad renal y problemas de circulación en las piernas.

Estos cuadros son:

- Hipertensión arterial.
- Glucosa (un tipo de azúcar) alta en la sangre.
- Niveles sanguíneos elevados de triglicéridos, un tipo de grasas.
- Bajo niveles sanguíneos de HDL, el colesterol bueno.
- Exceso de grasa alrededor de la cintura.

No todos los médicos están de acuerdo con la definición o la causa del síndrome metabólico. La causa puede ser resistencia a la insulina. La insulina es una hormona que produce su cuerpo para ayudar a convertir el azúcar proveniente de los alimentos en energía para el organismo. Si usted tiene resistencia a la insulina, se acumula un exceso de azúcar en la sangre, preparando el escenario para la aparición de la enfermedad.

CAPÍTULO II: ANÁLISIS DISCRIMINANTE.

2.1. Introducción del Análisis Discriminante.

El análisis discriminante se utiliza para clasificar a un grupo de individuos o unidades experimentales en dos o más poblaciones definidas de manera única. Cuando se va a efectuar una clasificación de unidades experimentales en una de varias categorías posibles por medio del análisis discriminante, debe tenerse una muestra de unidades experimentales de cada grupo posible de clasificación y posteriormente generar reglas para tal clasificación.

El análisis discriminante se parece mucho al de regresión y la única diferencia que existe entre estas dos técnicas multivariantes es que la variable dependiente es categórica en el análisis discriminante mientras que el análisis de regresión es continua. Esta diferencia tiene también implicaciones en cuanto a los objetivos que se persiguen al aplicar cada uno de estos métodos, ya que en el análisis de regresión lo que se persigue es predecir el valor de una variable llamada dependiente en base a un conjunto de variables llamadas predictoras, mientras que el análisis discriminante se desea predecir la pertenencia de una observación a una de los grupos posibles.

Al aplicar el AD debe tenerse presente los tipos de errores que podrían cometerse, los cuales se presentan a continuación:

- Clasificar una unidad experimental al grupo i cuando en realidad pertenece al grupo j.
- Clasificar una unidad experimental al grupo j cuando en realidad pertenece al grupo i.

El objetivo último del análisis discriminante es encontrar la combinación lineal de las variables independientes que mejor permite diferenciar (discriminar) a los grupos. Una vez encontrada esa combinación (la función discriminante) podrá ser utilizada para clasificar nuevos casos. Se trata de una técnica de análisis multivariante que es capaz de aprovechar las relaciones existentes entre una gran cantidad de variables independientes para maximizar la capacidad de discriminación.

El análisis discriminante es aplicable a muy diversas áreas de conocimiento. Se ha utilizado para distinguir grupos de sujetos patológicos y normales a partir de los resultados obtenidos en pruebas diagnósticas, como los parámetros hemodinámicos en el ámbito clínico médico o las pruebas psicodiagnósticas en el ámbito clínico psicológico. En el campo de los recursos humanos se aplica a la selección de personal para realizar un filtrado de los currículos previos a la entrevista personal. En banca se ha utilizado para atribuir riesgos crediticios y en las compañías aseguradoras para predecir la siniestralidad.

2.2. Modelo Matemático.

A partir de *G* grupos donde se asignan a una serie de individuos y de *p* variables medidas sobre ellos (x_1,…,x_p), se trata de obtener para cada sujetos una serie de puntuaciones que indican el grupo al que pertenecen (y_1,…,y_m), de modo que sean funciones lineal de x_1 ,…, x_p.

$$y_1 = w_{11}x_1 + \cdots + w_{1p}x_p + w_{10}$$

. . . . . .

$$y_m = w_{m1}x_1 + \cdots + w_{mp}x_p + w_{m0}$$

Donde $m = min(G-1, p)$, tales que discriminen o separen lo máximo posible a los *G* grupos. Estas combinaciones lineales de las *p* variables deben maximizar la varianza entre los grupos y minimizar la varianza dentro de los grupos.

2.3. Descomposición de la Varianza.

Se puede descomponer la variabilidad total de la muestra en variabilidad dentro de los grupos y entre los grupos.

Partimos de:

$$Cov\left(x_j, x_{j'}\right) = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{ij'} - \bar{x}_{j'})$$

Se puede considerar la media de la variable x_j en cada uno de los grupos $I_1, \ldots, I_G$, es decir,

$$\bar{x}_{kj} = \frac{1}{n_k}\sum_{i\in I_k} x_{ij}$$

Para $k = 1, \dots, G$.

De este modo, la media total de la variable x_j se puede expresar como función de las medias dentro de cada grupo. Así,

$$\sum_{i\in I_k} x_{ij} = n_k \bar{x}_{kj}$$

Entonces

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} = \frac{1}{n}\sum_{k=1}^{G}\sum_{i\in I_k} x_{ij} = \frac{1}{n}\sum_{k=1}^{G} n_k\, \bar{x}_{kj} = \sum_{k=1}^{G}\frac{n_k}{n}\bar{x}_{kj}$$

Así,

$$Cov\left(x_j, x_{j'}\right) = \frac{1}{n}\sum_{k=1}^{G}\sum_{i\in I_k}\left(x_{ij} - \bar{x}_j\right)\left(x_{ij'} - \bar{x}_{j'}\right)$$

Si en cada uno de los términos se sustituye:

$$\left(x_{ij} - \bar{x}_j\right) = \left(x_{ij} - \bar{x}_{kj}\right) + \left(\bar{x}_{kj} - \bar{x}_j\right)$$

$$\left(x_{ij'} - \bar{x}_{j'}\right) = \left(x_{ij'} - \bar{x}_{kj'}\right) + \left(\bar{x}_{kj'} - \bar{x}_{j'}\right)$$

Al hacer la simplificación se obtiene:

$$Cov\left(x_j, x_{j'}\right) = \frac{1}{n}\sum_{k=1}^{G}\sum_{i\in I_k}\left(x_{ij} - \bar{x}_{kj}\right)\left(x_{ij'} - \bar{x}_{kj'}\right) + \sum_{k=1}^{G}\frac{n_k}{n}\left(\bar{x}_{kj} - \bar{x}_j\right)\left(\bar{x}_{kj'} - \bar{x}_{j'}\right)$$

$$= w\left(x_j, x_{j'}\right) + f\left(x_j, x_{j'}\right)$$

Es decir, la covarianza total es igual a la covarianza dentro de grupos más la covarianza entre grupos. Si denominamos como $v\left(x_j, x_{j'}\right)$ a la covarianza total entre x_j y $x_{j'}$ (sin distinguir grupos), entonces lo anterior se puede expresar como:

$$v(x_j\ , x_{j'}) = \ w(x_j\ , x_{j'}) + f(x_j\ , x_{j'})$$

En notación matricial esto es equivalente a:

$$V = W + F$$

Dónde:

V= matriz de covarianzas total
W = matriz de covarianzas dentro grupos
F = matriz de covarianzas entre de grupos

2.4. Extracción de las Funciones Discriminantes.

La idea básica del Análisis Discriminante consiste en extraer a partir de $\boldsymbol{x_1}, \ldots, \boldsymbol{x_p}$ variables observadas en *G* grupos, *m* funciones $\boldsymbol{y_1}, \ldots, \boldsymbol{y_m}$ de forma:

$$y_i = \ w_{i1}x_1 + \cdots + \ w_{ip}x_p + \ w_{i0}$$

Donde $m = \min(\ G - 1, p)$, tales que $corr(y_i, y_j) = 0$ para todo $i \neq j$.

Si las variables $x_1, \ldots, x_p$ están tipificadas, entonces las funciones

$$y_i = \ w_{i1}x_1 + \cdots + \ w_{ip}x_p$$

Para $i = 1, \ldots, m$, se denominan funciones discriminantes canónicas.

Las funciones $y_1\ , \ldots, y_m$ se extraen de modo que

- y_1 sea la combinación lineal de $x_1\ , \ldots, x_p$ que proporciona la mayor discriminación posible entre los grupos.
- y_2 sea la combinación lineal de $x_1, \ldots, x_p$ que proporciona la mayor discriminación posible entre los grupos, después de y_1, tal que $Corr(y_1, y_2) = 0$.

En general, y_i es la combinación lineal de $x_1\ , \ldots, x_p$ que proporciona la mayor discriminación posible entre los grupos después de y_{i-1} y tal que $corr(y_i, y_j) = 0$ para $j = 1, \ldots, (i - 1)$.

2.5. Clasificación de los G Grupos.

2.5.1. Cálculo de la Función Discriminante.

El propósito del análisis discriminante consiste en aprovechar la información contenida en las variables independientes para crear una función Z combinación lineal de $X_1, \ldots, X_p$ capaz de diferenciar lo más posible a ambos grupos. La función discriminante de Fisher es de la forma:

$$Z_{jk} = a + w_1 X_{1k} + w_2 X_{2k} + \cdots + w_p X_{pk}$$

Donde:

Z_{jk} = Puntuación Z discriminante de la función discriminante j para el grupo k.

a = Constante.

w_i = Ponderación discriminante para la variable independiente i.

X_{ik} = Variable independiente i para el grupo k.

La puntuación discriminante representa la proyección de ese caso a lo largo del eje discriminante definido por la función. Se debe tener cuidado en que la función discriminante difiere de la función de clasificación, también conocida como la función discriminante lineal de Fisher. Las funciones de clasificación, una para cada grupo, pueden utilizarse al clasificar observaciones. En este método de clasificación, unos valores de la observación para las variables independientes se incluyen en las funciones de clasificación y se calcula una puntuación de clasificación para cada grupo para esa observación. La observación se clasifica entonces en el grupo con la mayor puntuación de clasificación. Utilizamos la función discriminante como el medio de clasificar porque ofrece una representación resumida y simple de cada función discriminante, simplificando el proceso de interpretación y la valoración de la contribución de las variables independientes.

2.5.2. Forma Matricial de la Función Discriminante.

La función discriminante en forma matricial:

$$\begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_N \end{pmatrix} = \begin{pmatrix} X_{11} & X_{21} & \cdots & X_{p1} \\ X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1N} & X_{2N} & \cdots & X_{pN} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{pmatrix}$$

Ahora expresando el modelo en función de las desviaciones a la media, resulta:

$$\begin{pmatrix} Z_1 & - & \overline{z}_1 \\ Z_2 & - & \overline{z}_2 \\ \vdots & \ddots & \vdots \\ Z_N & - & \overline{z}_N \end{pmatrix} = \begin{pmatrix} X_{11} & X_{21} & \cdots & X_{p1} \\ X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1N} & X_{2N} & \cdots & X_{pN} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{pmatrix}$$

Entonces función discriminante en diferencia es:

$$z = Xw$$

La variabilidad de la función discriminante (suma de cuadrados de las desviaciones de las variables discriminantes con respecto a su media) se expresa:

Suma de cuadrados explicada por esta función: $z'z = w'X'Xw$

Donde $X'X$ es una matriz simétrica que expresa las desviaciones cuadráticas con respecto a la media de las variables (suma de cuadrados total).

Se puede descomponer en suma de cuadrados entre grupos F y suma de cuadrados dentro de los grupos W:

$V = X'X$ (Matriz de suma de cuadrados y productos cruzados (varianzas-covarianzas) para el conjunto de observaciones).

$$V = X'X = F + W$$

Con lo cual,

$$z'z = w'X'Xw = w'(F + W)w = w'Fw + w'Ww$$

Los ejes discriminantes vienen dados por los vectores propios asociados a los valores propios de la matriz $(W^{-1}F)$ ordenados de mayor a menor. Las puntuaciones

discriminantes se corresponden con los valores obtenidos al proyectar cada punto del espacio G-dimensional de las variables originales sobre el eje discriminante.

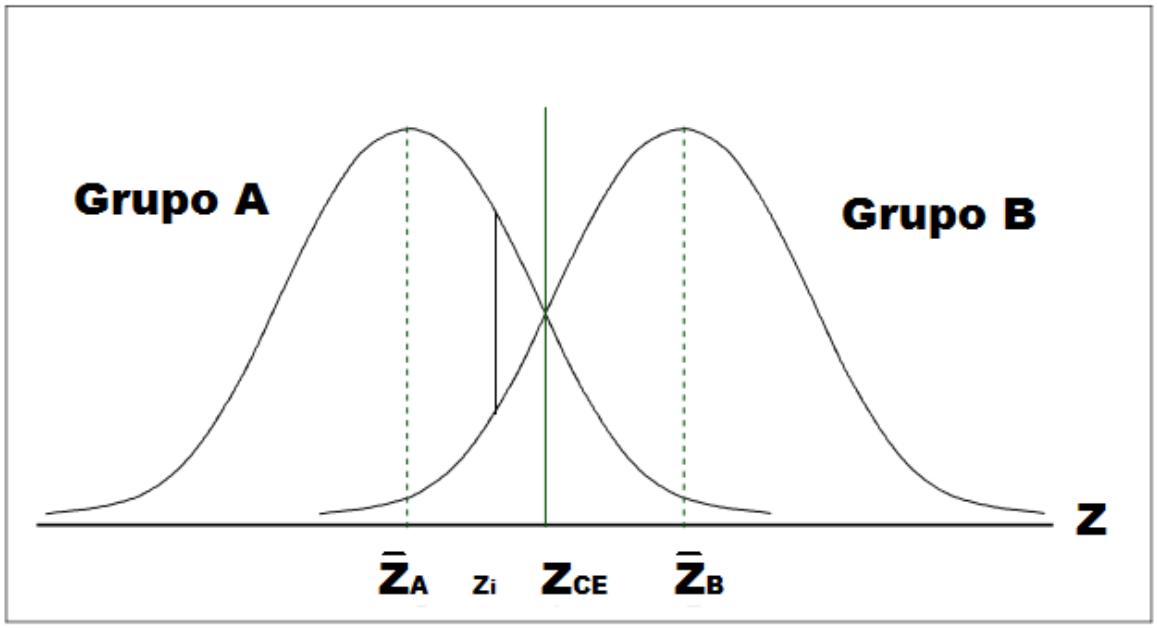
Los coeficientes *w* se obtienen:

$$Máx\ \lambda = \frac{w'Fw}{w'Ww} = \frac{\text{separación entre grupos}}{\text{separación dentro grupos}}$$

2.5.3. Criterio de Clasificación.

La determinación del punto de corte discriminante, antes de construir la matriz de clasificación el investigador debe determinar el punto de corte. El punto de corte discriminante es el criterio de clasificación al cual cada puntuación discriminante individual es comparada para determinar dentro de qué grupo debe ser clasificado cada individuo.

Figura 1. Histograma para clasificar cada observación en el grupo correcto, según el valor de la variable clasificadora.



En la figura 1 está representa sólo la función discriminante *Z*. Los grupos aparecen representados por sus histogramas y las proyecciones de los centroides $\overline{z}_A$ y $\overline{z}_B$ aparecen marcadas por una línea para cada grupo y el punto de corte discriminante Z_{CE} .

El punto de corte discriminante (Si los tamaños de grupo son iguales, la puntuación óptima es):

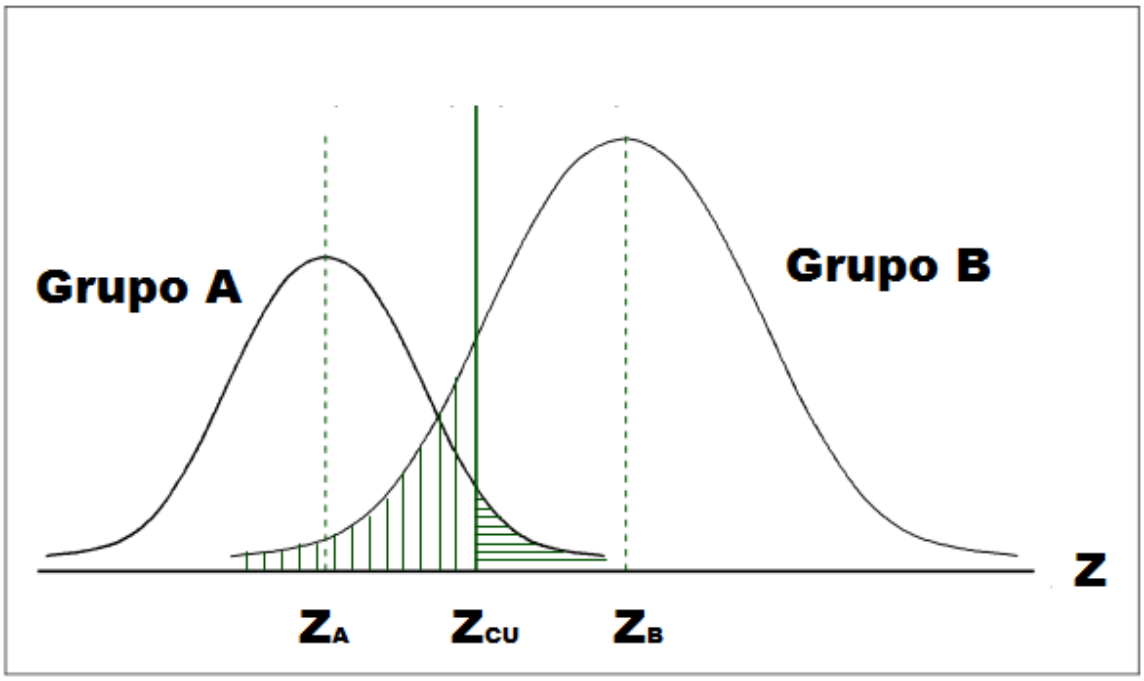
$$Z_{CE} = \frac{\overline{z}_A + \overline{z}_B}{2}$$

Se trata de minimizar los errores de clasificación:

- Si $z_i < Z_{CE}$ se clasifica en el grupo A.
- Si $z_i > Z_{CE}$ se clasifica en el grupo B.

Determinación del punto de corte para grupos de tamaño distintos. Si los grupos no son de igual tamaño y se supone que son representativos de las proporciones de la población, una media ponderada de los centroides de los grupos proporcionara una puntuación de corte óptima para una función discriminante.

Figura 2. Utilización de un punto de corte equidistante de ambos centroides ($N_A \neq N_B$)



En la Figura 2 puede verse con claridad que, si utilizamos el punto de corte Z_{CU} como punto de clasificación, la proporción de casos mal clasificados en el grupo de menor tamaño (zona rayada horizontalmente) será mucho menor que en el grupo de mayor tamaño (zona rayada verticalmente). Por tanto, con tamaños desiguales es preferible utilizar una regla de clasificación que desplace el punto de corte hacia el centroide del grupo de menor tamaño buscando igualar los errores de clasificación. Para calcular este punto de corte podemos utilizar una distancia ponderada:

$$Z_{CU} = \frac{N_A \overline{z}_A + N_B \overline{z}_B}{N_A + N_B}$$

Donde:

Z_{CU} = Valor de la puntuación de corte crítica para grupos de distinto tamaño.

N_A = Número del grupo A.

N_B = Número del grupo B.

$\overline{z}_A$ = Centroide del grupo A.

$\overline{z}_B$ = Centroide del grupo B.

2.5.4. Determinación del Criterio Basado en la Aleatoriedad.

Cuando los tamaños muéstrales son iguales, la determinación de la clasificación aleatoria es bastante simple; se obtiene dividiendo 1 por el número de grupos. La fórmula es C = 1/(número de grupos). Por ejemplo, en una función de dos grupos la probabilidad sería de 0.5; para una función de tres grupos la probabilidad sería de 0.33, y así sucesivamente.

Determinar la clasificación aleatoria es basarse en el tamaño muestral del grupo más grande, este criterio es conocido como el criterio de máxima aleatoriedad. Se determina calculando el porcentaje de la muestra completa representado por el más grande de los dos (o más) grupos. Por ejemplo, si los tamaños de los grupos son 65 y 35, el criterio de máxima aleatoriedad es el 65 por ciento de clasificaciones correctas. Por tanto, si la razón de aciertos por la función discriminante no excedió el 65%, entonces no nos ayudaría a predecir según este criterio. Este criterio debería utilizarse cuando el único objetivo del análisis discriminante es maximizar el porcentaje clasificado correctamente.

2.5.5. Medidas de Precisión Clasificatoria Fundamentadas Estadísticamente Relacionada con la Aleatoriedad.

Un contraste estadístico para contrastar la capacidad discriminatoria de la matriz de clasificación cuando se compara con un modelo de aleatoriedad es el estadístico Q de Press. Esta medida sencilla compara el número de clasificaciones correctas con el tamaño muestral total y el número de grupos. Se compara el valor hallado con un valor crítico (el valor de la chi-cuadrado para un grado de libertad al nivel de confianza

deseado). Si éste excede el valor crítico, la matriz de clasificación puede considerarse estadísticamente mejor que la aleatoriedad. El estadístico Q se calcula mediante la siguiente fórmula:

$$Q\ de\ Press = \frac{[N\ -(nG)]^2}{N(G-1)}$$

Donde:

$N =$ Tamaño muestral total.
$n =$ Número de observaciones correctamente clasificadas.
$G =$ Número de grupos.

2.5.6. Contrastes de Significación en el Análisis Discriminante.

En el análisis discriminante con G grupos, previamente, se deben realizar los siguientes contrastes:

a) Hipótesis de homocedasticidad.
b) Hipótesis de normalidad.
c) Hipótesis de diferencia entre las medias poblacionales de los G grupos.

La hipótesis de homocedasticidad asume que la matriz de covarianzas de los G grupos es constante igual a $\sum$.

La hipótesis de normalidad asume que cada uno de los grupos tiene distribución multivariante, es decir, $x_g \rightarrow N(\mu_g, \sum)$ para $g = 1,2, \dots, G$.

La respuesta que se da a la hipótesis (c) es decisiva para la justificación de la realización del análisis discriminante. En el caso de que la respuesta sea negativa carecería de interés continuar con el análisis, ya que significaría que las variables introducidas como variables clasificadoras no tienen capacidad discriminante significativa.

2.5.7. Contraste de Igualdad de Matrices de Varianzas Covarianzas.

Consideremos las poblaciones normales p-dimensionales $N(\mu_i, \sum_i)$ para $i = 1,2, \dots, G$. Se está interesado en realizar el siguiente contraste:

$$H_0 : \sum_1 = \sum_2 = \cdots = \sum_g$$

Este contraste se resuelve mediante la prueba de la razón de verosimilitud:

$$\lambda_R = \frac{|S_1|^{n_1/2} \times \cdots \times |S_G|^{n_G/2}}{|S_1|^{n/2}}$$

Donde S_i es la matriz de varianzas covarianzas de los datos de la población i, estimación máximo verosímil de $\sum_i$ y

$$n = n_1 + \cdots + n_G$$

$$S = \frac{1}{n}(n_1 S_1 + \cdots + n_G S_G) = \frac{W}{n}$$

Es la estimación máximo verosímil de $\sum$, matriz de covarianzas común bajo H_0. Se rechaza H_0 si el estadístico

$$-2\ln(\lambda_R) = n ln|S| - (n_1 ln|S_1| + \cdots + n_G ln|S_G|) \sim \chi_q^2$$

Es significativo, donde $q = G_P(p+1)/2 - p(p+1)/2 = (G-1)\,p(p+1)/2$ son los grados de libertad de la ji-cuadrado. Si se rechaza H_0, entonces resulta que no se dispone de unos ejes comunes para representar todas las poblaciones (la orientación de los ejes viene dada por la matriz de covarianzas).

Debido a que la prueba anterior puede ser sesgada, conviene aplicar la corrección en el estadístico de Barlett- Box,

$$c(n-G) ln|S| - \left((n_1 - 1) ln|\hat{S}_1| + (n_G - 1) ln|\hat{S}_G|\right)$$

Donde:

$$\hat{S}_i = \frac{n_i}{n_i - 1} S_i$$

y la constante c es

$$c = \left[1 - \left(\frac{2p^2 + 3p - 1}{6(p+1)(G-1)}\right)\left(\sum_{k=1}^{G} \frac{1}{n_G - 1} - \frac{1}{n-G}\right)\right]$$

2.5.8. Contraste de Igualdad de Varias Medias Multivariante.

Supóngase que se observa una muestra de tamaño *n* de una variable p dimensional que puede estratificarse en *G* grupos con n_g observaciones cada uno para $g = 1,2,\dots,G$. Un problema importante es contrastar que las medias de las *G* grupos son iguales. La hipótesis a contrastar es:

$$H_0 \ : \ \mu_1 = \mu_2 = \cdots = \mu_G \ = \mu$$

Donde, además, ∑ es la matriz de varianza covarianza, es definida positiva, e idéntica en los grupos. La hipótesis alternativa es:

$$H_1 : no\ todas\ las\ \mu\ son\ iguales$$

Con las mismas condiciones para ∑.

El test de la razón de verosimilitudes es:

$$\lambda = nl\mathrm{n}\left(\frac{|S|}{|S_w|}\right)$$

y

$$S = \frac{1}{n}\sum_{i=1}^{n}(x_i \ - \overline{x})(x_i \ - \overline{x})'$$

Donde:

$$x_i = \ (x_{i1}, x_{i2}, \dots, x_{ip})' \ \ y \ \ \overline{x} = (\overline{x}_1, \overline{x}_2, \dots, \overline{x}_p)'$$

$$S_w = \frac{1}{n}W$$

Donde:

$$W = \sum_{g=1}^{G}\sum_{h=1}^{H}\left(x_{hg} \ - \overline{x}_g\right)\left(x_{hg} \ - \overline{x}_g\right)'$$

La matriz *W* es conocida como la suma de cuadrados dentro de los grupos.

El estadístico λ tiene asintóticamente una distribución chi-cuadrada con g grados de libertad, donde $g = p(G-1)$.

Rechazamos H_0 a un nivel α si $\lambda > \chi^2_{\alpha,g}$.

2.5.9. Contrastes de Normalidad Multivariante.

Se Parte de que se tienen *n* vectores p-dimensionales, digamos, $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$ para *i = 1,2,…, n* y el objetivo es estudiar si esta muestra viene de una distribución normal p-dimensional.

Se parte de:

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

Donde:

$$\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)'$$

y la distancia de Mahalanobis d_{ij}^2 entre x_i y x_j es

$$d_{ij}^2 = (x_i - \bar{x})' S^{-1} (x_j - \bar{x})$$

Se define por A_p y K_p como el coeficiente de asimetría y curtosis multivariante respectivamente y,

$$A_p = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^3$$

$$K_p = \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2$$

Si los datos vienen de una normal multivariada, asintóticamente se verifica:

$$\frac{nA_p}{6} \sim \chi_f^2$$

$con\ f = \frac{1}{6}p(p+1)(p+2)$

$$K_p \sim N\left(p(p+2);\ \frac{8p(p+2)}{n}\right)$$

La potencia de esta prueba no es muy alta, a no ser que se tenga una muestra relativamente grande.

2.5.10. Método de Cálculo del Análisis Discriminante.

Se pueden utilizar dos métodos de cálculo para derivar una función discriminante: el método simultáneo (directo) y el método por etapas. La estimación simultánea implica el cálculo de la función discriminante donde todas las variables independientes son consideradas simultáneamente, sin considerar la capacidad discriminante de cada variable independiente. El método simultáneo es apropiado cuando, por razón teórica, el investigador quiere introducir todas las variables independientes en el análisis y no está interesado en observar resultados intermedios basados solamente en las variables que discriminan mejor.

La estimación por etapas es una alternativa al enfoque simultáneo. Incluye las variables independientes dentro de la función discriminante una a una, según su capacidad discriminatoria. El enfoque por etapas comienza eligiendo la variable que mejor discrimina. La variable inicial se empareja entonces con cada una de las variables independientes (una a una), y se elige la variable que más consigue incrementar la capacidad discriminante de la función en combinación con la primera variable. La tercera y posteriores variables se seleccionan de una manera similar. Mientras se incluyen variables adicionales, algunas variables seleccionadas previamente pueden ser eliminadas si la información que contienen sobre las diferencias del grupo está contenida en alguna combinación de otras variables incluidas en posteriores etapas. Al final, o bien todas las variables habrán sido incluidas en la función, o se habrá considerado que las variables excluidas no contribuyen significativamente a una mejor discriminación.

El método por etapas es útil cuando el investigador quiere considerar un número relativamente grande de variables independientes para incluir en la función. El conjunto

reducido es generalmente tan bueno como, y algunas veces mejor que, el conjunto completo de variables.

2.6. Aplicación del Teorema de Bayes.

La clasificación de los individuos se realiza utilizando el teorema de Bayes, que permite el cálculo de las probabilidades a posteriori a partir de estas probabilidades a priori y de la información muestral contenida en las puntuaciones discriminantes. En el caso general de G grupos, el teorema de Bayes establece que la probabilidad a posteriori de pertenencia a un grupo g con una puntuación discriminante Z, con probabilidades a priori π_g es:

$$Prob(g/Z) = \frac{\pi_g \, Prob(Z/g)}{\sum_{i=1}^{G} \pi_i \, Prob(Z/i)}$$

La probabilidad condicionada $Prob(Z/g)$ se obtiene calculando la probabilidad de la puntuación observada suponiendo la pertenencia a un grupo g.

Dado que el denominador $\sum_{i=1}^{G} \pi_i Prob(Z/i)$ es una constante, se utiliza también la forma equivalente:

$$Prob(g/Z) \propto \pi_g Prob(Z/g)\,; \text{ donde } \propto \equiv \text{proporcionalidad}$$

La clasificación de cada individuo, también puede realizar mediante la comparación de las probabilidades a posteriori. Así, se asignará un individuo al grupo para el cual sea mayor su probabilidad a posteriori. Se presenta el cálculo de probabilidades en el caso de dos grupos, de forma que sea fácilmente generalizable al caso de G grupos.

2.7. Análisis Discriminante en Poblaciones Desconocidas (Caso General).

Ahora se va a estudiar cómo aplicar la teoría anterior del análisis discriminante cuando en lugar de trabajar con poblaciones se dispone de muestras. Se aborda directamente el caso de *G* poblaciones posibles. Como caso particular, la discriminación clásica es para *G* = 2.

La matriz general de datos **X** de dimensiones $n \times p$, (n individuos y p variables), puede considerarse particionada ahora en *G* matrices correspondientes a las subpoblaciones.

Es posible llamar x_{ijg} a los elementos de estas submatrices, donde *i* representa el individuo, j la variable y *g* el grupo o submatriz.

Llámese n_g al número de elementos en el grupo g y el número total de observaciones es:

$$n = \sum_{g=1}^{G} n_g$$

Llámese $\boldsymbol{x'_{ij}}$ al vector fila $(1 \times p)$ que contiene los *p* valores de las variables para el individuo *i* en el grupo *g*, es decir, $\boldsymbol{x'_{ij}} = (x_{i1g}, \dots, x_{ipg})$. Entonces el vector de medias dentro de cada clase o subpoblación será:

$$\bar{X}_g = \frac{1}{n_g} \sum_{g=1}^{G} X_{ig}$$

y es un vector columna de dimensión *p* que contiene las *p* medias para las observaciones de la grupos *g*.

La matriz de varianzas y covarianzas para los elementos de los grupos g será:

$$\widehat{\boldsymbol{S}}_g = \frac{1}{n_g - 1} \sum_{i=1}^{n_g} (x_{ig} - \bar{x}_g)(x_{ig} - \bar{x}_g)'$$

Donde se ha dividido por $n_g - 1$ para tener estimaciones centradas de las varianzas y covarianzas.

Si se supone que las *G* subpoblaciones tienen la misma matriz de varianzas y covarianzas, su mejor estimación centrada con todos los datos será una combinación lineal de las estimaciones centradas de cada población con peso proporcional a su precisión. Por tanto:

$$\widehat{\boldsymbol{S}}_w = \sum_{g=1}^{G} \frac{n_g - 1}{n - G} \widehat{\boldsymbol{S}}_g$$

Y llamaremos **W** a la matriz de sumas de cuadrados dentro de los grupos que viene dada por:

$$W = (n - G)\widehat{\boldsymbol{S}}_w$$

Para obtener las funciones discriminantes se utiliza $\overline{\mathbf{X}}_{\mathrm{g}}$ como estimación de μ_g, y $\widehat{\boldsymbol{S}}_w$ como estimación de **V**.

En concreto, suponiendo iguales las probabilidades a priori y los costes de clasificación, se clasifica al elemento en el grupo que conduzca a un valor mínimo de la distancia de Mahalanobis entre el punto **x** y la media del grupo. Es decir, llamando $\widehat{\boldsymbol{w}}_g = \widehat{\boldsymbol{S}}_w^{-1}\overline{\mathbf{x}}_{\mathrm{g}}$ clasificaremos un nuevo elemento $\mathbf{x}_0$ en aquella población g donde:

$$\min_{\mathrm{g}}(\mathbf{x}_0 - \bar{\mathbf{x}}_{\mathrm{g}})'\hat{\mathbf{S}}_{\mathrm{w}}^{-1}(\mathbf{x}_0 - \bar{\mathbf{x}}_{\mathrm{g}}) = \min_{\mathrm{g}} \hat{\mathbf{w}}_{\mathrm{g}}'(\mathbf{x}_0 - \bar{\mathbf{x}}_{\mathrm{g}})$$

Es decir en el grupo *g*, en el que la distancia de Mahalanobis entre x_0 y $\overline{\mathrm{x}}_{\mathrm{g}}$ sea las más pequeña.

2.8. Discriminación Cuadrática. Discriminación de Poblaciones no Normales.

Si admitiendo la normalidad de las observaciones la hipótesis de igualdad de varianzas no fuese admisible, el procedimiento de resolver el problema es clasificar la observación en el grupo con máxima probabilidades a posteriori. Esto equivale a clasificar la observación $\mathbf{x}_0$ en el grupo donde se minimice la función:

$$\min_{\mathrm{j}\in(1,\ldots\ldots\mathrm{G})}\left[\frac{1}{2}\log|\mathbf{V}_{\mathrm{j}}| + \frac{1}{2}(\mathbf{x}_0 - \boldsymbol{\mu}_{\mathrm{j}})'\mathbf{V}_{\mathrm{j}}^{-1}(\mathbf{x}_0 - \boldsymbol{\mu}_{\mathrm{j}}) - \ln(C_j\pi_j)\right]$$

Cuando $\mathbf{V}_{\mathrm{j}}$ y $\boldsymbol{\mu}_{\mathrm{j}}$ son desconocidos se estiman por $\mathbf{S}_{\mathrm{w}}$ y $\bar{\mathbf{x}}_{\mathrm{j}}$ de la forma habitual. Ahora el término $\mathbf{x}_0'\mathbf{V}_{\mathrm{j}}^{-1}\mathbf{x}_0$ no puede anularse, al depender del grupo, y las funciones discriminantes no son lineales y tendrán un término de segundo grado. Suponiendo que los costes de clasificación son iguales en todos los grupos, clasificaremos nuevas observaciones con la regla:

$$\min_{\mathrm{j}\in(1,\ldots\ldots\mathrm{G})}\left[\frac{1}{2}\log|\widehat{\mathbf{V}}_{\mathrm{j}}| + \frac{1}{2}(\mathbf{x}_0 - \widehat{\boldsymbol{\mu}}_{\mathrm{j}})'\widehat{\mathbf{V}}_{\mathrm{j}}^{-1}(\mathbf{x}_0 - \widehat{\boldsymbol{\mu}}_{\mathrm{j}}) - \ln(\pi_{\mathrm{j}})\right]$$

En el caso particular de dos poblaciones y suponiendo las mismas probabilidades a priori clasificaremos una nueva observación en la población 2 si

$$\log\left|\widehat{\mathbf{V}}_1\right| + (\mathbf{x}_0 - \widehat{\boldsymbol{\mu}}_1)'\widehat{\mathbf{V}}_1^{-1}(\mathbf{x}_0 - \widehat{\boldsymbol{\mu}}_1) > \log\left|\widehat{\mathbf{V}}_2\right| + (\mathbf{x}_0 - \widehat{\boldsymbol{\mu}}_2)'\widehat{\mathbf{V}}_2^{-1}(\mathbf{x}_0 - \widehat{\boldsymbol{\mu}}_2)$$

Que equivale a

$$\mathbf{x}_0'(\widehat{\mathbf{V}}_1^{-1} - \widehat{\mathbf{V}}_2^{-1})\mathbf{x}_0 - 2\mathbf{x}_0'(\widehat{\mathbf{V}}_1^{-1}\widehat{\boldsymbol{\mu}}_1 - \widehat{\mathbf{V}}_2^{-1}\widehat{\boldsymbol{\mu}}_2) > c \quad (*)$$

Donde

$$c = \log(\left|\widehat{\mathbf{V}}_2\right| / \left|\widehat{\mathbf{V}}_1\right|) + \widehat{\boldsymbol{\mu}}_2'\,\widehat{\mathbf{V}}_2^{-1}\widehat{\boldsymbol{\mu}}_2 - \widehat{\boldsymbol{\mu}}_1'\,\widehat{\mathbf{V}}_1^{-1}\widehat{\boldsymbol{\mu}}_1$$

Llamando

$$\widehat{\mathbf{V}}_{\mathrm{d}}^{-1} = (\widehat{\mathbf{V}}_1^{-1} - \widehat{\mathbf{V}}_2^{-1})$$

Y

$$\widehat{\boldsymbol{\mu}}_{\mathrm{d}} = \widehat{\mathbf{V}}_{\mathrm{d}}^{-1}(\widehat{\mathbf{V}}_1^{-1}\widehat{\boldsymbol{\mu}}_1 - \widehat{\mathbf{V}}_2^{-1}\widehat{\boldsymbol{\mu}}_2)$$

Y definiendo las nuevas variables

$$\mathbf{z}_0 = \widehat{\mathbf{V}}_{\mathrm{d}}^{1/2}\mathbf{x}_0$$

Y llamando $\mathbf{z}_0 = (\mathbf{z}_{01}, \dots, \mathbf{z}_{0\mathrm{p}})'$ y definiendo el vector $\mathbf{m} = (\mathrm{m}_1, \dots \dots, \mathrm{m}_\mathrm{p})' = \widehat{\mathbf{V}}_{\mathrm{d}}^{1/2}(\widehat{\mathbf{V}}_1^{-1}\widehat{\boldsymbol{\mu}}_1 - \widehat{\mathbf{V}}_2^{-1}\widehat{\boldsymbol{\mu}}_2)$, la ecuación $(*)$ puede escribirse:

$$\sum_{\mathrm{i}=1}^{\mathrm{p}} \mathrm{z}_{0\mathrm{i}}^2 - 2\sum_{\mathrm{i}=1}^{\mathrm{p}} \mathrm{z}_{0\mathrm{i}}\,\mathrm{m}_\mathrm{i} > c$$

Esta es una ecuación de segundo grado en las nuevas variables z_{i0}. Las regiones resultantes con estas funciones de segundo grado son típicamente disjuntas y a veces difíciles de interpretar en varias dimensiones.

El número de parámetros a estimar en el caso cuadrático es mucho mayor que en el caso lineal. En el caso lineal hay que estimar *Gp + p(p + 1)/2* y en el caso cuadrático *G(p+p(p+1)/2)*. Por ejemplo con 10 variables y 4 grupos pasamos de estimar 95 parámetros en el caso lineal a 260 en el caso cuadrático. Este gran número de

parámetros hace que, salvo en el caso en que tenemos muestras muy grandes, la discriminación cuadrática sea bastante inestable y, aunque las matrices de covarianzas sean muy diferentes, se obtengan con frecuencia mejores resultados con la función lineal que con la cuadrática. Un problema adicional con la función discriminante cuadrática es que es muy sensible a desviaciones de la normalidad de los datos. La evidencia disponible indica que la clasificación lineal es en estos casos más robusta. Recomendamos siempre calcular los errores de clasificación con ambas reglas utilizando validación cruzada y en caso de que las diferencias sean muy pequeñas quedarse con la lineal.

CAPÍTULO III: APLICACIÓN DEL ANÁLISIS DISCRIMINANTE.

3.1 Introducción de la Aplicación del Análisis Discriminante.

En el presente capítulo tendrá los objetivos siguiente:

Objetivo General

Crear un modelo matemático-estadístico que evalúe las relaciones entre los factores de riesgo para la aparición de la Diabetes Mellitus, en la población del Bajo Lempa en el municipio de Jiquilisco, El Salvador.

Objetivos Específicos

- Identificar los factores de riesgo que inciden en el padecimiento de la Diabetes Mellitus.
- Proporcionar al Instituto Nacional de Salud (INS), los factores que inciden en la prevalencia de la DM, validados técnica y científicamente.
- Determinar la prevalencia de Diabetes Mellitus de acuerdo a su distribución por: edad, sexo y ocupación en la región del Bajo Lempa, municipio de Jiquilisco.

Sobre los datos para el Estudio.

De la base de datos del Estudio Nefrolempa se obtuvo una variedad de indicadores, y dentro de las cuales se encontraba la Diabetes Mellitus. Para este trabajo se utiliza la Base de Datos del Estudio Nefrolempa, en concreto, se analizan 7 variables, las cuales se detallan a continuación:

Variables Dependiente:

- Y = glucecat1 = Glucemia diagnóstica

Variables Independientes:

- X_1 = colesterl = Valores de Colesterol en la sangre
- X_2 = triglicérido = Triglicéridos
- X_3 = imc = Índice de masa corporal
- X_4 = tadiast = Tensión arterial diastólica

- X_5 = tsistolic = Tensión arterial sistólica
- X_6 = edad = Edad en años del entrevistado

En dicho estudio la muestra que se tomó fue de 1215 personas (534 hombres, 681 mujeres). El factor discriminante (o variable Dependiente "Y") tiene tres atributos: "Normal", "Prediabetes" y "Diabetes Mellitus".

3.2. Depuración de la Base de Datos.

Previo al análisis y obtención de resultado, se procedió a verificar la calidad de cada una de las variables a utilizar. En la base de datos se identificaron algunos valores inconsistentes, en las variables: colesterl (Valores de Colesterol en la sangre), imc (Índice de masa corporal), tadiast (Tensión arterial diastólica) y tsistolic (Tensión arterial sistólica). Por ejemplo los valores de colesterol en la sangre de una persona no pueden ser cero y en la base de datos aparecían dichos valores. En total, los valores inconsistentes fueron 80, los cuales se depuraron. Obteniendo una base de datos final de 1,135 personas, con la cual se trabajó y se corresponden los resultados de este trabajo.

3.3 Selección de una Muestra Aleatoria (Validar el Modelo).

Se realizó un muestreo aleatorio simple extrayendo el 20% en cada uno de los grupos: Personas Sanas, Prediabetes y Diabetes Mellitus de la base de datos depurada, utilizando el software SPSS se obtuvo dicha muestra, la cual se presenta en la siguiente tabla:

Tabla 1: Resultados muestreo aleatorio simple

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Normal	178	72.7	72.7	72.7
Prediabetes	51	20.8	20.8	93.5
Diabetes Mellitus	16	6.5	6.5	100.0
Total	245	100.0	100.0	

Por lo tanto, en la tabla 1 se puede observar que el tamaño de la muestra resultante es de 245 personas. Obteniendo una base de datos final con 890 personas, con la que se hará el estudio del análisis discriminante.

3.4 Prueba de los Supuestos del Análisis discriminante.

Test de Normalidad de las variables independientes.

En la Tabla 2 se presentan las pruebas no paramétricas de normalidad en el grupo 1 (personas sanas). Se observa que el nivel de significancia es mayor que 0.05 (Sig. asintót. (bilateral)) para todas las variables, y así no se puede rechazar la hipótesis de normalidad de las variables, lo que indica que la normalidad es razonable.

Tabla 2: Prueba de Kolmogorov-Smirnov para una muestra

		Colesterol	Triglicéridos	Edad en años	Indice de masa corporal	Tensión arterial diastólica	Tensión arterial sistólica
N		659	659	659	659	659	659
Parámetros normales[a,b]	Media	187.2516	140.9029	34.79	34.2577	69.39	108.6297
	Desviación típica	45.87918	88.21971	16.269	43.41197	11.239	17.92835
Diferencias más extremas	Absoluta	.058	.137	.151	.416	.120	.108
	Positiva	.058	.137	.131	.416	.120	.108
	Negativa	-.032	-.111	-.151	-.333	-.087	-.074
Z de Kolmogorov-Smirnov		1.498	3.511	3.878	10.671	3.083	2.782
Sig. asintót. (bilateral)		**.322**	**.210**	**.071**	**.123**	**.097**	**.239**

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

En la Tabla 3 se presentan las pruebas no paramétricas de normalidad en el grupo 2 (personas diagnosticadas con prediabetes). Se observa que el nivel de significancia es mayor que 0.05 (Sig. asintót. (bilateral)) para todas las variables, y así no se puede rechazar la hipótesis de normalidad de las variables, lo que indica que la normalidad es razonable.

Tabla 3: Prueba de Kolmogorov-Smirnov para una muestra

		Colesterol	Triglicéridos	Edad en años	Indice de masa corporal	Tensión arterial diastólica	Tensión arterial sistólica
N		178	178	178	178	178	178
Parámetros normales[a,b]	Media	270.7902	329.8427	43.87	28.1972	73.01	118.4017
	Desviación típica	46.94144	96.27660	17.715	17.95245	11.631	20.66309
Diferencias más extremas	Absoluta	.080	.131	.104	.288	.083	.143
	Positiva	.080	.131	.104	.288	.074	.143
	Negativa	-.055	-.066	-.072	-.272	-.083	-.080
Z de Kolmogorov-Smirnov		1.073	1.743	1.388	3.845	1.113	1.912
Sig. asintót. (bilateral)		**.400**	**.325**	**.056**	**.060**	**.168**	**.251**

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

En la Tabla 4 se presentan las pruebas no paramétricas de normalidad en el grupo 3 (personas diagnosticadas con diabetes mellitus). Se observa que el nivel de significancia es mayor que 0.05 (Sig. asintót. (bilateral)) para todas las variables, y así no se puede rechazar la hipótesis de normalidad de las variables, lo que indica que la normalidad es razonable.

Tabla 4: Prueba de Kolmogorov-Smirnov para una muestra

		Colesterol	Triglicéridos	Edad en años	Indice de masa corporal	Tensión arterial diastólica	Tensión arterial sistólica
N		53	53	53	53	53	53
Parámetros normales[a,b]	Media	366.5532	463.8491	52.38	27.8396	77.43	121.0943
	Desviación típica	25.82549	86.36403	12.971	5.28154	11.224	23.28586
Diferencias más extremas	Absoluta	.141	.267	.085	.058	.127	.072
	Positiva	.104	.267	.085	.058	.127	.066
	Negativa	-.141	-.094	-.064	-.049	-.081	-.072
Z de Kolmogorov-Smirnov		1.026	1.940	.618	.424	.921	.524
Sig. asintót. (bilateral)		.244	.311	.086	.200	.364	.523

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

En la tabla 5, se puede observar que los contrastes de igualdad de medias entre los grupos para cada variable (en los casos: Colesterol, Triglicéridos, Edad en años, Tensión arterial diastólica y Tensión arterial sistólica se rechaza la hipótesis nula, p-valor (Sig.) < 0.05, es decir, los grupos, en media son diferentes), y en el caso del Índice de masa corporal se acepta la hipótesis nula, p-valor (Sig.) > 0.05.

Tabla 5: Pruebas de igualdad de las medias de los grupos

	Lambda de Wilks	F	Grados de Libertad 1	Grados de Libertad 2	Sig.
Colesterol	.898	50.281	2	887	.000
Triglicéridos	.879	60.935	2	887	.000
Edad en años	.909	44.199	2	887	.000
Indice de masa corporal	**.995**	**2.206**	**2**	**887**	**.111**
Tensión arterial diastólica	.962	17.423	2	887	.000
Tensión arterial sistólica	.944	26.390	2	887	.000

En la tabla 6 se observa el contraste de igualdad de matrices de varianza-covarianza por medio de la prueba de M de Box para contrastar la hipótesis de homocedasticidad. Puede observarse que el valor p=0.10 (≥ 0.05) y por lo tanto, no se rechaza la hipótesis nula de igualdad de varianzas-covarianzas. Por tanto, concluir que las matrices de varianza-covarianza poblacionales correspondientes a cada grupo son iguales entre si.

Tabla 6: Resultados de la prueba

M de Box		92.972
F	Aprox.	7.641
	gl1	12
	gl2	94435.671
	Sig.	.10

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

3.5 Aplicación del Análisis Discriminante.

La tabla 7 ofrece un resumen con el total de casos procesados, el número de casos válidos para el análisis y el número de casos excluidos. En este caso, no se han excluidos casos para el análisis.

Tabla 7: Resumen del procesamiento para el análisis de casos

Casos no ponderados		N	Porcentaje
Válidos		890	100.0
Excluidos	Códigos de grupo para perdidos o fuera de rango	0	.0
	Perdida al menos una variable discriminante	0	.0
	Perdidos o fuera de rango ambos, el código de grupo y al menos una de las variables discriminantes.	0	.0
	Total excluidos	0	.0
Casos Totales		890	100.0

La tabla 8 ofrece un resumen del número de casos válidos en cada variable discriminante. La información de esta tabla posee un interés especial, pues un número desigual de casos en cada uno de los grupos puede afectar a la clasificación. En el estudio se puede observar que las personas con Diabetes Mellitus representan menos del 25% del total de personas analizadas.

También, se muestran los estadísticos descriptivos: media y desviación típica. Las medias de cinco variables introducidas como independientes en el análisis (Colesterol, Triglicéridos, Edad, Tensión arterial diastólica y sistólica) son mayores en el grupo de Diabetes Mellitus que en relación con los otros dos grupos (Normal, Prediabetes). En el caso de la variable Índice de masa corporal no se cumple esta observación.

Tabla 8: Estadísticos de grupo

Glucemia diagnóstica		Media	Desv. típ.	N válido (según lista)	
				No ponderados	Ponderados
Normal	Colesterol	187.2516	45.87918	659	659
	Triglicéridos	140.9029	88.21971	659	659
	Edad en años	34.7891	16.26937	659	659
	Índice de masa corporal	34.2577	43.41197	659	659
	Tensión arterial diastólica	69.3945	11.23860	659	659
	Tensión arterial sistólica	108.6297	17.92835	659	659
Prediabetes	Colesterol	270.7902	46.94144	178	17
	Triglicéridos	329.8427	96.27660	178	178
	Edad en años	43.8652	17.71538	178	178
	Índice de masa corporal	28.1972	17.95245	178	178
	Tensión arterial diastólica	73.0056	11.63061	178	178
	Tensión arterial sistólica	118.4017	20.66309	178	178
Diabetes Mellitus	Colesterol	366.5532	25.82549	53	53
	Triglicéridos	463.8491	86.36403	53	53
	Edad en años	52.3774	12.97072	53	53
	Índice de masa corporal	27.8396	5.28154	53	53
	Tensión arterial diastólica	77.4340	11.22413	53	53
	Tensión arterial sistólica	121.0943	23.28586	53	53
Total	Colesterol	214.6368	67.81482	890	890
	Triglicéridos	197.9225	134.73222	890	890
	Edad en años	37.6517	17.17475	890	890
	Indice de masa corporal	32.6634	38.31404	890	890
	Tensión arterial diastólica	70.5955	11.52424	890	890
	Tensión arterial sistólica	111.3264	19.38383	890	890

La tabla de variables introducidas/excluidas (Tabla 9) muestra un resumen de todos los pasos llevados a cabo en la construcción de la función discriminante y recuerda los criterios utilizados en la selección de variables. En cada paso se informa de la variable que ha sido incorporada al modelo y, en su caso, de la variable o variables que han sido expulsadas. En el presente estudio, todos los pasos llevados a cabo han sido de incorporación de variables: en el primer paso se incorpora la variable Colesterol; en el segundo se incorpora la variable Triglicéridos; en el tercero se incorpora la variable Tensión arterial sistólica. En ninguno de los 3 pasos ha habido expulsión de variables. Si alguna de las variables previamente incorporadas hubiera sido expulsada en algún paso posterior, la tabla mostraría una columna adicional indicando tal circunstancia.

Las notas a pie de tabla recuerdan algunas de las opciones establecidas para el análisis: la selección de variables se ha llevado a cabo utilizando el método de la distancia de Mahalanobis, el número máximo de pasos permitidos es 12, el valor del estadístico *F* para incorporar variables es 3.84 (criterio de entrada), el valor del estadístico *F* para excluir variables es 2.71 (criterio de salida) y, por último, en la nota *d* se informa que se ha alcanzado alguno de los criterios de parada (los niveles del estadístico *F*, el criterio de tolerancia y la *V* mínima de Rao), por lo que alguna de las variables independientes inicialmente propuestas no ha sido incluida en el modelo final.

Tabla 9: Variables introducidas/excluidas[a,b,c,d]

Paso		Mín. D cuadrado					
				F exacta			
	Introducidas	Estadístico	Entre grupos	Estadístico	gl1	gl2	Sig.
1	Colesterol	3.420	Normal y Prediabetes	479.364	1	887	0
2	Triglicéridos	4.986	Prediabetes y Diabetes Mellitus	101.689	2	886	0
3	Tensión arterial sistólica	4.986	Prediabetes y Diabetes Mellitus	67.728	3	885	0

En cada paso se introduce la variable que maximiza la distancia de Mahalanobis entre los grupos más cercanos.

a. El número máximo de pasos es 12.

b. La F parcial mínima para entrar es 3.84.

c. La F parcial máxima para salir es 2.71

d. El nivel de F, la tolerancia o el VIN son insuficientes para continuar los cálculos.

.La tabla 10 muestra el estadístico lambda de Wilks global para el modelo generado en cada paso, independientemente de que se haya optado por otro estadístico como método de selección de variables. Según se sabe, este estadístico permite valorar el grado de diferenciación entre los grupos tomando como referencia las variables independientes incluidas en cada paso. También se puede observar que a medida se van incorporando nuevas variables al modelo en cada paso, los valores de la lambda de Wilks global y del estadístico *F* asociado a ella van disminuyendo.

Tabla 10: Lambda de Wilks

Paso	Número de variables	Lambda	gl1	gl2	gl3	F exacta			
						Estadístico	gl1	gl2	Sig.
1	1	.443	1	2	887	558.418	2	887	.000
2	2	.353	2	2	887	303.033	4	1772	.000
3	3	.349	3	2	887	204.593	6	1770	.000

Los autovalores o valores propios asociados con la primera y segunda funciones discriminantes canónicas, se muestran en el tabla 11. Son dos funciones discriminantes debido a que el min(3-1,6)=2. El valor propio asociado a la primera función es 1.810 y explica el 98.9% de la varianza en los datos. Este primer valor propio es muy alto, por lo tanto, la función asociada aporta mucha información y discriminará muy bien. Con respecto a las correlaciones canónicas ambas con valores grandes, cercanos a uno, esto quiere decir que la segunda función discrimina bien, a pesar que solo explique el 1.1% de la varianza.

Tabla 11: Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	1.810[a]	98.9	98.9	.803
2	.021[a]	1.1	100.0	.143

a. Se han empleado las 2 primeras funciones discriminantes canónicas en el análisis.

En la tabla 12 se presentan los coeficientes de las funciones discriminantes (coeficientes no estandarizados), con los que se construyen las funciones discriminantes canónicas. Con estas funciones se obtendrán las puntuaciones discriminantes para cada individuo.

Tabla 12: Coeficientes de las funciones canónicas discriminantes

	Función	
	1	2
Colesterol	.013	-.020
Triglicéridos	.007	.009
Tensión arterial sistólica	.006	.022
(Constante)	-4.734	-.006

Coeficientes no tipificados

Ahora, se determina si las funciones generadas son estadísticamente significativas, es decir si el conjunto de las funciones permite que las medias de los grupos estén separadas. En la tabla lambda de Wilks (Tabla 13) se observa que el valor de lambda de Wilks para el conjunto formado por las dos funciones es 0.349; el segundo valor obtenido cuando se elimina la primera función es 0.980, y corresponde a la segunda función. Ambos valores son cercanos a uno (1) por lo que se concluye que ambas funciones tienen un alto poder discriminatorio. Esto se corrobora al ver que los p-valores asociados al estadístico chi-cuadrado son menores que 0.05 (columna "sig") por lo que se rechaza las hipótesis nulas de $\lambda_1 = \lambda_2 = 0$ y $\lambda_2 = 0$, y concluimos que ambas funciones son significativas.

Tabla 13: Lambda de Wilks

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a la 2	.349	933.523	6	0
2	.980	18.224	2	0

En la matriz de estructura (Tabla 14) se observan las correlaciones intra-grupo combinadas entre las variables discriminantes y cada una de las 2 funciones. El coeficiente más alto de cada variable aparece marcado con un asterisco que indica cuál es la función con la que más correlaciona esa variable (lo que no significa que sea ésa la función en la que más discrimina la variable). Si existe alta colinealidad (alta relación entre las variables independientes), los coeficientes de esta tabla puede ser muy distintos de los coeficientes estandarizados, como de hecho sucede. La variable Colesterol y Triglicéridos, es la que tiene la mayor correlación con la función 1. En la función 2, la variable Tensión arterial sistólica es la que tiene la correlación más alta con esta función.

Tabla 14: Matriz de estructura

	Función	
	1	2
Colesterol	.832*	-.535
Triglicéridos	.832*	.459
Edad en años[a]	.233*	.024
Indice de masa corporal[a]	-.049*	.005
Tensión arterial sistólica	.177	.374*
Tensión arterial diastólica[a]	.171	.270*

*. Mayor correlación absoluta entre cada variable y cualquier función discriminante.

a. Esta variable no se emplea en el análisis.

La tabla 15 muestra la ubicación de los centroides en cada una de las funciones discriminantes. La primera función distingue fundamentalmente a las personas diagnosticadas con Prediabetes y Diabetes Mellitus (cuyos centroides está ubicado en la parte positiva), y las personas sanas (cuyo centroide se encuentran en la parte negativa). En la segunda función, el centroide de las personas diagnosticadas con Prediabetes se sitúa en la parte positiva, mientras que el de las personas sanas se sitúa en la parte negativa; el de las personas diagnosticadas con Diabetes Mellitus queda en la parte central. Dado que la primera función ha conseguido explicar el máximo de las diferencias existentes entre las personas diagnosticadas con DM y el resto, es lógico que la segunda función discrimine precisamente entre los dos grupos que han quedado más próximos en la primera.

Tabla 15: Funciones en los centroides de los grupos

Glucemia diagnóstica	Función	
	1	2
Normal	-.747	-.029
Prediabetes	1.640	.228
Diabetes Mellitus	3.782	-.403

Funciones discriminantes canónicas no tipificadas evaluadas en las medias de los grupos

Hasta aquí se ha discutido el proceso de construcción o estimación del modelo. Para valorar la capacidad predictiva del modelo estimado se debe prestar atención a los resultados de la clasificación.

La tabla 16 ofrece las probabilidades a priori. Estas probabilidades indican que se ha dado la misma importancia relativa a todos los grupos: 0.333 (a pesar de que las personas Normal constituyen más del 70% de la muestra).

Tabla 16: Probabilidades a priori iguales para los grupos

Glucemia diagnóstica	Previas	Casos utilizados en el análisis	
		No ponderados	Ponderados
Normal	.333	659	659
Prediabetes	.333	178	178
Diabetes Mellitus	.333	53	53
Total	1.000	890	890

La tabla 17 ofrece las probabilidades a priori. Estas probabilidades están basadas en los tamaños de los grupos. Enseguida veremos qué ocurre si utilizamos probabilidades previas basadas en los tamaños de los grupos.

Tabla 17: Probabilidades a priori según el tamaño de los grupos

Glucemia diagnóstica	Previas	Casos utilizados en el análisis	
		No ponderados	Ponderados
Normal	.740	659	659
Prediabetes	.200	178	178
Diabetes Mellitus	.060	53	53
Total	1.000	890	890

Función de clasificación asumiendo probabilidades iguales para los grupos:

Las funciones de clasificación lineales de Fisher se muestran en la tabla 18.

Tabla 18: Coeficientes de la función de clasificación (probabilidades a priori iguales).

	Glucemia diagnóstica		
	Normal	Prediabetes	Diabetes Mellitus
Colesterol	.083	.109	.149
Triglicéridos	-.001	.017	.025
Tensión arterial sistólica	.290	.309	.308
(Constante)	-24.522	-36.918	-52.917

Funciones discriminantes lineales de Fisher

De esta forma, las funciones de clasificación son:

Función de clasificación para personas sanas:

$$Z_1 = -24.522 + 0.083(Colesterol) - 0.001(Triglicéridos) + 0.290(TAS)$$

Función de clasificación para personas diagnosticadas con Prediabetes:

$$Z_2 = -36.918 + 0.109(Colesterol) + 0.017(Triglicéridos) + 0.309(TAS)$$

Función de clasificación para personas diagnosticadas con DM:

$$Z_3 = -52.917 + 0.149(Colesterol) + 0.025(Triglicéridos) + 0.308(TAS)$$

3.5.1 Función Discriminante asumiendo probabilidades a priori proporcionales al tamaño de la muestra:

Las probabilidades previas o probabilidades a priori se basan en el conocimiento de la población o el tamaño de los grupos. La probabilidad a priori que se asigna a cada grupo es proporcional a su tamaño. Siendo N el tamaño de la muestra y n_g el tamaño de un grupo cualquiera, la probabilidad a priori asignada a ese grupo es n_g/N. Con esta opción, si un caso posee una puntuación discriminante equidistante de los centroides de dos grupos, el caso es clasificado en el grupo de mayor tamaño. Mediante sintaxis, es posible asignar a cada grupo probabilidades a priori personalizadas.

Los coeficientes propuestos por Fisher se utilizan únicamente para la clasificación. Por medio de esta opción se obtiene una función de clasificación para cada uno de los grupos: Personas sanas, prediabetes y diabetes mellitus.

Las funciones de clasificación lineales de Fisher se muestran en la siguiente tabla 19.

Tabla 19: Coeficientes de la función de clasificación (probabilidades previas basada en los tamaños de los grupos).

	Glucemia diagnóstica		
	Normal	Prediabetes	Diabetes Mellitus
Colesterol	.083	.109	.149
Triglicéridos	-.001	.017	.025
Tensión arterial sistólica	.290	.309	.308
(Constante)	-23.724	-37.429	-54.639

Funciones discriminantes lineales de Fisher

De esta forma, las funciones de clasificación para probabilidades previas basada en los tamaños de los grupos son:

Función de clasificación para personas sanas:

$$Z_1 = -23.724 + 0.083(Colesterol) - 0.001(Triglicéridos) + 0.290(TAS)$$

Función de clasificación para personas diagnosticadas con Prediabetes:

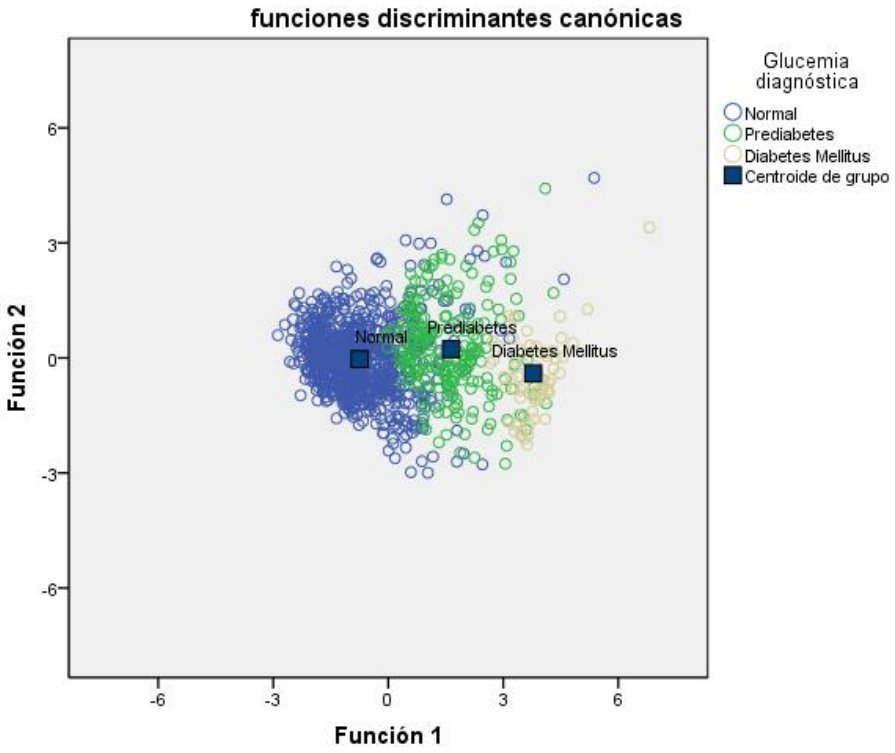
$$Z_2 = -37.429 + 0.109(Colesterol) + 0.017(Triglicéridos) + 0.390(TAS)$$

Función de clasificación para personas diagnosticadas con DM:

$$Z_3 = -54.639 + 0.149(Colesterol) + 0.025(Triglicéridos) + 0.308(TAS)$$

La figura 1 muestra el diagrama de dispersión de todos los casos utilizados en el análisis sobre el plano definido por las dos funciones discriminantes. Los casos están identificados por la Glucemia de las personas. La mayor utilidad de este gráfico radica en la posibilidad de identificar casos atípicos difíciles de clasificar.

Figura 1. Diagrama de dispersión de los tres grupos en las dos funciones discriminantes.



Por último, la matriz de confusión de la tabla 20 ofrece los resultados de la clasificación (probabilidades previas iguales). La tabla indica que se ha clasificado correctamente el 88.0% de los casos agrupados. En el grupo de personas de Diabetes Mellitus se consigue el porcentaje más alto de clasificación correcta, 96.2 %, frente a un porcentaje del 88.5% en el grupo de personas sanas y del 83.7% en el grupo Prediabetes. (Esta circunstancia resulta especialmente llamativa pues, a pesar de que la regla de clasificación se basa en probabilidades a priori iguales para todos los grupos, el porcentaje de clasificación correcta más alto se da precisamente en el grupo de menor tamaño). Basándose en los porcentajes de clasificación correcta de cada grupo se puede afirmar que las personas sanas se confunden, mayoritariamente, con las personas diagnosticadas con diabetes mellitus; y que las personas diagnosticadas con prediabetes y diabetes mellitus no se confunden con las personas sanas, sino entre sí.

Tabla 20: Resultados de la clasificación[a,b](probabilidades previas iguales).

			Glucemia diagnóstica	Grupo de pertenencia pronosticado			
				Normal	Prediabetes	Diabetes Mellitus	Total
Casos seleccionados	Origina	Recuento	Normal	583	72	4	659
			Prediabetes	9	149	20	178
			Diabetes Mellitus	0	2	51	53
		%	Normal	88.5	10.9	.6	100.0
			Prediabetes	5.1	83.7	11.2	100.0
			Diabetes Mellitus	0	3.8	96.2	100.0
Casos no seleccionados	Origina	Recuento	Normal	157	21	0	178
			Prediabetes	2	44	5	51
			Diabetes Mellitus	0	0	16	16
		%	Normal	88.2	11.8	0	100.0
			Prediabetes	3.9	86.3	9.8	100.0
			Diabetes Mellitus	0	0	100.0	100.0

a. Clasificados correctamente el 88.0% de los casos agrupados originales seleccionados.

b. Clasificados correctamente el 88.6% de casos agrupados originales no seleccionados.

La matriz de confusión ofrece los resultados que muestra la tabla 21. El porcentaje de clasificación correcta ha subido del 88.0 % al 88.2 %. Al variar los datos basado en los tamaños de los grupos con la nueva regla de clasificación, ha aumentado el porcentaje de clasificación correcta de las personas más numerosos (las personas sanas), pero algunas de las personas prediabetes y diabetes mellitus se confunden con las personas sanas. La tasa de clasificación correcta del grupo diagnosticada con prediabetes y diabetes mellitus se ha reducido considerablemente. Probablemente las probabilidades previas podrían ser mejor calibradas y ello permitiría obtener mejores resultados en la clasificación.

Como se puede observar, este modelo discriminante ha clasificado correctamente el 88.2% a nivel global de las 890 personas, habiendo clasificado el 94.8% correctamente de 659 personas sanas en el grupo G1, el 65.2% correctamente de 178 para el G2 de las personas diagnosticada con prediabetes y el 83% correctamente de 53 personas diagnosticada con diabetes mellitus en el grupo G3.

La probabilidad de error o probabilidad de clasificar una persona incorrectamente es de 11.8% a nivel global de las 890 personas, habiendo clasificado el 5.2% incorrectamente de 659 personas sanas, clasificado incorrectamente en las personas diagnosticada con prediabetes el 4.9% y el 0.3% en las personas con diabetes mellitus. La probabilidad de clasificar incorrectamente 178 personas diagnosticas con prediabetes es de 34.8%, clasificado incorrectamente en las personas sanas el 29.2% y el 5.6% en las personas con diabetes mellitus. La probabilidad de clasificar una persona con diabetes mellitus incorrectamente es de 17% de 53 personas, clasificado incorrectamente en las personas sanas el 0% y el 17% en las personas diagnosticada con prediabetes.

Para la clasificación de los nuevos datos como se puede observar que se han clasificado correctamente el 88.6% a nivel global de las 245 personas, habiendo clasificado el 95.5% correctamente de 178 personas sanas en el grupo G1, el 62.7% correctamente de 51 para el G2 de las personas diagnosticada con prediabetes y el 93.8% correctamente de 16 personas diagnosticada con diabetes mellitus en el grupo G3. La probabilidad de error o probabilidad de clasificar una persona incorrectamente es de 11.4% a nivel global de las 245 personas.

Tabla 21: Resultados de la clasificación[a,b] (probabilidades previas basada en los tamaños iguales).

			Glucemia diagnóstica	Grupo de pertenencia pronosticado			
				Normal	Prediabetes	Diabetes Mellitus	Total
Casos seleccionados	Origina	Recuento	Normal	625	32	2	659
			Prediabetes	52	116	10	178
			Diabetes Mellitus	0	9	44	53
		%	Normal	94.8	4.9	.3	100.0
			Prediabetes	29.2	65.2	5.6	100.0
			Diabetes Mellitus	.0	17.0	83.0	100.0
Casos no seleccionados	Origina	Recuento	Normal	170	8	0	178
			Prediabetes	16	32	3	51
			Diabetes Mellitus	0	1	15	16
		%	Normal	95.5	4.5	.0	100.0
			Prediabetes	31.4	62.7	5.9	100.0
			Diabetes Mellitus	.0	6.3	93.8	100.0

a. Clasificados correctamente el 88.2% de los casos agrupados originales seleccionados.

b. Clasificados correctamente el 88.6% de casos agrupados originales no seleccionados.

Características del modelo:

Utilizando los resultados de la clasificación con probabilidades basada en el tamaño de los grupos, se obtiene:

Donde:

- G1: personas sana.
- G2: persona diagnosticada con prediabetes.
- G3: persona diagnosticada con diabetes mellitus

La probabilidad de clasificar una persona correctamente es:

P = P (ser del G1 y clasificado en G1) + P (ser del G2 y clasificado en G2)+

+ P (ser del G3 y clasificado en G3)

$$P = \frac{625}{890} + \frac{116}{890} + \frac{44}{890} = \frac{785}{890} = 0.882$$

La probabilidad de error o probabilidad de clasificar una persona incorrectamente es:

P = P (ser del G1 y clasificado en G2) + P (ser del G2 y clasificado en G1)+

+ P (ser del G1 y clasificado en G3) + P (ser del G2 y clasificado en G3) +

+ P (ser del G3 y clasificado en G1) + P (ser del G3 y clasificado en G2)

$$P = \frac{32}{890} + \frac{52}{890} + \frac{2}{890} + \frac{10}{890} + \frac{0}{890} + \frac{9}{890} = \frac{105}{890} = 0.118$$

La probabilidad de clasificar una persona sana en el grupo de la persona sana (G1) dado que es del grupo de las personas sanas (G1), es:

$$P = P(Clasificado\ en\ G1/es\ de\ G1) = \frac{P(ser\ del\ G1\ y\ Clasificado\ en\ G1)}{P(ser\ de\ G1)} =$$

$$= \frac{\frac{625}{890}}{\frac{659}{890}} = \frac{625}{659} = 0.9484$$

La probabilidad de clasificar una persona diagnosticada con prediabetes en el grupo de prediabetica (G2) dado que es del grupo prediabetica (G2), es:

$$P = P(Clasificado\ en\ G2/es\ de\ G2) = \frac{P(ser\ del\ G2\ y\ Clasificado\ en\ G2)}{P(ser\ de\ G2)} =$$

$$= \frac{\frac{116}{890}}{\frac{178}{890}} = \frac{116}{178} = 0.65168$$

La probabilidad de clasificar una persona con diabetes mellitus en el grupo de diabetes (G3) dado que es del grupo diabetes (G3), es:

$$P = P(Clasificado\ en\ G3/es\ de\ G3) = \frac{P(ser\ del\ G3\ y\ Clasificado\ en\ G3)}{P(ser\ de\ G3)} =$$

$$= \frac{\frac{44}{890}}{\frac{53}{890}} = \frac{44}{53} = 0.83018$$

Como se puede observar, en estas tres últimas características, señalan que el modelo que clasifica mejor es el grupo G1 (Personas Sanas).

3.6 Tablas de Contingencia.

Analizar la distribución de una variable con relación a otra u otras es una tarea corriente en Salud Pública, vinculada, la mayoría de las veces, a la búsqueda de un patrón que indique la relación, (o la falta de ella) entre las variables estudiadas. Este es un proceso clave en la identificación de las posibles causas de los problemas de salud, y también de factores que, aun cuando no puedan ser finalmente considerados causales, resulten estar asociados a estos daños y constituyan importantes elementos prácticos para la identificación de grupos con mayores riesgos de padecer determinado daño.

Las tablas de contingencia (tablas de doble entrada) son una herramienta fundamental para este tipo de análisis. Están compuestas por filas (horizontales), para la información de una variable y columnas (verticales) para la información de otra variable. Estas filas y columnas delimitan celdas donde se vuelcan las frecuencias de cada combinación de las variables analizadas. En su expresión más elemental, las tablas tienen solo 2 filas y 2 columnas (tablas de 2x2). Como se puede ver en la tabla 19.

3.6.1 Identificar los factores de riesgo que inciden en el padecimiento de la Diabetes Mellitus.

Para el cálculo de los factores de riesgo que inciden en el padecimiento de diabetes mellitus, se realizó por la razón de momios (RM) o razón de oportunidades en inglés, odds ratio (OR), es una medida estadística utilizada en estudios epidemiológicos transversales y de casos y controles, así como en los metaanálisis (conjunto de herramientas estadísticas, que son útiles para sintetizar los datos de una colección de estudios). En términos formales, se define como la posibilidad de que una condición de salud o enfermedad se presente en un grupo de población frente al riesgo de que ocurra en otro. En epidemiología, la comparación suele realizarse entre grupos humanos que presentan condiciones de vida similares, con la diferencia de que uno se encuentra expuesto a un factor de riesgo (m_i) mientras que el otro carece de esta

característica (m_o). Por lo tanto, la razón de momios o de posibilidades es una medida de tamaño de efecto.

El cálculo del Odds Ratio está dado por la razón de los casos expuestos, con la casilla a, por lo controles no expuestos de la casilla d, en el numerador y los casos no expuestos de la casilla c por los controles expuestos de la casilla b, en el denominador. Por medio de la tabla 22 se puede calcular el OR.

Tabla 22: de contingencia estándar de 2x2 (OR)

		Glucemia Diagnóstica		
		Diabético	No Diabético	Total
	Enfermo	a	b	m_i
	No Enfermo	c	d	m_o
Total		De casos (n_i)	No casos (n_o)	

$$OR = \frac{a * d}{b * c}$$

En la tabla 23 se puede calcular el OR por medio de las tablas de contingencia, donde una persona con hipertensión tiene 2.3 veces más ventaja de tener diabetes mellitus que una persona sin hipertensión.

Tabla 23: Hipertensión arterial escala ordinal versus Glucemia.

		Glucemia diagnóstica		
		Diabético	No Diabético	Total
Hipertensión arterial escala ordinal	Con hipertensión	157	273	430
	sin hipertensión	141	564	705
Total		298	837	1135

Nota: Sin hipertensión $< 120/80$ mmHg. y Con hipertensión $\geq 140/90$ mmHg.

$$OR = \frac{ad}{bc} = \frac{(157)(564)}{(273)(141)} = 2.30$$

En la tabla 24 se puede calcular el OR por medio de las tablas de contingencia, donde una persona con colesterolemia tiene 3.43 veces más ventaja de tener diabetes mellitus que una persona sin colesterolemia.

Tabla 24: Colesterol categorizada versus Glucemia diagnóstica

		Glucemia diagnóstica		
		Diabético	No Diabético	Total
colesterol categorizada	con colesterolemia	193	292	485
	sin colesterolemia	105	545	650
Total		298	837	1135

Nota: Sin colesterolemia $< 200\ mg/dl.$ y Con colesterolemia $\geq 200\ mg/dl$

$$OR = \frac{ad}{bc} = \frac{(193\ )(545)}{(292)(105)} = 3.43$$

En la tabla 25 se puede calcular el OR por medio de las tablas de contingencia, donde una persona con trigliceridemia tiene 3.66 veces más ventaja de tener diabetes mellitus que una persona sin trigliceridemia.

Tabla 25: Triglicérido categorizada versus Glucemia diagnóstica.

		Glucemia diagnóstica		
		Diabético	No Diabético	Total
triglicérido categorizada	Con trigliceridemia	197	291	488
	Sin trigliceridemia	101	546	647
Total		298	837	1135

Nota: Sin trigliceridemia $< 150\ mg/dl.$ y Con trigliceridemia $\geq 150\ mg/dl.$

$$OR = \frac{ad}{bc} = \frac{(197\ )(546)}{(291)(101)} = 3.66$$

En la tabla 26 se puede calcular el OR por medio de las tablas de contingencia, donde una persona con sobre peso y obesidad tiene 1.63 veces más ventaja de tener diabetes mellitus que una persona con peso normal.

Tabla 26: imc categorizada versus Glucemia diagnóstica.

		Glucemia diagnóstica		
		Diabético	No Diabético	Total
imc categorizada	Sobre peso y obesidad	186	423	609
	peso normal	112	414	526
Total		298	837	1135

Nota: P. normal (18.5 – 24.99). Sobre peso (25 – 29.99). Obesidad ≥ 30

$$OR = \frac{ad}{bc} = \frac{(186\ )(414)}{(423)(112)} = 1.63$$

En la tabla 27 se puede calcular el OR por medio de las tablas de contingencia, donde una persona mayor de 41 años tiene 3.92 veces más ventaja de tener diabetes mellitus que una persona de 18 a 40 años.

Tabla 27: Edad categorizada versus Glucemia diagnóstica

		Glucemia diagnóstica		
		Diabético	No Diabético	Total
edad categorizada	41 o mas	186	249	435
	18 - 40	112	588	700
Total		298	837	1135

$$OR = \frac{ad}{bc} = \frac{(186)(588)}{(249)(112)} = 3.92$$

3.6.2 Determinar la prevalencia de Diabetes Mellitus de acuerdo a su edad, sexo y ocupación en la región del Bajo Lempa, municipio de Jiquilisco.

En la tabla 28 se puede calcular la prevalencia por medio de las tablas de contingencia, donde una persona mayor de 41 años tiene 2.68 veces más ventaja de tener diabetes mellitus que una persona de 18 a 40 años.

Tabla 28: Edad categorizada versus Glucemia diagnóstica

		Glucemia diagnóstica		
		Diabético	No Diabético	Total
edad categorizada	41 o mas	186	249	435
	18 - 40	112	588	700
Total		298	837	1135

$$Prevalencia\ expuesto\ \geq 41\ edad = \frac{186}{435} = 0.427 \cong 0.43 = 43\%$$

$$Prevalencia\ no\ expuesto\ de\ 18 -\ 40\ \ edad = \frac{112}{700} = 0.16 = 16\%$$

La razón de prevalencia corresponde a:

$$RP = \frac{0.43}{0.16} = 2.68$$

En la tabla 29 se puede calcular la prevalencia por medio de las tablas de contingencia, donde una mujer tiene 1.04 veces más ventaja de tener diabetes mellitus que un hombre.

Tabla 29: Sexo versus Glucemia diagnóstica

		Glucemia diagnóstica		
		Diabético	No Diabético	Total
Sexo	femenino	170	466	636
	masculino	128	371	499
Total		298	837	1135

$$Prevalencia\ expuesto\ = \frac{170}{636} = 0.267 \cong 0.27 = 27\%$$

$$Prevalencia\ no\ expuesto\ = \frac{128}{499} = 0.256 \cong 0.26 = 26\%$$

La razón de prevalencia corresponde a:

$$RP = \frac{0.27}{0.26} = 1.04$$

En la tabla 30 se puede calcular la prevalencia por medio de las tablas de contingencia, donde un agricultor tiene 2.19 veces más ventaja de tener diabetes mellitus que un desempleado.

Tabla 30: Ocupación Versus Glucemia diagnóstica

		Glucemia diagnóstica		
		Diabético	No Diabético	Total
Ocupación	estudiante	18	186	204
	ama de casa	116	248	364
	desempleado	5	26	31
	otros	32	140	172
	agricultor	127	237	364
Total		298	837	1135

$$Prevalencia\ expuesto\ agricultor = \frac{127}{364} = 0.348 \cong 0.35 = 35\%$$

$$Prevalencia\ no\ expuesto\ desempleado\ = \frac{5}{31} = 0.161 \cong 0.16 = 16\%$$

La razón de prevalencia corresponde a:

$$RP = \frac{0.35}{0.16} = 2.19$$

Ahora, calculando la prevalencia de todas las personas que trabajan (agricultor, ama de casa y otros), y las personas que no trabajan (estudiante, desempleado). Por medio de la tabla 27, se tienen los siguientes resultados:

$$Prevalencia\ expuesto\ de\ personas\ que\ trabajan = \frac{275}{900} = 0.305 \cong 0.31 = 31\%$$

$$Prevalencia\ expuesto\ de\ personas\ que\ no\ trabajan = \frac{23}{235} = 0.097 = 10\%$$

La razón de prevalencia corresponde a:

$$RP = \frac{0.31}{0.1} = 3.1$$

Por lo tanto, una persona que trabaja tiene 3.1 veces más ventaja de tener diabetes mellitus que una persona que no trabaja.

3.7 Comparación de los métodos Análisis Discriminante y Regresión Logística para el caso de dos grupos.

Para construir los dos grupos se unieron las categorías: Prediabetes y Diabetes Mellitus, de esta manera se obtienen los 2 grupos que son: Personas sanas y diabética.

De manera general, según los resultados, los métodos de Análisis Discriminante y Regresión Logística, han clasificado correctamente el 89.9% para el A.D y 84.4% para la regresión logística, en este sentido teniendo una diferencia. Es decir que la efectividad de clasificar, pero cuando intervienen variables independientes categóricas en el modelo, puede ser más adecuado el modelo de regresión logística, ya que este permite la incorporación de dichas variables al modelo; mientras que en el análisis discriminante se requieren que las variables independientes sean cuantitativas. A continuación se presenta los resultados de clasificación y la validación de los nuevos datos en ambos métodos.

3.7.1 Análisis Discriminante.

Funciones de clasificación (Análisis Discriminante) para probabilidades a priori basada en los tamaños para los dos grupos.

Tabla 31: Coeficientes de la función de clasificación

	Glucemia diagnóstica	
	Normal	Diabético
Colesterol	.070	.096
Triglicéridos	-.004	.014
Tensión arterial sistólica	.291	.310
(Constante)	-22.379	-36.419

Funciones discriminantes lineales de Fisher

Función de clasificación para personas sanas:

$$Z_1 = -22.379 + 0.070(Colesterol) - 0.004(Triglicéridos) + 0.291(TAS)$$

Función de clasificación para personas diabética:

$$Z_2 = -36.419 + 0.096(Colesterol) + 0.014(Triglicéridos) + 0.310(TAS)$$

La validación de este modelo la analizamos con la siguiente tabla 32 de resultados de la clasificación, para el análisis discriminante. Como se puede observar, este modelo discriminante ha clasificado correctamente el 89.9% a nivel global de las 890 personas, habiendo clasificado el 95.6% correctamente de 659 personas sanas en el grupo G1 y el 73.6% correctamente de 231 para el grupo G2 de las personas diagnosticada con diabetes mellitus.

La probabilidad de error o probabilidad de clasificar una persona incorrectamente es de 10.1% a nivel global de las 890 personas, habiendo clasificado el 4.4% incorrectamente de 659 personas sanas, clasificado incorrectamente en las personas diagnosticada con diabetes mellitus. La probabilidad de clasificar incorrectamente 231 personas diagnosticas con diabetes mellitus es de 26.4%, clasificado incorrectamente en las personas sanas.

Para la clasificación de los nuevos datos como se puede observar que se han clasificado correctamente el 89.4% a nivel global de las 245 personas.

Tabla 32: Resultados de la clasificación[a,b]

			Glucemia diagnóstica	Grupo de pertenencia pronosticado		
				Sana	Diabético	Total
Casos seleccionados	Original	Recuento	Sana	630	29	659
			Diabético	61	170	231
		%	Sana	95.6	4.4	100.0
			Diabético	26.4	73.6	100.0
Casos no seleccionados	Original	Recuento	Sana	172	6	178
			Diabético	20	47	67
		%	Sana	96.6	3.4	100.0
			Diabético	29.9	70.1	100.0

a. Clasificados correctamente el 89.9% de los casos agrupados originales seleccionados.

b. Clasificados correctamente el 89.4% de casos agrupados originales no seleccionados.

3.7.2 Regresión Logística.

Funciones de clasificación (Regresión Logística) para probabilidades a priori basada en los tamaños para los dos grupo.

Tabla 33: Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1[a]	colescat1	3.791	.532	50.818	1	.000	44.287
	triglicat1	20.087	1694.185	.000	1	.991	5.290E8
	imccat1	-.647	.241	7.203	1	.007	.524
	htaord1	.474	.227	4.350	1	.037	1.606
	edadcat1	.631	.224	7.954	1	.005	1.880
	Constante	-23.432	1694.186	.000	1	.989	.000

a. Variable(s) introducida(s) en el paso 1: colescat1, triglicat1, imccat1, htaord1, edadcat1.

Usando un método de selección de variables hacia adelante, se determina en el primer paso, quedando las todas variables Los parámetros son significativamente distintos de cero al 5%.

El modelo de clasificación es:

$$\hat{\pi}_i = \frac{1}{1 + e^{-3.791(\text{colestcat1}) + 0.647(\text{imccat1}) - 0.474(\text{htaord1}) - 0.631(\text{edadcat1})}}$$

Si $\hat{\pi}_i \geq 0.5$ para el individuo i, entonces este se clasifica en el grupo de las personas diabéticas, en caso contrario en el grupo de las personas sanas.

La validación de este modelo la analizamos con la siguiente tabla 34 de resultados de la clasificación, para la regresión logística. Como se puede observar, este modelo de regresión ha clasificado correctamente el 84.4% a nivel global de las 890 personas, habiendo clasificado el 85.1% correctamente de 659 personas sanas en el grupo G1 y el 82.3% correctamente de 231 para el grupo G2 de las personas diagnosticada con diabetes mellitus. Para la clasificación de los nuevos datos como se puede observar que se han clasificado correctamente el 84.9% a nivel global de las 245 personas.

Tabla 34: de clasificación[c]

Observado			Pronosticado					
			Casos seleccionados[a]			Casos no seleccionados[b]		
			Glucemia diagnóstica		Porcentaje correcto	Glucemia diagnóstica		Porcentaje correcto
			Sana	Diabético		Sana	Diabético	
Paso 1	Glucemia diagnóstica	Sana	561	98	85.1	154	24	86.5
		Diabético	41	190	82.3	13	54	80.6
	Porcentaje global				84.4			84.9

a. Casos seleccionados VALIDACION EQ 1

b. Casos no seleccionados VALIDACION NE 1

c. El valor de corte es .500

Conclusiones.

Mediante, la matriz de clasificación de probabilidades a priori basada en los tamaños iguales, se obtuvo que el 89.9% de las personas fueron clasificadas correctamente por medio de las funciones discriminantes lineales de Fisher, por lo que se comprueba que estas funciones poseen un alto poder discriminante y pueden ser utilizadas para futuras pruebas de clasificación de individuos nuevos.

Mediante, la matriz de clasificación de probabilidades a priori basada en los tamaños iguales, se muestra que para los nuevos datos, las personas fueron clasificadas correctamente el 89.4% a nivel global de las 245 personas.

Los factores de riesgo (OR) que inciden más en el padecimiento de la diabetes mellitus, en la región del Bajo Lempa, municipio de Jiquilisco son: La edad, el colesterol y los triglicérido. El factor de riesgo que menos incide es el índice de masa corporal.

La prevalencia de Diabetes Mellitus: Una persona mayor de 41 años de edad, tiene 2.68 veces más ventaja de tener diabetes mellitus que una persona de 18 a 40 años de edad, en la región del Bajo Lempa, municipio de Jiquilisco.

La prevalencia de Diabetes Mellitus en las mujeres: Una mujer tiene 1.04 veces más ventaja de tener diabetes mellitus que un hombre, en la región del Bajo Lempa, municipio de Jiquilisco.

La prevalencia de Diabetes Mellitus en los agricultores: Un agricultor tiene 2.19 veces más ventaja de tener diabetes mellitus que un desempleado, en la región del Bajo Lempa, municipio de Jiquilisco.

La prevalencia de Diabetes Mellitus de los trabajadores: Un trabajador tiene 3.1 veces más ventaja de tener diabetes mellitus que uno que no trabaja, en la región del Bajo Lempa, municipio de Jiquilisco.

Referencias Bibliográficas.

[1] Arnaiz Quintana A. Tierras pagadas a precio de sangre. Testimonios y retratos del Bajo Lempa Usuluteco. 2nd ed. San Salvador: Editorial Catalunya; 2008. Spanish.

[2] Asociación Americana del corazón. A.H.A. Guías Clínicas. 2010

[3] Barceló A. et al. The cost of diabetes in Latin America and the Caribbean. Bulletin of the World Health Organization. 2003; 81: 19-27.

[4] Barquera S. Prevención de la diabetes: Un problema mundial. Salud Pública Méx. 2003; 45(5):413-414.

[5] DIAMOND Project Group. Incidence and trends of childhood type 1 diabetes worldwide 1990-1999. Diabet Med 2006; 23 (8): 857-866.

[6] Díez-Tejedor E, Fuentes B, Gil Núñez A, Gil Peralta A, Matías Guiu J. Guía para el tratamiento preventivo de la isquemia cerebral. Neurología 2002; 17(Supl. 3):61-75.

[7] Health in the Americas, 2007. Volume I – Regional. World Health Organization.

[8] J. F. Hair, Jr., R. E. Anderson, R. L. Tatham, W. C. Black, 1999, Análisis Multivariante, quinta edición, Pearson Prentice Hall.

[9] Leo P. Krall & Richard S. Beaser. Centro Joslin para la Diabetes. Joslin Diabetes Manual. Lea & Febiger. Filadelfia, 1989. Pág. 34.

[10] OMS. Iniciativa de Diabetes para las Américas (DIA): Plan de Acción para América Latina y El Caribe 2001-2006 [OPS/OMS]. División de prevención y control de enfermedades/ Programa de enfermedades no-transmisibles/Organización Panamericana de la Salud/Organización Mundial de la Salud. Washington DC; 2001.

[11] Organización Panamericana de la Salud Iniciativa Centroamericana de Diabetes mellitus (CAMDI): Encuesta de diabetes mellitus, hipertensión y factores de riesgo de enfermedades crónicas. Belice, San José, San Salvador, Ciudad de Guatemala, Managua y Tegucigalpa, 2009

[12] Peña, Daniel. (2002), Análisis de Datos Multivariante, primera edición, McGraw-Hill \Interamericana de España.

[13] Rubio, R. y M. X., Frojan (2004) “Análisis discriminante a la adhesión al tratamiento en la diabetes mellitus insulinodependiente”. Psicotherma. Volumen 16, nº 4, pp. 548-554

[14] Stegmayr B, Asplund K. Diabetes as a risk factor for stroke. A population perspective. Diabetologia 1995; 38:1061-8.

[15] World Health Organization. Prevention of diabetes mellitus. Report of a WHO Study Group. Geneva: World Health Organization; 1994. No. 844

Referencias Consultadas en Internet.

[1] Asociación Americana de Diabetes – ADA [Internet]. [Citado 18 de Junio de 2015]. Recuperado a partir de: http://www.diabetes.org/es/

[2] Asociación Americana de Diabetes – ADA [Internet]. [Citado 18 de Junio de 2015]. Recuperado a partir de: http://www.diabetes.org/es/vivir-con-diabetes/tratamiento-y-cuidado/medicamentos/insulina/lo-basico-sobre-la-insulina.html

[3] Asociación Salvadoreña de Diabetes – ASADI [Internet].[Citado 8 de Junio de 2015].Recuperado a partir de: http://www.asadi.com.sv/noticias/

[4] Becton, Dickinson and Company (BD) – BD [Internet]. [Citado 23 de Septiembre de 2015]. Recuperado a partir de: http://bd.com/mx/diabetes/main.aspx?cat=3258&id=3296

[5] Biblioteca Nacional de Medicina de los EE.UU (NIH) – MedlinePlus [Internet]. [Citado 8 de Junio de 2015]. Recuperado a partir de: https://www.nlm.nih.gov/medlineplus/spanish/metabolicsyndrome.html

[6] Dan L. Longo, Larry Jameson, Anthony S. Fauci, Stephen L. Hauser, Joseph Loscalzo. Harrison Principios de Medicina Interna, 18a edición – Harrison Medicina [Internet]. [Citado 19 de Junio de 2015]. Recuperado a partir de: http://harrisonmedicina.mhmedical.com/content.aspx?bookid=865&sectionid=68954974

[7] Grupo de Diabetes de la Sociedad Andaluza de Medicina Familiar y Comunitaria - Grupo de Diabetes de la SAMFYC [Internet]. [Citado 14 de Junio de 2015].

Recuperado a partir de: http://www.grupodiabetessamfyc.es/index.php/guia-clinica/aspectos-generales/epidemiologia.html

[8] Grupo Pacifico - [Internet]. [Citado 8 de Junio de 2015]. Recuperado a partir de: https://www.pacifico365.com/descargar/Folleto-diabetes.pdf

[9] Hernández W. Nacimiento y Desarrollo del río Lempa [Internet]. San Salvador: SNET; 2005. p.14. Disponible en: http://www.snet.gob.sv/Geologia/NacimientoEvolucionRLempa.pdf

[10] Infomed Red de Salud de Cuba – SLD [Internet]. [Citado 8 de Junio de 2015]. Recuperado a partir de: http://articulos.sld.cu/diabetes/files/2009/07/cronologia_de_la_diabetes_mellitus.pdf

[11] International Diabetes Federation – IDF [Internet]. [Citado 18 de Junio de 2015]. Recuperado a partir de: http://www.idf.org/sites/default/files/attachments/GDP-Spanish.pdf

[12] International Diabetes Federation – IDF [Internet]. [Citado 8 de Junio de 2015]. Recuperado a partir de: http://www.idf.org/sites/default/files/SP_6E_Atlas_Full.pdf

[13] Médicos de El Salvador - [Internet]. [Citado 8 de Junio de 2015]. Recuperado a partir de: http://www.medicosdeelsalvador.com/Detailed/Art_culos_M_dicos/Cirug_a_General/Diabetes_Mellitus_2668.html

[14] Ministerio de Salud de la Nación Argentina – MSAL [Internet]. [Citado 18 de Junio de 2015]. Recuperado a partir de: http://www.msal.gov.ar/images/stories/bes/graficos/0000000070cnt-2012-08-02_guia-prevencion-diagnostico-tratamiento-diabetes-mellitus-tipo-2.pdf

[15] Ministerio de Salud de la Nación Argentina – MSAL [Internet]. [Citado 18 de Junio de 2015]. Recuperado a partir de: http://www.msal.gov.ar/images/stories/bes/graficos/0000000070cnt-2012-08-02_guia-prevencion-diagnostico-tratamiento-diabetes-mellitus-tipo-2.pdf

[16] NHS Choices - [Internet]. [Citado 8 de Junio de 2015]. Recuperado a partir de: http://www.nhs.uk/translationspanish/Documents/Diabetes_Spanish_FINAL.pdf

[17] Organización Mundial de la Salud – OMS [Internet]. [Citado 8 de Junio de 2015]. Recuperado a partir de: http://apps.who.int/iris/bitstream/10665/66040/1/WHO_NCD_NCS_99.2.pdf?ua=1.

[18] Universidad Autónoma de Madrid – UAM [Internet]. [Citado 10 de Julio de 2015]. Recuperado a partir de: http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/DISCRIMINANTE/analisis-discriminante.pdf

[19] Universidad Complutense de Madrid – UCM [Internet]. [Citado 10 de Junio de 2015]. Recuperado a partir de: http://pendientedemigracion.ucm.es/info/socivmyt/paginas/D_departamento/materiales/analisis_datosyMultivariable/23discr_SPSS.pdf

[20] Universidad de El Salvador – UES [Internet]. [Citado 8 de Junio de 2015]. Recuperado a partir de: https://www.google.com.sv/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CBwQFjAAahUKEwiM5qGD693HAhXPEZIKHWU1CeA&url=http%3A%2F%2Fwww.medicina.ues.edu.sv%2Findex.php%3Foption%3Dcom_docman%26task%3Ddoc_download%26gid%3D255%26Itemid%3D85&usg=AFQjCNEFiM6ckQrjCkWcQCC5AnrK2gBBAQ&bvm=bv.101800829,d.aWw&cad=rja

[21] Universidad de El Salvador – UES [Internet]. [Citado 8 de Junio de 2015]. Recuperado a partir de: http://168.243.33.153/infolib/tesis/50106382.pdf

[22] Universidad de El Salvador – UES [Internet]. [Citado 8 de Junio de 2015]. Recuperado a partir de: http://ri.ues.edu.sv/832/1/10136794.pdf

[23] Universidad de El Salvador – UES [Internet]. [Citado 8 de Junio de 2015]. Recuperado a partir de: http://ri.ues.edu.sv/2117/1/Realizar_seguimiento_farmacoterapeutico_a_pacientes_del_Club_de_diab%C3%A9ticos_del_Hospital_Nacional_Dr._Juan_Jos%C3%A9_Fern%C3%A1ndez_Zacamil._Aplicando_el_m%C3%A9todo_Dader.pdf

[24] Universidad Dr. José Matías Delgado - UJMD [Internet]. [Citado 8 de Junio de 2015]. Recuperado a partir de: http://www.redicces.org.sv/jspui/bitstream/10972/739/1/0000061-ADTESLE.pdf