



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

**Modelos estocásticos aplicados a determinación de factores prevalentes
asociados a grupos homogéneos de pacientes con diabetes mellitus tipo 2,
en la ciudad de Guaranda año 2021**

MANUEL MESIAS MONAR SOLORZANO

Trabajo de titulación modalidad: Proyecto de investigación y Desarrollo, presentado
ante el Instituto de Posgrado y Educación Continua de la **ESPOCH**, como requisito
parcial para la obtención del grado de:

MAGÍSTER EN MATEMÁTICAS MENCIÓN MODELACIÓN Y DOCENCIA

Riobamba – Ecuador

Octubre 2022

©2022, Manuel Mesias Monar Solórzano

Se autoriza la reproducción total o parcial, con fines académicos por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento, siempre y cuando se reconozca el Derecho del Autor.



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

EL TRIBUNAL DEL TRABAJO DE TITULACIÓN CERTIFICA QUE:

El Trabajo de Titulación modalidad Proyecto de investigación y desarrollo, titulado: Modelos estocásticos aplicados a determinación de factores prevalentes asociados a grupos homogéneos de pacientes con diabetes mellitus tipo 2, en la ciudad de Guaranda año 2021, de responsabilidad del señor Manuel Mesias Monar Solorzano, ha sido prolijamente revisado y se autoriza su presentación.

Dra. Silvia Mariana Haro Rivera Mag.
PRESIDENTE



Ing. Sergio Alberto Castillo Páez. Ph. D.
DIRECTOR



Ing. Guanga Chunata Deysi Margoth Mag.
MIEMBRO



Ing. Juan Carlos Paz Espinoza Mag.
MIEMBRO



Riobamba, octubre 2022

DERECHOS INTELECTUALES

Yo, Manuel Mesias Monar Solórzano, soy responsable de las ideas, doctrinas y resultados expuestos en el **Trabajo de Titulación modalidad Proyecto de Investigación y Desarrollo**, y el patrimonio intelectual generado por la misma pertenece exclusivamente a la Escuela Superior Politécnica de Chimborazo.



Manuel Mesias Monar Solórzano

No. Cedula: 0201761954

DECLARACIÓN DE AUTENTICIDAD

Yo, Manuel Mesias Monar Solórzano, declaro que el presente proyecto de investigación es de mi autoría y que los resultados del mismo son auténticos y originales. Los textos constantes en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autor, asumo la responsabilidad legal y académica de los contenidos de este Trabajo de Titulación de maestría.



Manuel Mesias Monar Solórzano

No. Cedula: 0201761954

DEDICATORIA

Este trabajo va en virtud de mi esfuerzo y superación profesional, y dedico a Dios por darme las ganas y la facultad de superación a mi madre quien es la persona que con su compañía y fortaleza acompaña todos mis pasos, a mis hermanos Juan y Fredy por ser las personas que están a mi lado dando fuerzas y a mis sobrinos Martin y Samuel que con su ternura incentivan a ser cada día mejor.

Manuel Monar

AGRADECIMIENTO

Para la realización de este trabajo definitivamente fue el conjunto de múltiples actividades que por sí solas no se hubiesen hecho realidad, por lo consiguiente agradezco primero a la Escuela Superior Politécnica de Chimborazo, al Instituto de Posgrado y Educación Continua y su programa de maestría en matemáticas mención Modelación y Docencia que me abrieron las puertas y me dieron la oportunidad de crecer profesionalmente, a mi tutor Dr. Sergio Castillo Páez por ser la guía, mentor y director de mi tesis, a los Ingenieros Deysi Guanga y Juan Carlos Paz Espinoza como evaluadores de mi trabajo que en todo momento me brindaron su ayuda, a la Ing. Sonia Salazar decana de la Facultad de Ciencias Agropecuarias de la universidad Estatal de Bolívar quien con su insistencia y consejos incentivaron a terminar este trabajo, a mis compañeros ing. Verónica Arguello y Carlos Pérez cómplices de esta aventura.

TABAL DE CONTENIDO

RESUMEN	xv
ABSTRACT.....	xvi
CAPÍTULO I	1
1. INTRODUCCIÓN.....	1
1.1. Planteamiento Del Problema.....	1
1.2. Formulación del problema.	2
1.3. Preguntas directrices.	2
1.4. Justificación de la Investigación.	2
1.5. Objetivos de la Investigación	3
1.5.1. Objetivo general	3
1.5.2. Objetivos específicos.....	3
1.6. Hipótesis de investigación.....	3
1.6.1. Hipótesis de trabajo:	3
1.6.2. Hipótesis de Nula:	3
1.6.3. Identificación de Variables	4
CAPÍTULO II	5
2. MARCO TEÓRICO.	5
2.1. Antecedentes del Problema.....	5
2.2. Bases Teóricas.....	11
2.2.1. Modelación Matemática.	11
2.2.2. Modelos estocásticos	12
2.2.6. Análisis de datos categóricos.....	15
2.2.7. Análisis de Componentes principales categóricos (CAPTCA)	25

CAPÍTULO III.....	35
3. METODOLOGÍA.....	35
3.1. Tipo y diseño de investigación.....	35
3.2. Métodos de investigación.....	35
3.3. Enfoque de la Investigación.....	35
3.4. Alcance de la Investigación.....	36
3.5. Población de Estudio.....	36
3.6. Tratamiento de los Datos.....	36
CAPÍTULO IV	38
4. RESULTADOS según OBJETIVOS PLANTEADOS	38
4.1 Descripción estadística de factores que inciden en el crecimiento mediante análisis exploratorio de datos.....	38
4.2 Identificación de factores mediante tabla de contingencias que inciden en el crecimiento con personas con diabetes.....	49
4.2.1. Anova para determinar chi cuadrado de las variables sociodemográficas.....	49
4.2.1. Anava para determinar chi cuadrado de los factores modificables.....	51
4.2.1. Anova para determinar chi cuadrado de los factores no modificables.....	55
4.3 Modelos estadísticos mediante Análisis de componentes principales para datos categóricos de factores de crecimiento de los pacientes de diabetes mellitus tipo 2.....	58
CONCLUSIONES.....	75
RECOMENDACIONES.....	76
GLOSARIO	
BIBLIOGRAFÍA	
ANEXOS	

ÍNDICE DE TABLAS

Tabla 1-2: Matriz de tabla de contingencia	18
Tabla 2-4: Análisis exploratorio de la variable Lugar de atención medica de personas que usan la red de salud publica	39
Tabla 3-4: Lugar de Atención según los Centros y Subcentros de la Red Pública de Salud	40
Tabla 4-4: Clasificación de los usuarios por su localidad de residencia.....	41
Tabla 5-4: Análisis exploratorio de datos de la variable localidad de la población	41
Tabla 6-4: Análisis exploratorio de datos de la variable sexo del paciente	43
Tabla 7-4: distribución de frecuencias de la variable sexo del paciente.....	43
Tabla 8-4: Análisis exploratorio de datos de la variable Autoidentificación del paciente ...	44
Tabla 9-4: Distribución de frecuencia de la variable autoidentificación	45
Tabla 10-4: Análisis exploratorio de datos de la variable complejidad del diagnóstico	46
Tabla 11-4: Distribución de frecuencias de la variable complejidad del diagnostico	46
Tabla 12-4: Análisis exploratorio de la variable clasificación de la IMC	47
Tabla 13-4: Distribución de frecuencias de la variable clasificación del IMC.....	48
Tabla 14-4: Tabla de contingencia entre las variables el nivel de ingreso vs destino de la población (Rural, Urbano).	49
Tabla 15-4: Pruebas de chi-cuadrado.....	50
Tabla 16-4: Tabla de contingencia destino de la población (Rural, Urbano) y formación académica.....	50
Tabla 17-4: Chi cuadrado de la tabla de contingencia nivel de formación y destino de la población.....	51
Tabla 18-4: Tabla de contingencia de las variables actividad física y Destino de la Población	51

Tabla 19-4: Pruebas de chi-cuadrado de la variable actividad física y destino de la población	52
Tabla 20-4: Tabla cruzada Tabla cruzada Clasificación del IMC*Destino de la Población	53
Tabla 21-4: Pruebas de chi-cuadrado de las variables.....	54
Tabla 22-4: Tabla cruzada Consume por lo menos dos tabacos al día*Destino de la Población.	54
Tabla 23-4: Pruebas de chi-cuadrado variables consumo de tabacos oblación	54
Tabla 24-4: Tabla de contingencia edad del paciente vs destino de la población (Rural, Urbano).	55
Tabla 25-4: Chi cuadrado de la variable Edad y destino de la población.....	56
Tabla 26-4: Tabla cruzada de la variable antecedentes familiares vs destino de la población	56
Tabla 27-4: Pruebas de chi-cuadrado de la variable destino de la población y antecedentes familiares	57
Tabla 28-4: Tabla de contingencia entre la variable autoidentificación y destino de la población.....	57
Tabla 29-4: Pruebas de chi-cuadrado entre la variable autoidentificación y destino de la población.....	58
Tabla 30-4: Dimensiones del Modelo (CAPTCA)	58
Tabla 31-4: Variables transformadas de correlaciones.....	72

ÍNDICE DE FIGURAS

Figura 1-2 Clasificación de procesos estocásticos.....	12
Figura 2-4: Mapa de casos con diabetes en la ciudad de Guaranda.....	39
Figura 3-4: Curva de distribución de frecuencias de la atención de pacientes en Centros y Subcentros de la red pública de salud.....	40
Figura 4: Curva de distribución de la variable localidad de la población.....	42
Figura 5-4: curva de distribución de la variable sexo.....	44
Figura 6-4: Curva de distribución de la variable tipo de diagnóstico.....	45
Figura 7-4: Curva de distribución de la variable complejidad del diagnóstico.....	47
Figura 8-4: Curva de distribución normal de la variable clasificación del IMC.....	48
Figura 9-4: Cuantificación de puntos variable lugar de atención de pacientes con diabetes mellitus tipo 2.....	59
Figura 10-4: Cuantificación de puntos de categoría variable autoidentificación.....	60
Figura 11-4: Cuantificación de categoría variable Destino de la población.....	61
Figura 12-4: Cuantificación de puntos de categoría variable sexo del paciente.....	62
Figura 13-4: Cuantificación de puntos de categoría variable tipo de diagnóstico.....	63
Figura 14-4: Cuantificación de puntos de categorías variable complejidad del diagnóstico.....	64
Figura 15-4: Cuantificación de puntos de categoría de la variable Clasificación del IMC.....	65
Figura 16-4: Cuantificación categórica de puntos de la variable realiza actividad física.....	66
Figura 17-4: Cuantificación de puntos de categoría de la variable consumo de tabacos.....	67
Figura 18-4: Cuantificación de puntos de categoría de la variable Familiar diagnosticado con diabetes Realizado por: Monar Manuel, 2022.....	68
Figura 19-4: Cuantificación de puntos de categoría de la variable nivel de ingresos familiares.....	69
Figura 20-4: Cuantificación de puntos de categoría de la variable formación académica.....	70

Figura 21-4: Puntos de categoría de la viable edad paciente	71
Figura 22-4: Grafica de componentes principales	73

ÍNDICE DE ANEXOS

Anexo A: captura de la base de datos de análisis.

Anexo B: Captura de la encuesta realizada a los pacientes

Anexo C: Código del estudio de variables en el software SPSS

Anexo D: Análisis de componentes principales para datos categóricos

Anexo E: Resumen del modelo

Anexo F: historial de interacciones

RESUMEN

El objetivo de esta investigación es determinar mediante modelos estocásticos aplicados a los factores asociados a grupos homogéneos de pacientes con diabetes mellitus tipo 2, en la ciudad de Guaranda. El número de personas con diabetes mellitus tipo II, que lo padecen tienen una progresión de incremento significativo, los factores para su propagación son diversos y relacionadas con aspectos de la vida cotidiana y condiciones genéticas. Se identificó los factores de crecimiento en pacientes con diabetes y se diseñó modelaciones estadísticas de los factores de crecimiento. Mediante la estadística descriptiva se identificó y clasificó los factores prevalente mediante variables cualitativas categóricas, y se realizó un análisis de datos categóricos con la herramienta CAPTCA del software SPSS donde se concluyó que el mayor número de personas con diabetes en la ciudad de Guaranda en la zona urbana con un 53% y los hombre en el grupo de mestizo dentro del factor de autoidentificación son los que tiene mayor incidencia, debido a tiene más mayor factores asociados como el consumo de tabaco, sedentarismo, lo que incrementa el grado de índice de masa corporal.

Palabras Clave: <MATEMÁTICAS>, <MODELACIÓN ESTOCÁSTICA>, <FACTORES ASOCIADOS>, <CRECIMIENTO DE PACIENTE DIABETES>.



30-09-2022

0138-DBRA-UPT-IPEC-2022

ABSTRACT

The objective of this research is to determine through stochastic models applied to the factors associated with homogeneous groups of patients with diabetes mellitus type 2, in the city of Guaranda. The number of people with type II diabetes mellitus, who suffer from it have a significant increase progression, the factors for its propagation are diverse and related to aspects of daily life and genetic conditions. The growth factors in patients with diabetes were identified and statistical modeling of the growth factors was designed. Using descriptive statistics, the prevalent factors were identified and classified by categorical qualitative variables, and a categorical data analysis was performed with the CAPTCA tool of SPSS software where it was concluded that the largest number of people with diabetes in the city of Guaranda in the urban area with 53% and men in the metis group within the factor of self-identification are those with the highest incidence, due to having more associated factors such as smoking, sedentary lifestyle, which increases the degree of body mass index.

Keywords: <MATEMATICS), <STOCASTIC MODELLATION>, <ASSOCIATED FACTORS>, <GROWTH OF DIABETES PATIENT>.

CAPÍTULO I

1. INTRODUCCIÓN

1.1. Planteamiento Del Problema.

La Diabetes Mellitus tipo II es considerado un problema de salud pública, catalogada como una causa primaria de muertes para quienes los padecen según, (Carabetti & Aníbal, 2017). La prevalencia de la diabetes en el mundo está en aumento, el número creciente de personas que lo padecen tienen una progresión de incremento significativo, los factores para su propagación son diversos y relacionadas con aspectos de la vida cotidiana, su afectación en las personas es significativo su tratamiento es permanente y costoso, la complicación no solo incide en la persona que lo padece, sino que cambia radicalmente el estilo de vida de sus familiares.

La diabetes es una enfermedad metabólica altamente asociada a un mayor riesgo cardiovascular y con aumento progresivo e incremento de las dificultades en las personas que lo padecen, su prevalencia es complicada para las personas de edad avanzada y que tiene diferentes hábitos que agrava esta enfermedad, es así que el riesgo de desarrollar esta enfermedad se incrementa según según pasan los años de edad (Palacios et al., 2012). Adicionalmente (Leiva et al., 2018) manifiesta que la diabetes mellitus tipo dos (DMT2) lo considera una de las cuatro enfermedades crónicas no transmisibles de mayor prioridad en la actualidad, su prevalencia está siendo duplicada a partir de los años 1980, pasando de una 4,7% al 8,5% en la población adulta y adicionalmente este aumento se da en los países de bajos y medianos ingresos, dificultado la calidad de vida.

Existen diverso factores relacionados con el desarrollo de esta enfermedad, algunos de estos factores son no modificables como la edad, sexo, historial familiar de DMT2, región de origen, para (Leiva et al., 2018) la raza e historial familiar acompaña una gran disposición genética a sufrir de DMT2 en aquellos individuos con padre diabético su posibilidad se incrementa en un 40% si los dos padres son diabéticos la posibilidad se eleva a un 70% , también existen diferentes grupos étnicos que tiene mayor posibilidad de desarrollar DMT2 estos factores su dificultad se incrementa ya que existen hábitos y factores modificables relacionados con el estilo de vida dado por el

aumento de peso corporal, inactividad física, tabaquismo consumo de alcohol y consumo excesivo de sal y azúcar, el sobre peso y la obesidad representan los más importantes factores de desarrollo de DMT2 y esa dificultad va en aumento progresivo a nivel mundial y sobre todo en Latinoamérica.

La diabetes al ser un enfermedad y problema de salud pública se hace necesario e importante el estudio e identificación de factores de riesgo asociados a esta enfermedad clasificándole en grupos homogéneas y determinando su crecimiento en relación con dichos grupos y factores, la moderación estadística usada como herramienta cuantitativa permitirá verificar estos resultados y sacar conclusiones.

1.2. Formulación del problema.

¿Es factible identificar los factores de crecimiento del número de personas con diabetes mellitus tipo 2, en la ciudad de Guaranda año 2022?

1.3. Preguntas directrices.

1. ¿Es factible identificar los factores de crecimiento con diabetes en la ciudad de Guaranda en el año 2022?
2. ¿Identificar los factores más significativos de las personas con diabetes mellitus tipo 2?
3. ¿Identificar los factores más relacionados con el crecimiento de números de pacientes utilizando modelos estadísticos?

1.4. Justificación de la Investigación.

La presente investigación trata de analizar de manera crítica, investigativa y relacional sobre la prevalencia de la diabetes mellitus, a través de la identificación de los factores de riesgo, su clasificación y valoración con esto se procede a un modelamiento estocástico para determinar su crecimiento y afectación, la construcción de modelos estadísticos ayudara de una mejor manera a comprender la dinámica de los factores, la interacción de los mismos y su forma de involucramiento entre las variables asociadas a la enfermedad y la prevalencia para que se dé el crecimiento.

En el presente proyecto de titulación, se busca aplicar los modelos estadísticos, probabilísticos adecuados al campo de la salud en especial a su afectación de las personas con Diabetes Mellitus tipo 2 de la Ciudad de Guaranda, y ser un precedente en vista que no hay trabajos relacionados con esta área en la ciudad.

1.5. Objetivos de la Investigación

1.5.1. Objetivo general

Determinar los factores prevalentes asociados a grupos homogéneos de pacientes con diabetes mellitus tipo 2, en la ciudad de Guaranda año 2021.

1.5.2. Objetivos específicos.

- Describir los factores que inciden en el crecimiento de personas con diabetes mellitus tipo 2 en la ciudad de Guaranda año 2021.
- Identificar los factores de crecimiento en pacientes con diabetes mellitus tipo 2 ciudad de Guaranda en el año 2021.
- Diseñar modelaciones estadísticas de factores de crecimiento con grupos de afectación de los pacientes de diabetes mellitus tipo 2, en la población de la ciudad de Guaranda, año 2021.

1.6. Hipótesis de investigación

1.6.1. Hipótesis de trabajo:

La modelación estocástica mediante el uso del software SPSS, permite determinar los factores de crecimiento en pacientes con diabetes Mellitus tipo II de la ciudad de Guaranda.

1.6.2. Hipótesis de Nula:

La modelación estocástica mediante el uso del software SPSS, no permite determinar el crecimiento de personas con diabetes Mellitus tipo II de la ciudad de Guaranda.

1.6.3. Identificación de Variables

Variable independiente:

Factores de crecimiento en pacientes con diabetes mellitus tipo 2.

Variable dependiente:

Grupos homogéneos de pacientes con diabetes mellitus tipo II.

CAPÍTULO II

2. MARCO TEÓRICO.

2.1. Antecedentes del Problema.

(Quisiguiña, 2011) define a la Diabetes Mellitus como “Un desorden metabólico crónico caracterizado por niveles persistentemente elevados de glucosa en la sangre, como consecuencia de una alteración en la secreción y/o acción de la insulina. La Diabetes mellitus tipo 2 es caracterizada por resistencia insulínica, que habitualmente se acompaña de un déficit relativo de insulina”.

A nivel mundial la incidencia de la diabetes como enfermedad crónica está teniendo una elevada incidencia, considerándose como problema emergente mundial, por lo tanto, un problema de salud pública, por su afectación a personas indiferentemente de la edad, sexo, clase social o área geográfica, su elevada mortalidad y costos sanitarios han generado en una pandemia, los pacientes con esta afectación en el 2013 fueron de 382 millones y se prevé para el 2035 padecerán cerca de 592 millones de pacientes siendo los países de bajo y medianos ingresos los que se enfrentaran a este padecimiento (Rizo & Sandoval, 2016).

La Organización Mundial de la Salud (OMS), concluye que la diabetes mellitus tipo 2 es una problemática de la salud pública, con gran repercusión social, económica y sobre poner aspectos deficientes en la salud y calidad de vida de los pacientes y familiares en el entorno comunitario, en este sentido el factor de riesgo con mayor prevalencia a nivel mundial, cerca del 90% corresponde al sobre peso y la falta de actividad física, incrementado el gasto sanitario desproporcionado, la presencia de cardiopatía y evento cerebro vascular

La diabetes se está convirtiendo en una de las emergencia mundiales para el siglo XXI, cada año el incremento de personas que viven con esta afección que puede desencadenar con complicaciones a lo largo de la vida de las personas que lo padecen, 418 millones de personas especialmente adultos mayores con tolerancia a la glucosa, muchos países no son conscientes de

la problemática mundial de esta epidemia además del impacto social y económico (Federation, 2015)

En los últimos años se ha observado una alarmante incremento de la prevalencia de diabetes tipo 2 en los centros pediátricos de todo el mundo, esta nueva problemática de diabetes en niños tiene un conjetura a nivel mundial, complicando por el incremento de los factores ambientales y genéticos, prevaleciendo los pacientes por lo general obesos, y los grupos raciales muestran una tendencia de riesgo (Frenk & Márkez, 2006).

La Diabetes mellitus tipo dos, está catalogada como una causa principal de hospitalización, consulta y mortalidad, entre las prevalencias principales para el incremento de esta enfermedad está el sobrepeso que refleja un 43,5% una problemática para aquello es la falta de ejercicios, acrecentado esta dificultad a los factores de riesgos modificables crónicas no transmisibles. La diabetes está asociada también a factores de riesgos como la edad, índice de masa corporal, antecedentes familiares entre otros (Falconi et al., 2017).

La diabetes mellitus por su elevada morbimortalidad se encuentra entre las cuatro principales enfermedades no transmisibles que se han convertido en uno de los mayores desafíos para el desarrollo del siglo XXI, clásicamente la diabetes mellitus tipo 2 (DM2) ha sido considerada una condición de los adultos, pero las evidencia y estudios constantes demuestran que tiene una incidencia creciente en niños y adolescentes (Gómez et al., 2019). En América Latina aún no se cuenta con un registro oficial de la incidencia de niños y adolescentes con DM2, es decir que no existen estudios epidemiológicos que traten al respecto pero según fuentes oficiales del 2 al 3 % de las personas con DM2 se encuentran en edad pediátrica (Gómez et al., 2019).

La falta de diagnóstico oportuno es una importante causa de desarrollo prematuro de las formas más severas de DM2 (Columbié et al., 2016). Se ha demostrado que el manejo temprano de pacientes con DM2 retrasa y reduce la incidencia de complicaciones derivadas de la enfermedad, así mismo, es posible reducir sensiblemente los impactos de DM2 mediante intervención sobre factores de riesgo frecuentes en población sana (Columbié et al., 2016).

El comportamiento de DM2 a nivel mundial muestra mayores tasas de incidencia en hombres que en mujeres, especialmente en edades mayores a 30 años sin embargo, se ha reportado que tras el

establecimiento de DM2, en las mujeres se observa un mayor incremento de riesgo de enfermedad cardiovascular, demencia, cáncer y enfermedad renal en comparación con los hombres, sin duda estas disparidades se atribuyen principalmente a diferencias biológicas y conductuales (Leiva et al., 2018).

El impacto de las complicaciones microvasculares y macrovasculares en sus dos fases de morbilidad, mortalidad sumado, a esto el cambio de la calidad de vida de los habitantes de los países de ingresos medios y bajos está considerado uno de los principales problemas sociosanitarios del mundo actual, el cambio de vida, sedentarismo, y falta de ejercicio, el consumo de una dieta rica en carbohidratos y una elevada ingesta calórica, han modificado la incidencia y la prevalencia de la diabetes mellitus (Calderón, 2007).

Los factores de riesgo modificables como prevalencia e incidencia en la diabetes mellitus tipo 2 están identificados a la obesidad que es el sobre peso, calculado en función del índice de masa corporal (IMC), la relación directa del aumento del ICM está estrechamente ligada con la DM2 de modo que cada aumento unitario del ICM se asocia con un incremento del riesgo del 12%, por cada incremento eleva en un 4,5% el riesgo de desarrollar diabetes en los próximos 10 años. La actividad física es otro factor modificable que tiene una relación directa con el riesgo de desarrollo de esta prevalencia, múltiples estudios muestran que la actividad física acompaña un descenso de la incidencia de la diabetes, el ejercicio mejora la sensibilidad a la insulina con reducción. Como el riesgo de padecer la enfermedad aumenta con el número de factores de riesgo presentes en el individuo, y con el fin de mejorar la identificación aquí se detallan los siguientes:

Factores de riesgo no modificables

- Edad. La prevalencia de DM2 aumenta a partir de la mediana edad, y es mayor en la tercera edad (Ezkurra, 2016).
- Raza/etnia. El riesgo de desarrollar DM2 es menor en individuos de raza caucásica que en hispanos, asiáticos, negros y grupos nativos americanos (indios, alaskaños, hawaianos, etc.), que además presentan una evolución más rápida a diabetes mellitus (DM) (Ezkurra, 2016).
- Otro factor de riesgo son los antecedentes de DM2 en un familiar de primer grado. Los individuos con padre o madre con DM2 tienen entre dos y tres veces (cinco o seis si ambos

padres presentan la condición) mayor riesgo de desarrollar la enfermedad (Bohórquez et al., 2020).

- Antecedente de DM gestacional. Las mujeres con antecedentes de DM gestacional tienen alrededor de 7,5 veces mayor riesgo de DM2 en comparación con las mujeres sin la condición (Gómez et al., 2019).
- El síndrome del ovario poliquístico. Este síndrome se asocia a alteraciones en la regulación de la glucosa en diferentes poblaciones, por ejemplo en Estados Unidos alcanza hasta un 40 % la incidencia en mujeres con síndrome de ovario poliquístico ya que se ve afectada la regulación de su glucosa a partir de los 40 años; así mismo un meta análisis reveló que mujeres con dicho síndrome padecen tres veces mayor riesgo de DM gestacional (Leiva et al., 2018).

Factores de riesgo modificables

- Obesidad, sobrepeso y obesidad abdominal. Estos factores aumentan el riesgo de intolerancia a la glucosa y DM2 en todas las edades, las mismas que inducen la resistencia a la insulina, sin duda es preocupante conocer que más del 80 % de los casos de DM2 se atribuyen a la obesidad, aunque si esta condición se revierte, también se puede disminuir el riesgo y mejorar el control glicémico en pacientes con DM diagnosticada; es importante mencionar que el riesgo relativo ajustado por edad para DM fue de 6 a 1 veces mayor para las mujeres con IMC mayor 35 kg/m² que para aquellas con IMC menor 24,9 kg/m². Otra característica, como es el perímetro de la cintura, al aumentar 1 cm de este, el riesgo de padecer DM2 y de glicemia basal alterada se eleva (Palacios et al., 2012).
- El sedentarismo: Este estilo de vida reduce el gasto de energía y promueve el aumento de peso, lo que eleva el riesgo de DM2, es importante mencionar que la conducta sedentaria predominante es ver la televisión por tiempo prolongado, la misma que puede desarrollar obesidad y DM con el tiempo. A diferencia de practicar actividad física de intensidad moderada reduce la incidencia de nuevos casos de DM2 (Ezkurra, 2016). independientemente de la presencia o ausencia de intolerancia a la glucosa, como han demostrado diversos estudios.
- El Tabaquismo: Consumir continuamente varias dosis diarias de tabaco también se asocia a presentar un mayor riesgo de DM2 dependiente, sin embargo dejar de fumar también

puede reducir el riesgo de DM y el beneficio es evidente cinco años después del abandono y se equipara al de los que nunca fumaron después de 20 años (Ezkurra, 2016).

- Los Patrones dietéticos: sin duda una dieta con un predominante consumo de carnes rojas, lácteos altos en grasa, bebidas azucaradas, alimentos procesados como dulces y postres se asocian a un mayor riesgo de padecer DM2 independientemente del IMC; el riesgo de DM2 fue evidentemente mayor entre los sujetos que consumen este tipo de alimentos en su dieta y además son obesos (Gómez et al., 2019). Contrastando la información antes mencionada, se puede indicar que las personas que siguen una dieta con un mayor aporte de verduras, frutas, pescados blancos y azules, aves y cereales integrales, presentan una significativa reducción del riesgo de padecer DM (Bohórquez et al., 2020).
- Los Trastornos de regulación de la glucosa: Llamados también estados intermedios de hiperglucemia o prediabetes, presentan glicemias basales alteradas, tolerancia a la glucosa y aumento de la hemoglobina glicosilada, su presencia aislada o conjuntamente evidencian un mayor riesgo de DM2 (Ezkurra, 2016).
- Condicionantes clínicos asociados a mayor riesgo de DM2: En este grupo están los pacientes que padecen enfermedades coronarias e insuficiencia cardíaca avanzada, estos presentan un mayor riesgo de desarrollar DM, adicional al padecer hipertensión arterial, infarto agudo de miocardio y el ictus también están son factores de riesgo para padecer DM2 (Ezkurra, 2016).
- Otros factores. Se ha comprobado una relación en U entre el peso al nacer y el riesgo de presentar DM2; un peso alto o bajo al nacer se asocia similarmente con mayor riesgo de DM2 durante la vida. Neonatos prematuros, cualquiera que sea su peso, también pueden estar en mayor riesgo de padecer DM2 (Gómez et al., 2019).

Al hablar del cuidado de la salud, se puede ampliar y conocer los beneficios de la dieta mediterránea en la que predomina un alto consumo de frutas, verduras, cereales integrales y fuentes de grasa como nueces y aceite de oliva, este tipo de dieta reduce la aparición de DM2 hasta un 40 % sin necesidad de reducir peso; en cuanto a los componentes específicos de los alimentos, el consumo de productos lácteos bajos en grasa, fibra, nueces, café descafeinado y té verde a largo plazo también disminuyen el riesgo de presentar DM2 (Columbié et al., 2016).

Adicionalmente cabe destacar que la lactancia materna exclusiva se asocia con un descenso del riesgo de DM, porcentualmente se indica que existe 15 % de reducción por cada año de lactancia materna, hasta 15 años después posterior del último parto; sin embargo en las madres con DM gestacional no se evidencio beneficios (Ezkurra, 2016).

(Guevara & Carrasco, 2001), en su estudio sobre modelación estadística de DBO y OD. Caso de Estudio, donde como base en los diferentes modelos lineales y no lineales permite modelar y analizar los aspectos químicos de la demanda bioquímica del suelo y el oxígeno directo como aspectos importantes de la calidad del agua, donde se concluyó con la identificación de parámetros de calidad después de la descarga en las riberas de los ríos.

(Rivera & Valencia, 2019), en su investigación sobre simulación estocástico para determinar el valor presente neto y costo de incertidumbre en una planta eólica donde hace una simulación sobre los costos de inversión y los riegos en la construcción de una planta eólica, en base a la densidad de las probabilidades del valor presente del proyecto, donde a través de histogramas se pudo obtener la densidad del valor presente del proyecto.

(Villalba Mata, 2011) por medio de su investigación doctoral sobre la construcción y utilización de un modelo estocástico para la simulación de estrategias de manejo invernal en rebaños de vacas nodrizas, donde la utilización de modelación estocástica, dada en una distribución de probabilidades.

(Neciosup Vera, 2018) en su investigación sobre los modelos lineales mixtos aplicado a clases de datos latentes longitudinales en el sector de la salud, donde a través de la aplicación de un proceso estocástico, y la simulación para evaluar el desempeño de indicadores en la salud, donde el autor concluye que el modelo lineal mixto es ideal para modelar simultáneamente un proceso longitudinal.

(Montesinos López & Hernández Suárez, 2007) a través de su investigación describe la importancia de los modelos matemáticos en la comprensión y dinámica de la transmisión de las enfermedades infecciosas, el autor enfatiza la revisión bibliográfica relacionada al tema para encontrar hallazgos relacionados al tema. Donde el autor concluye que el empleo de los modelos

matemáticos en áreas de las enfermedades infecciosas, está creciendo por significativamente por lo útil en la toma de decisiones.

2.2. Bases Teóricas.

2.2.1. Modelación Matemática.

La importancia de la modelación matemática permite establecer una dinámica y comprensión de los factores y variables endógenas y exógenas, así como el dinamismo de la interacción del crecimiento, el diseño de medidas y modelos eficaces de control, el uso de modelos cuantitativos en especial los modelos matemáticos está ganando notoriedad en la investigación científica, por el dinamismo explicativo de los factores de crecimiento de enfermedades contagiosas, esto permite idear programas de control e identificación de patrones. (Montesinos López & Hernández Suárez, 2007).

La construcción de modelos matemáticos es una herramienta que se utiliza en los estudios del área de la salud sobre todo en problemas de medicina, biología, fisiología, bioquímica, epidemiología, farmacocinética, entre otras áreas del conocimiento, la modelación matemática y estadística permite la descripción, explicación, predicción de fenómenos y procesos en dichas áreas, aunque su aplicación es limitada por la falta de conocimientos información acerca de los principios básicos del modelamiento matemático .

En el sector de la salud, la cantidad de datos observados de las unidades con el objetivo de describir el proceso estocástico, para ellos se enfatizará en los modelos lineales mixtos (LMM) que permitirá identificar el patrón de variabilidad.

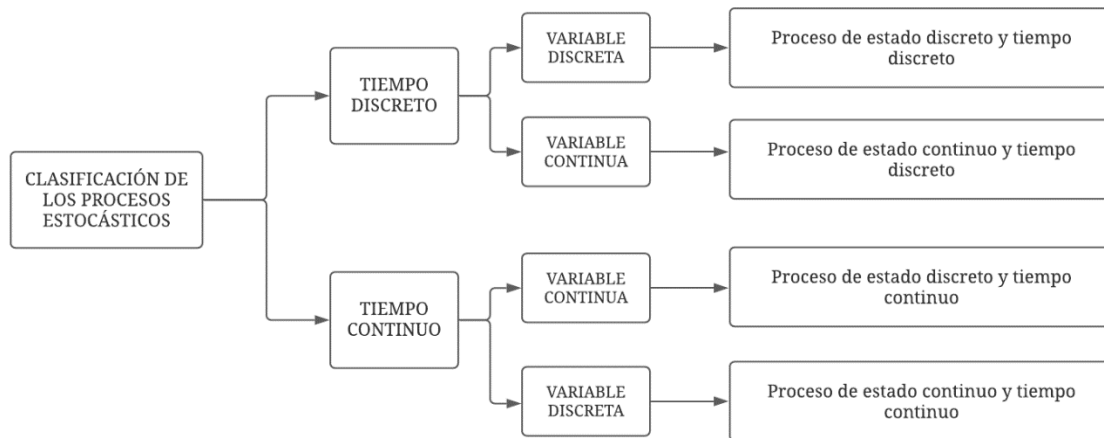


Figura 1-2 Clasificación de procesos estocásticos

Fuente: (Támara, 2019).

Elaborado por: Manuel Monar, 2022

2.2.2. Modelos estocásticos

2.2.2.1. Análisis Exploratorio de Datos

El análisis exploratorio de datos permite establecer el “Por qué” y “Para qué sirve” la información recolectada en los diferentes fases de la investigación, y consiste en el primer proceso de análisis y ordenamiento de la información estadística, mediante el conjunto de técnicas estadísticas con la finalidad de obtener un entendimiento básico de los datos y su representatividad de las variables analizadas, los métodos sistemáticos que permiten organizar y preparar los datos, encontrar errores y fallos de los datos dentro de su diseño, recolección y sistematización técnica conocida como **missing** y adicional la identificación de los casos atípicos que podrían modificar nuestras técnicas de medidas de centralización y medidas de distribución **outliers** y comprobación de los supuestos subyacentes en mayor parte de las técnicas multivariantes normalidad, linealidad, homocedasticidad (Támara, 2019).

Las etapas de la preparación de los datos consisten en los siguientes pasos

- a) Preparar los datos haciendo los accesibles a cualquier técnica estadística.
- b) Mediante el análisis descriptivo-numéricos de las variables individuales hacer un mapa gráfico.
- c) Mediante la relación de variables hacer un mapa grafico que permita el grado de interrelaciones existente entre ellas.
- d) Evaluar supuestos básicos sobre normalidad, linealidad y homocedasticidad de las variables de estudio.
- e) Identificación de casos atípicos y evaluar su impacto dentro del análisis estadístico.

2.2.2.2. *Estadística descriptiva.*

La estadística descriptiva es la rama de las matemáticas que tiene como finalidad la recolección , presentación y categorización del conjunto de datos u observaciones y describe apropiadamente las diversas características de ese conjunto, y este con junto de variables con valores numéricos que adoptan un carácter cuantitativo se les denomina variables estadísticas, y puede ser de dos tipo cualitativas o categóricas que no se pueden medir numéricamente ejemplo de ellos son las variables de nacionalidad, color de piel, sexo, y las cuantitativas denominadas si por su valor numérico dentro de su descripción ejemplo de ellos son las variables de edad, ingresos mensuales, precio del producto (Triola, 2013)

2.2.2.3. *No colinealidad o multicolinealidad*

En los modelos lineales múltiples los predictores son independientes, no debe haber colinealidad entre ellos, esto ocurre cuando los predictores están linealmente relacionados con uno u otros predictores del modelo, o a su vez son combinaciones lineales de otros productores. Como consecuencia de la colinealidad no se puede identificar de forma precisa el efecto individual en las variables colineales sobre la variable de respuesta, esto da como resultado el incremento de la varianza de los coeficientes de regresión estimados imposibilitando establecer la significancia estadística.

No existe un método estadístico para determinar la existencia de colinealidad o multicolinealidad entre los predictores de un modelo de regresión, sin embargo, hay reglas prácticas para tratar de determinar en qué medida afecta la estimación y contraste de los modelos.

- Si el coeficiente de determinación R^2 es alto, pero ninguno de los predictores resulta significativo, hay indicios de colinealidad.
- Calcular una matriz de correlación en la que se estudia la relación lineal entre predictores en pares, el no tener coeficientes de correlación altos no implica la inexistencia de multicolinealidad. Puede existir relación lineal casi perfecta entre tres o más variables y las correlaciones simples no sean mayores que 0.5.
- Generar un modelo de regresión lineal simple entre cada uno de los predictores frente al resto. Si en alguno de los modelos el coeficiente de determinación R^2 es alto, estaría señalando una posible colinealidad.
- Tolerancia (TOL) y factores de inflación de la varianza (VIF). Cálculo de los dos parámetros para la cuantificación el inverso del otro aplicando la siguiente fórmula.

$$VIF_{\hat{\beta}_j} = \frac{1}{1-R^2}$$

$$Tolerancia_{\hat{\beta}_j} = \frac{1}{VIF_{\hat{\beta}_j}}$$

Donde R^2 se obtiene de la regresión del predictor X_j sobre los otros predictores. Esta es la opción más recomendada, los límites de referencia que se suele emplear son:

VIF = 1 ausencia total de colinealidad

$1 < VIF < 5$ la regresión tiene cierta afectación de colinealidad

$5 < VIF < 10$ Precaución alta colinealidad

Los límites recomendados para la tolerancia de la colinealidad están recomendados entre 1 y 0.1

En el caso de encontrar colinealidad las posibles soluciones son, excluir uno de los predictores problemáticos conservando el predictor que más relevancia posee a juicio del investigador, la segunda opción consiste en combinar las variables colineales en un único predictor.

2.2.2.4. *Independencia.*

Verificar si los errores son independientes entre observaciones, se aplica la prueba de Durbin – Watson, este test puede variar entre 0 y 4, donde el valor cercano a 2 significa que los residuos no

están correlacionados, por otro lado, un valor superior a 2 indica una correlación positiva entre errores subyacentes, mientras que un valor menor a 2 representa una correlación negativa entre errores.

2.2.2.5. Distribución normal de los residuos

La distribución de los residuos debe ser de forma normal con media cero. Para comprobar se recurre a histogramas, a los cuartiles normales o test de hipótesis de normalidad.

2.2.2.6. Homocedasticidad.

Para comprobar homocedasticidad, la varianza de los residuos debe ser constante en todo el rango de observaciones, una distribución cónica es un claro identificador de falta de homocedasticidad. También se puede recurrir a contraste de homocedasticidad como el test de Breusch-Pagan.

2.2.6. Análisis de datos categóricos

Los datos categóricos miden a la variable en escala de categoría cualitativa o continua como por ejemplo variables como género, estado civil o afiliación política, también pueden clasificarse en variables dependientes o independientes, la escala de medida de la variable respuesta categórica es fundamental para la elección del análisis estadístico apropiado y estas pueden ser: dicotómicas, ordinales, nominales y de recuento. En cualquier área del conocimiento, es muy común encontrar situaciones en donde los datos recogidos son observaciones de variables categóricas cuyos niveles o categorías son empleados en la discriminación o identificación de las unidades muestrales en estudio (Arnau Gras, 1996).

Es importante indicar que la variable categórica es una característica para la cual la escala de medida consiste de un conjunto de categorías (di Rienzo et al., 2008). En esta situación los datos se presentan como frecuencias de observaciones que ocurren en la misma categoría (di Rienzo et al., 2008).

Dentro de la escala categórica se distinguen tres tipos principales de variables.

Nominales: son aquellas cuyos niveles no están naturalmente ordenados, por ejemplo, color del tegumento de semillas de maíz, variedad de un cultivo, raza de animales, etc. (Martín Fernández, 2001).

Ordinales: son aquellas cuyas distintas categorías tienen un orden natural, por ejemplo, grado de ataque de una plaga (sin ataque, controlable, no controlable), diagnóstico de enfermedades (seguro, probable, improbable), etc (Martín Fernández, 2001).

De intervalo: son aquellas variables de tipo numérico que tienen una distinción entre dos niveles, por ejemplo, edad de los individuos (entre 15-20,21-25 y 26,30 años) diámetro de los árboles (10-20,31-40 y 41-50 cm), etc (Martín Fernández, 2001).

Cabe mencionar que el análisis de datos categóricos se ocupa del estudio de variables que no son medibles (color, nacionalidad, enfermedades, sexo, afiliación política, etc.), denominadas también atributos o caracteres cualitativos (Arnau Gras, 1996). Podemos distinguir entre datos en escala nominal (sexo, estado civil, distintas ramas de actividad económica, profesión, ideología política, ...) y datos en escala ordinal (nivel de estudios, estratificación de familias por su capacidad de consumo, nivel de autoestima,), cuando podemos establecer un determinado orden o rango entre las observaciones (Díaz Monroy et al., 2018). En estos casos no tiene sentido el empleo de promedios, tales como la media aritmética; cuando las observaciones se nos ofrecen en una escala nominal, solo la moda puede utilizarse como medida resumen; y si éstas responden a una escala ordinal, podría determinarse, además del valor modal, también la mediana (di Rienzo et al., 2008). Una cuestión más interesante es el estudio de la existencia o no de asociación entre dos atributos, y de medidas similares a las de correlación para los casos en que variables no numéricas están relacionadas entre sí. Para atributos en escala nominal estableceremos los llamados coeficientes de contingencia (Martín Fernández, 2001). Cuando los caracteres estudiados pueden ordenarse de acuerdo con una cierta escala, es posible definir unos coeficientes de correlación que midan el grado de asociación entre ellos de manera parecida a como se mide la asociación entre variables cuantitativas; estos coeficientes están basados en los rangos u ordenes de las observaciones (Arnau Gras, 1996).

3.2.6.1. *Hipótesis y estimación.*

Los resultados de la regresión tengan la confiabilidad necesaria, es decir que los datos no sean insesgados, o que los resultados se acerquen a la realidad esperada, la varianza debe ser lo más mínima posible, para ello:

- a) La relación entre variables sea lineal. Con ello no significa que forzosamente tenga que ser una línea recta sino también la linealidad puede ser a través de alguna transformación.
- b) Las perturbaciones es decir las variables que provocan efectos al no estar incluida en el modelo deben ser de media cero, homocedástico y no auto correlacionadas. Dos formas de la estimación de una regresión lineal. Primero por mínimos cuadrados ordinarios (MCO), que consiste en resolver la ecuación.

$$\hat{b} = (X'X)^{-1}X'y$$

Donde \hat{b} es el vector de estimación de los coeficientes, y es el valor de la dependiente, X es la matriz de variables dependientes y X' es la respuesta de X. y la segunda formula esta dada por la máxima verosimilitud (ML). Que consiste en maximizar la ecuación.

$$\ln L(Y) = -\frac{n}{2} \ln(2\pi^2) - \frac{n}{2} \ln(2\sigma^2) - \frac{(Y-X\beta)'(Y-X\beta)}{2\sigma^2}$$

Derivando respecto de β y de σ^2 e igualando a 0. Ambos procedimientos llevan al mismo resultado, de complicarse la resolución, también puede resolverse por máxima verosimilitud.

2.2.6.2. *Tablas de contingencia*

Una tabla de contingencia o también conocida como tabla de clasificación cruzada es una forma común de resumir datos, es decir, se centra en estudiar si existe asociación entre una variable fila o una variable columna, calculando la intensidad de asociación. Se consideran de manera formal que X e Y dos variables categóricas con I y J categorías respectivamente (Sánchez R, 2005).

Cuando las casillas de una tabla contienen frecuencias observadas, la tabla se denomina tabla de contingencia dicho termino fue introducido en 1904 por Pearson, la tabla de contingencia recibe ese nombre cuando las características en estudio no son cuantitativas. Estas tablas se integran por filas o columnas mismas que forman celdas donde se registra la frecuencia absoluta o relativa de cada una de las categorías que se van a analizar.

Una ventaja de la tabla de contingencia es que ofrece información sistematizada y lista para así facilitar al investigador la búsqueda entre la relación de esas variables, debe ser utilizada para que su lectura se acerque a la explicación de la hipótesis por tanto es necesario que tenga un mismo formato para que así se pueda dar lectura a las columnas como variable independiente y a las filas como variable dependiente.

Tabla 1-2: Matriz de tabla de contingencia

		FACTOR 1				
FACTOR 2		NIVEL 1	NIVEL 2	...	NIVEL C	X_i
	NIVEL 1	X_{11}	X_{12}	...	X_{1c}	$X_{1.}$
		E_{11}	E_{12}	...	E_{1c}	$E_{1.}$
	NIVEL 2	X_{21}	X_{22}	...	X_{2c}	$X_{2.}$
		E_{21}	E_{22}	...	E_{2c}	$E_{2.}$
				...		
	NIVEL r	x_{r1}	x_{r1}	...	x_{rc}	$x_{r.}$
		E_{r1}	E_{r2}	...	E_{rc}	$E_{r.}$
	X.J	$x_{.1}$	$x_{.2}$...	$x_{.c}$	$x_{...=n}$
		$E_{.1}$	$E_{.2}$...	$E_{.c}$	

Las variables de clasificación son x e y, el tamaño de la muestra es n y los resultados se muestran por filas y columnas no métricas o categóricas, donde las entradas son las frecuencias de las respuestas que se incluyen en cada celda. Nos ofrecen información más precisa para poder explicar un fenómeno superando así a las tablas de distribución de frecuencia debido a que tiene un cruce de información entre las variables., además, son necesarias para un análisis estadístico porque

demuestran asociación entre variables. Existen medidas de asociación que indican el grado de relación asociativa que tienen las variables.

1. Medida de asociación es la Phi usada en tablas de 2 x 2 para escala nominal, está basada en Ji-cuadrado.

Toma valores entre 0 y 1 las cuales indican mínimo y máximo grado de asociación. El coeficiente de contingencia es una extensión de Phi, debido a que esta toma valores entre 0 y C, donde si r y c son el número de categorías de cada variable:

$$C(\text{maximo}) = \frac{\min(r - 1, c - 1)}{1} + \min(r - 1, c - 1)$$

2. Medida Lambda es una asociación donde sus parámetros son iguales a los anteriores el 0 y 1. Siendo una medida simétrica ya que mide la asociación sin saber cuál es la variable y asimétrica porque considera a cada una de las variables como la dependiente y da un valor de asociación. Incluso esta medida sirve para medir el tamaño de la asociación:

$$\lambda_{yx} = (\mathfrak{R}máx, f_{jk} - máx. f_k) / (n - máx. f_k)$$

2.2.6.3. Test Chi-cuadrado.

El test de Chi cuadrado proporciona un estadístico también conocido como X^2 o ji-cuadrado que permite constatar la hipótesis que los criterios utilizados de clasificación estas pueden ser dos variables categóricas son independientes, para ello se comprueba las frecuencias observadas y contratadas con las frecuencias esperadas estimando de igual manera con las frecuencias esperadas.

$$(\text{frecuencia esperada})_{ij} = \frac{(\text{total de la fila } i) * (\text{total de la columna } j)}{\text{numero total de casos}}$$

Donde:

i= a una fila cualquiera

j= una columna cualquiera

ij= una casilla cualquiera.

Bajo la condición de independencia a frecuencia esperada de una casilla se obtiene dividiendo el producto de las frecuencias marginales correspondientes a esa casilla por el número total de casos y el X^2 se obtiene de la siguiente manera.

$$x^2 = \sum_i \sum_j \frac{(n_{ij} - m_{ij})^2}{m_{ij}}$$

2.2.6.4. *Distribuciones marginales y condicionadas*

Distribuciones marginales

Una distribución marginal de x permite estudiar el comportamiento de la variable x con independencia de cómo se comporte con respecto a Y .

Se la define como las distribuciones por separado de cada una de las componentes de la variable aleatoria bidimensional.

La función de distribución conjunta de (X, Y) , $F(x,y)$ definiendo la distribución marginal de x como:

$$F_X(x) = P(X \leq x) = F(x, +\infty) = P(X \leq x, Y \leq +\infty), x \in \mathbb{R}$$

La función de distribución marginal de Y :

$$F_Y(y) = P(Y \leq y) = F(+\infty, y) = P(X \leq +\infty, Y \leq y), y \in \mathbb{R}$$

Caso discreto

Función masa de probabilidad marginal de X :

$$P_X(x_i) = P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j P_{ij} = P_{i.}$$

Función de masa de probabilidad marginal de Y :

$$P_Y(y_j) = P(Y = y_j) = \sum_i P(X = x_i, Y = y_j) = \sum_i P_{ij} = P_{.j}$$

Funciones de distribución marginal:

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i, Y = y_j) = \sum_{x_i \leq x} P_{i.}, x \in \mathbb{R}$$

$$F_Y(y) = P(Y \leq y) = \sum_{y_j \leq y} P(X = x_i, Y = y_j) = \sum_{y_j \leq y} P_{.j}, y \in \mathbb{R}$$

Caso continuo

Funciones de densidad marginales de X :

$$f_x(x) = \int_{-\infty}^{+\infty} f(x, y) dy, x \in \mathbb{R}$$

Funciones de densidad marginales de Y:

$$f_y(y) = \int_{-\infty}^{+\infty} f(x, y) dx, y \in \mathbb{R}$$

Funciones de distribución marginal:

$$F_X(x) = P(X \leq x) = P(X \leq x, Y \in \mathbb{R}) = \int_{-\infty}^x \int_{-\infty}^{+\infty} f(u, v) dudv, x \in \mathbb{R}$$

$$F_Y(y) = P(Y \leq y) = P(X \in \mathbb{R}, Y \leq y) = \int_{-\infty}^{+\infty} \int_{-\infty}^y f(u, v) dudv, y \in \mathbb{R}$$

Distribuciones condicionadas

Las distribuciones condicionales se obtienen al fijar el valor de otra variable.

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}, P(B) \neq 0$$

Caso discreto

Función masa de probabilidad de X condicionada al valor de yj de Y:

$$P_{X/Y}(x_i/y_j) = p_{i/j} = P(X = x_i/Y = y_j) = \frac{P(X = x_i \cap Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{.j}}$$

Función masa de probabilidad de Y condicionada al valor xi de X:

$$P_{Y/X}(y_j/x_i) = p_{j/i} = P(Y = y_j/X = x_i) = \frac{P(X = x_i \cap Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_{i.}}$$

Caso continuo

Función de densidad de X dado que Y vale y=

$$f(x/y) = \frac{f(x, y)}{f_Y(y)}, x \in \mathbb{R}$$

Función de densidad de Y dado que X vale x:

$$f(y/x) = \frac{f(x, y)}{f_X(x)}, y \in \mathbb{R}$$

Funciones de distribución asociadas:

$$F(x/y) = P(X \leq x/Y = y) = \int_{-\infty}^x \frac{f(x, y)}{f_Y(y)} dx, x \in \mathbb{R}$$

$$F(y/x) = P(Y \leq y/X = x) = \int_{-\infty}^y \frac{f(x, y)}{f_X(x)} dy, y \in \mathbb{R}$$

2.2.6.5. *Análisis de correlación de Pearson.*

El indicador de correlación de Pearson, desinado para variables cuantitativas es un índice que mide el grado de covariación entre las distintas variables que se relacionan linealmente esto implica la relación fuertemente lineal entre algunas de ellas, el coeficiente de correlación es un índice de fácil ejecución e interpretación debido que se obtiene valores absolutos entre cero y uno, esto es si se tiene dos variables X y Y y se define como la relación r_{xy} entonces:

$$0 \leq r_{xy} \leq 1$$

Se ha especificado los términos valores absolutos, pero si se contempla el intervalo para el coeficiente de correlación de Pearson entre -1 y +1, es también importante que la magnitud de la relación viene dada por el signo del valor numérico, en este sentido una relación + 1 la correlación es perfecta positiva y una correlación -1 es perfecta negativa, la fórmula utilizada para el coeficiente de correlación de Pearson viene dada por la siguiente expresión:

$$r_{xy} = \frac{\sum z_x z_y}{N}$$

Esto es, el coeficiente de correlación de Pearson hace referencia a la media de los productos cruzados de las puntuaciones estandarizadas de X y Y, su valor oscila entre 0 y 1, las puntuaciones estandarizadas muestran precisamente la posición en desviaciones tipo de un individuo respecto a su media y como el individuo está separado respecto a su media, cuando las sus puntuaciones son iguales ($z_x = z_y$) en este caso la fórmula de la correlación se transforma en:

$$r_{xy} = \frac{\sum z_x z_y}{N} = \frac{\sum z_x z_x}{N} = \frac{\sum z_x^2}{N} = 1$$

Donde: z_x equivale a la varianza, que como se sabe vale 1.

Cuando la correlación es perfecta negativa los valores de z_x y z_y , son exactamente iguales, pero de signo contrario, resultado de los productos cruzados de z_x y z_y negativos, en este caso, el valor de la correlación es el mismo, pero con signo negativo.

$$r_{xy} = \frac{\sum z_x z_y}{N} = \frac{\sum z_x z_x}{N} = \frac{\sum z_x^2}{N} = 1$$

Cuando la correlación es nula, para un valor obtenido de X se podrá obtener cualquier valor de Y, es decir un valor determinado de z_x la misma cantidad de valores positivos y negativos de z_y resultado de ello la suma de productos cruzados valdrá cero.

$$r_{xy} = \frac{\sum z_x z_y}{N} = 0$$

$$r_{xy} = \frac{\frac{\sum xy}{N} - \bar{x}\bar{y}}{S_x S_y}$$

Donde:

$$\begin{aligned} r_{xy} &= \frac{\sum z_x z_y}{N} = \frac{\sum \frac{x - \bar{x}}{S_x} * \frac{y - \bar{y}}{S_y}}{N} = \frac{\sum (x - \bar{x})(y - \bar{y})}{N S_x S_y} = \frac{\sum (xy - x\bar{y} - \bar{x}y + \bar{x}\bar{y})}{N S_x S_y} \\ &= \frac{1}{S_x S_y} \left(\frac{\sum xy}{N} - \frac{\bar{y} \sum x}{N} - \frac{\bar{x} \sum y}{N} + \frac{N \bar{x}\bar{y}}{N} \right) = \frac{1}{S_x S_y} \left(\frac{\sum xy}{N} \bar{x}\bar{y} - \bar{x}\bar{y} + \bar{x}\bar{y} \right) = \frac{\frac{\sum xy}{N} - \bar{x}\bar{y}}{S_x S_y} \end{aligned}$$

2.2.6.6. *Distribuciones marginales*

Una distribución marginal de x permite estudiar el comportamiento de la variable x con independencia de cómo se comporte con respecto a Y. Se la define como las distribuciones por separado de cada una de las componentes de la variable aleatoria bidimensional. la función de distribución conjunta de (X, Y), F(x,y) definiendo la distribución marginal de x como:

$$F_X = (x) = P(X \leq x) = F(x, +\infty) = P(X \leq x, Y \leq +\infty), x \in \mathbb{R}$$

La función de distribución marginal de Y:

$$F_Y = (y) = P(Y \leq y) = F(+\infty, y) = P(X \leq +\infty, Y \leq y), y \in \mathbb{R}$$

Caso discreto

Función masa de probabilidad marginal de X:

$$P_X(x_i) = P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j P_{ij} = P_{i.}$$

Función de masa de probabilidad marginal de Y:

$$P_Y(y_j) = P(Y = y_j) = \sum_i P(X = x_i, Y = y_j) = \sum_i P_{ij} = P_{.j}$$

Funciones de distribución marginal:

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i, Y = y_j) = \sum_{x_i \leq x} P_{i.}, x \in \mathbb{R}$$

$$F_Y(y) = P(Y \leq y) = \sum_{y_j \leq y} P(X = x_i, Y = y_j) = \sum_{y_j \leq y} P_{.j}, y \in \mathbb{R}$$

Caso continuo

Funciones de densidad marginales de X:

$$f_x(x) = \int_{-\infty}^{+\infty} f(x, y) dy, x \in \mathbb{R}$$

Funciones de densidad marginales de Y:

$$f_y(y) = \int_{-\infty}^{+\infty} f(x, y) dx, y \in \mathbb{R}$$

Funciones de distribución marginal:

$$F_X(x) = P(X \leq x) = P(X \leq x, Y \in \mathbb{R}) = \int_{-\infty}^x \int_{-\infty}^{+\infty} f(u, v) du dv, x \in \mathbb{R}$$

$$F_Y(y) = P(Y \leq y) = P(X \in \mathbb{R}, Y \leq y) = \int_{-\infty}^{+\infty} \int_{-\infty}^y f(u, v) du dv, y \in \mathbb{R}$$

Distribuciones condicionadas

Las distribuciones condicionales se obtienen al fijar el valor de otra variable.

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}, P(B) \neq 0$$

Caso discreto

Función masa de probabilidad de X condicionada al valor de y_j de Y:

$$P_{X/Y}(x_i/y_j) = p_{i/j} = P(X = x_i/Y = y_j) = \frac{P(X = x_i \cap Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{.j}}$$

Función masa de probabilidad de Y condicionada al valor xi de X:

$$P_{Y/X}(y_j/x_i) = p_{j/i} = P(Y = y_j/X = x_i) = \frac{P(X = x_i \cap Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_{i.}}$$

Caso continuo

Función de densidad de X dado que Y vale y=

$$f(x/y) = \frac{f(x, y)}{f_Y(y)}, x \in \mathbb{R}$$

Función de densidad de Y dado que X vale x:

$$f(y/x) = \frac{f(x, y)}{f_X(x)}, y \in \mathbb{R}$$

Funciones de distribución asociadas:

$$F(x/y) = P(X \leq x/Y = y) = \int_{-\infty}^x \frac{f(x, y)}{f_Y(y)} dx, x \in \mathbb{R}$$

$$F(y/x) = P(Y \leq y/X = x) = \int_{-\infty}^y \frac{f(x, y)}{f_X(x)} dy, y \in \mathbb{R}$$

2.2.7. Análisis de Componentes principales categóricas (CAPTCA)

Cuando se trata de hacer una linealidad de variables cualitativas, uno de los métodos a utilizar para el análisis de causalidad o correlación mediante la materialización es el método de componentes principales categóricas, para ello debemos suponer que se tiene variables X_1, X_2, \dots, X_n las cuales son dicotómicas, basadas estas en ausencia o presencia de caracteres cualitativos. Se obtiene un individuo W_i quedando caracterizado por la presencia o ausencia de n caracteres:

El objetivo de los análisis de componentes principales es la reducción de un conjunto original de variables en un conjunto más pequeño de componentes no correlacionados que representen la mayor parte de la información encontrada en las variables originales (IBM Corporation, 2021). La técnica es más útil cuando un extenso número de variables impide una interpretación eficaz de las relaciones entre los objetos (sujetos y unidades). Al reducir la dimensionalidad, se interpreta un pequeño número de componentes en lugar de un extenso número de variables. (Díaz Monroy et al., 2018). El análisis estándar de componentes principales asume relaciones lineales entre las variables numéricas; por otra parte, el método de escalamiento óptimo permite escalar las variables a diferentes niveles (IBM Corporation, 2021). Las variables categóricas se cuantifican de forma óptima en la dimensionalidad especificada. Como resultado, se pueden modelar relaciones no lineales entre las variables (di Rienzo et al., 2008).

Las variables X_1, \dots, X_n mismas que son dicotómicas basadas en ausencia de $-$ o presencia de $+$, con un individuo W_i queda caracterizado como:

$$X_1 \ X_2 \ X_3 \ \dots \ X_n$$

$$\omega_i \ + \ - \ - \ \dots \ +$$

El número de caracteres presentes sobre n :

$$n = a + b + c + d$$

La asociación se mide mediante un coeficiente de similaridad S_{ij} :

$$S_{ij} = f(a, b, c)$$

Varios autores propusieron coeficientes de similaridad con estas propiedades:

Sokal y Michener $\frac{a+d}{n}$

Sokal y Sneath $\frac{a}{a+2(b+c)}$

Jaccard $\frac{a}{a+b+c}$

Rusell y Rao $\frac{a}{b}$

Dice $\frac{2a}{(a+b)(a+c)}$

Kulezynski $\frac{a}{b+c}$

Harman $1 - \frac{2(b+c)}{n}$

Roger y Tanimoto $\frac{n-(b+c)}{n+(b+c)}$

Ochial $\frac{a}{\sqrt{(a+b)}\sqrt{(a+c)}} n$

Hay otros coeficientes que expresan la dependencia estocástica entre W_i y W_j , el valor 0 se identifica como noción de independencia estocástica que existe entre las variables.

Ejemplo:

Yule $\frac{ad-bc}{ad+bc}$

Pearson $\frac{ad-dc}{[(a+b)(b+d)(a+d)(c+d)]^{\frac{1}{2}}}$

Para definir el coeficiente de similaridad se recoge una matriz simétrica:

$$\mathbf{S} = \begin{pmatrix} S_{11} & S_{12} & \dots & S_{1N} \\ S_{21} & S_{22} & \dots & S_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ S_{N1} & S_{N2} & \dots & S_{NN} \end{pmatrix}$$

Para la elección de un determinado coeficiente depende del tipo de datos que se va a representar y el peso que se quiere dar a las frecuencias a, b, c, d.

Por tanto, es necesario tener en cuenta que:

- a. S_{ij} no debe ser función de d.
- b. El coeficiente debe ser tal que la matriz S
- c. El coeficiente debe verificar la propiedad ultra métrica.

2.2.7.1. *Coordenadas principales categóricas construidas sobre una similaridad.*

Tienen por objeto realizar una clasificación ejemplo: especies, géneros, familias, sobre los individuos W_1, \dots, W_n tomando como información la matriz de similaridad S .

Trata de obtener una representación geométrica de los N individuos que se presentan en relación a una distancia razonablemente compatible con la similaridad que existe entre las variables.

Distancia al cuadrado entre W_i y W_j :

$$d_{ij}^2 = d^2(W_i + W_j) = S_{ii} + S_{jj} - 2S_{ij}$$

1. $d^2(W_i, W_j) = 0 \quad i: 1, \dots, \dots, N$
2. $d^2(W_i, W_j) = d^2(W_j, W_i). \forall i, j$ la distancia es simétrica.
3. Si S_{ij} es una similaridad, $0 \leq S_{ij} \leq 1$, entonces

$$d_{ij}^2 = 2(1 - S_{ij})$$

4. Si S_{ij} es una correlación $-1 \leq S_{ij} \leq +1$, entonces d varía de 0.
5. Identificando S_{ij} como un producto escalar entre W_i y W_j , se corresponde formalmente con la norma al cuadrado de $(W_i - W_j)$.

$$|W_i - W_j|^2 = W_i * W_i + W_j * W_j - 2W_i * W_j = S_{ii} + S_{jj} - 2S_{ij}$$

Al ser estas variables dicotómicas, los individuos no se pueden representar directamente con puntos en el espacio R_n . El método de coger consiste en encontrar una matriz de datos que defina una configuración de puntos en R_n de manera que su distancia euclídea coincida con la distancia d_{ij} .

Para ello vamos ahora a construir una tabla de datos que defina unas coordenadas para los individuos, en el espacio euclídeo R_n , compatible con la distancia d_{ij} .

$$Su_i = \lambda_i u_i$$

Siendo $Su_i = (u_{1i}, u_{2i}, \dots, u_{Ni})'$ vector propio λ_i normalizado es decir $\sum_{h=1}^N u_{hi}^2 = \lambda_i$

Se asigna a los individuos $\{W_i\}$ las siguientes coordenadas, formadas por los vectores de S ,

$$S_{ij} = \sum_{h=1}^N u_{ih}u_{jh}$$

La distancia 2 euclidea entre los puntos W_i y W_j :

$$D_{ij}^2 = \sum_{h=1}^N (u_{ih}u_{jh})^2 = \sum_{h=1}^N u_{ih}^2 + \sum_{h=1}^N u_{jh}^2 - 2u_{ih}u_{jh} = S_{ii} + S_{jj} - 2S_{ij}$$

Coinciden exactamente con las distancias D_{ij}^2 definidas por la similaridad.

El paso siguiente del análisis consistiría en representar los elementos de N puntos en un espacio de dimensión $d < N$ con dispersión máxima.

La dispersión de las columnas desciende de izquierda a derecha. Además, si S es definida positiva, el rango de.... Es $(N-1)$, los N puntos están en una variedad lineal de dimensión $(N-1)$ y la última columna de la matriz W tiene todos los elementos iguales. Si S es semidefinida positiva, esto también ocurre con otras columnas anteriores a la última. Las filas de la matriz W se la llaman matriz de componentes principales categóricas de los individuos W_1, W_2, \dots, W_n .

$$\begin{array}{l} W_1 \quad (W_{11} \quad W_{12}) \\ W_2 \quad (W_{21} \quad W_{22}) \\ \dots \\ W_N \quad (W_{N1} \quad W_{N2}) \end{array}$$

La representación gráfica tiene máxima resolución en dos dimensiones, en el sentido del segundo teorema fundamental.

El análisis de componentes principales categóricas es el análisis de componentes principales aplicado a una tabla de datos compatible con la distancia.

$$d_{ij}^2 = S_{ii} + S_{jj} - 2S_{ij}$$

2.2.7.2. Algoritmo Para Obtener Componentes Principales Categóricas.

Las dos diagonalizaciones presentadas anteriormente pueden reducirse a una sola. Transformando la matriz de asociaciones en otra matriz $T = (t_{ij})$:

$$t_{ij} = S_{ij} + S_i + S_j + S$$

La media se verifica de la siguiente manera:

$$d_{ij}^2 = S_{ii} + S_{jj} - 2S_{ij} = t_{ii} + t_{jj} - 2t_{ij}$$

La matriz T es simétrica y se verifica:

- a) La suma de toda la columna y de toda la fila de T es 0.
- b) $\text{Rang}(T) = \text{rang}(S) - 1$
- c) T tiene al menos un valor propio nulo.

Resumiendo para obtener las coordenadas principales se efectúan los siguientes pasos:

- 1. Obtención de la matriz similaridad S .
- 2. Transformación de S en la matriz T .

$$t_{ij} = S_{ij} - S_i - S_j - S$$

- 3. Diagonalización de la matriz T

$$T = W * W^0$$

$$U_1 > U_2 \dots \dots \dots U_n = 0$$

- 4. Las coordenadas principales, para una representación en dimensiones d , son las d primeras columnas de W .

La dispersión de los N individuos en un subespacio de dimensión d es:

$$\sum_{i,j=1}^N d_{ij}^2 = 2 N (U_1 + U_2 \dots \dots U_d)$$

El análisis de coordenadas principales permite representar objetos con referencia a una distancia deducida de un coeficiente de similaridad. Además, representa distancias con máxima resolución en dimensión reducida en lugar de representar direcciones.

El criterio de este análisis es maximizar la suma de cuadrados de las distancias.

Es un método de representación de datos por análisis de componente principal sobre la tabla de datos W_{ij} , obtenida por diagonalización de la matriz T .

2.2.7.3. *Aplicaciones y Ejemplos del ACPC.*

El análisis de Componentes principales Categóricas (ACPC) es, en realidad, un método métrico de Análisis de Proximidades (AP), técnica de análisis de datos que había sido introducida mucho antes por forgeron (1952,1958), desarrollada por Shepard (1962) y refinada por Kruskal (1964) y otros actores (Guttman, Benze-cri, LIngoes, Carroll, Young). Sin embargo, como observa Gower (1966), El ACPC es mucho más sencillo de cálculo que el AP, que exige resolver un costoso proceso iterativo.

El ACPC es una técnica que ha tenido una notable aceptación y de la que se han hecho numerosas aplicaciones, debido, principalmente, a los motivos siguientes:

1. El ACPC se reduce a la diagonalización de la matriz T , deducida de la matriz de similaridad S . La solución está algebraicamente bien determinada.
2. Cuando S tiene una buena estructura métrica (en el sentido de que es definida positiva, o en todo caso, los valores propios negativos son de escasa relevancia), describe bien la estructura de datos.

La solución en AP es tan solo numérica, resultado de un proceso iterativo. En general no puede describirse explícitamente en función de los datos. El AP solamente supera al ACPC cuando la información es claramente no métrica u ordinal.

2.2.7.4. *Representación de datos del ACPC.*

Las posibilidades del ACPC como método de clasificación en Biología agrupación de especies, razas, poblaciones, etc. han sido resaltadas por Pielou (1969), Blackith y Reyment (1971), Dallot e Ibanez (1972) y otros.

El ACPC destaca como método de representación de datos por el hecho de admitir cualquier tipo de variable, incluyendo las dicotómicas (específicamente que presente problemas sistemática biológica), bastando calcular una matriz de similaridad apropiada.

2.2.7.5. *Análisis de matrices de correlaciones con muchas variables*

El ACPC es utilizado como método de agrupación de variables para eliminar algunas de ellas como fase previa a un análisis factorial o para formar grupos de variables estrechamente relacionadas.

La distancia:

$$d^2(X_i, X_j) = 2(1 - r_{ij})$$

El ACPC nos proporciona una configuración euclídea P_1, \dots, P_n , en dimensión m , con la propiedad de que dos variables X_i y X_j aparecen tanto más próximas cuanto mayor es su correlación r_{ij} .

Esto nos permite definir grupos de variables fuertemente correlacionadas entre sí.

2.2.7.6. *Representación de individuos cuyas interdistancias son conocidas.*

Sean W_1, \dots, W_n individuos de los que se conocen algún tipo de distancia:

$$d_{ij} = d(W_i, W_j)$$

El ACPC permite obtener N puntos del espacio euclídeo R_m .

$$P_i = (x_{i1}, x_{jm})$$

Cuyas interdistancias verifiquen:

$$d^2(P_i, P_j) = \sum_{h=1}^m (x_{ih} - x_{jh})^2 = d_{ij}^2$$

La similaridad es:

$$S_{ij} = S_{jj} = 0, \quad S_{ij} = \frac{1}{2} d_{ij}^2$$

Es evidente que:

$$d_{ij}^2 = S_{ii} + S_{jj} - 2S_{ij}$$

Por tanto, se obtendrán las principales coordenadas de los individuos W_1, \dots, W_n . La dimensión es:

$$m = \text{rang}(T) = \text{rang}(s) < 1$$

Ejemplo. El análisis de componentes principales categórico se puede utilizar para representar gráficamente la relación entre la categoría laboral, la división laboral, la provincia, el número de desplazamientos (alto, medio y bajo) y la satisfacción laboral. Observará que con dos dimensiones se puede explicar una gran cantidad de varianza. La primera dimensión podría separar la categoría laboral de la provincia, mientras que la segunda dimensión podría separar la división laboral del número de desplazamientos. También podrá observar que la alta satisfacción laboral está relacionada con un número medio de desplazamientos (IBM Corporation, 2021).

Estadísticos y gráficos. Frecuencias, valores perdidos, nivel de escalamiento óptimo, modo, varianza explicada por: las coordenadas del centroide, las coordenadas de vector, total por variable y total por dimensión; cargas en componentes para las variables cuantificadas por los vectores, cuantificaciones y coordenadas de categoría, historial de iteraciones, correlaciones entre las variables transformadas y los autovalores de la matriz de correlaciones, correlaciones entre las variables originales y los autovalores de la matriz de correlaciones, puntuaciones de objetos, gráficos de categorías, gráficos de categorías conjuntas, gráficos de transformación, gráficos de residuos, gráficos de centroides proyectados, gráficos de objetos, diagramas de dispersión biespaciales, diagramas de dispersión triespaciales y gráficos de las cargas en componentes (IBM Corporation, 2021).

Consideraciones sobre los datos del Análisis de componentes principales categórico

Datos. Los valores de las variables de cadena se convierten en enteros positivos por orden alfabético ascendente. Los valores perdidos del usuario, los valores perdidos del sistema y los valores menores que 1 se consideran valores perdidos; se puede añadir una constante o recodificar las variables con valores inferiores a 1 para evitar que se pierdan los mismos (Arnau Gras, 1996).

Supuestos. Los datos deben contener al menos tres casos válidos. El análisis se basa en datos enteros positivos. La opción de discretización categorizará de forma automática una variable con valores fraccionarios, agrupando sus valores en categorías con una distribución casi “normal” y convertirá de forma automática los valores de las variables de cadena en enteros positivos. Se pueden especificar otros esquemas de discretización (di Rienzo et al., 2008).

Procedimientos relacionados. Si se escalan todas las variables a nivel numérico, el análisis se corresponderá con el análisis de componentes principales estándar. Hay características de representación alternativas que están disponibles si se utilizan las variables transformadas en un análisis de componentes principales lineal estándar. Si todas las variables tienen un nivel de escalamiento nominal múltiple, el análisis de componentes principales categórico es idéntico al análisis de correspondencias múltiple (HOMALS). Si hay conjuntos de variables que son de interés, se debe utilizar el análisis de correlación canónica categórico (no lineal) (IBM Corporation, 2021).

CAPÍTULO III

3. METODOLOGÍA

3.1. Tipo y diseño de investigación.

El tipo de investigación que se realizara es cuantitativas, porque basara su enfoque en variables de tipo numérica y nominal para la elaboración de los modelos estocásticos, para lo cual se hará un primer clasificación de las variables en cuanto a grupos según los clúster y se culminara con la elaboración de modelos lineales predictivos, todo esto basado en los métodos de investigación como el inductivo, deductivo y analítico para determinar los factores que inciden el en la propagación de personas con diabetes tipo II en la ciudad de Guaranda. Además, que hay que mencionar el enfoque descriptivo observacional debido a que se trabajara con las historias clínicas de los pacientes de diabetes.

Por otro lado, el examinar el comportamiento de los datos con el uso de la estadística estocástica a través de los modelo lineales y binomiales, esto permitirá dar respuesta al crecimiento de personas que padecen la enfermedad y podrá hacer una predicción pronostico en el tiempo, lo que se estará dando respuesta a la problemática planteada en la problemática de la investigación

3.2. Métodos de investigación.

La presente investigación para su análisis y descomposición del problema que es el crecimiento de personas que padecen la enfermedad de diabetes mellitus tipo II en la ciudad de Guaranda utilizará el método inductivo, donde descompondrá en parte muy pequeñas para identificar los factores de crecimiento, y el método deductivo analítico para formular los modelos estocásticos

3.3. Enfoque de la Investigación.

El enfoque que tomara la investigación es de cuantitativa porque permitirá a partir de una clasificación de variables nominales y ordinales la formulación de modelos estocásticos de

crecimiento y predictoras, que son sistemas dinámicos lineales y categorizados a través de clúster en base a las distancias entre datos.

3.4. Alcance de la Investigación.

El presente trabajo de investigación tendrá un alcance correlacional porque busca medir la interacción de las variables de factor de riesgo modificables y no modificables sobre las personas que tiene diabetes y medir su crecimiento de afectación. Este trabajo también tendrá un alcance de tipo explicativo porque plantearan las condiciones para darse el crecimiento de la afectación de las personas con diabetes mellitus tipo II.

3.5. Población de Estudio.

En esta investigación se trabajó con los datos de los pacientes que se hacen atender en la red medica del Ministerio de Salud Distrito Bolívar. Para la selección de los datos se utilizó la base de registros de personas de salud que integral el total de personas que se hacen atender en el Servicio de Salud Pública del Ecuador de la Plataforma de Registro de Atención de Salud (PRAS) 2020.

3.6. Tratamiento de los Datos.

Una vez visualizados los datos que consta la base que está estructurada de 680 observaciones y 17 variables además se aplicó un cuestionario compuesto de 5 preguntas que permitieron levantar información sobre hábitos y consumos de los pacientes, esta información no constaba en la base de datos, mediante la aplicación de estadística descriptiva e inferencial se determinó la normalidad de los datos mediante los histogramas de cada una de las variables.

La técnica que permitió la reducción de las variables y los factores mediante aplicación de la herramienta análisis de componentes principales para datos categóricos (CAPTCA) en el software SPSS, que usa un modelo estadístico basado el cálculo mediante una matriz por auto vectores y autovalores para determinar regresiones lineales e identifica variables y factores donde se determinó un perfil del paciente con más afectación en el crecimiento.

El análisis de componentes principales para datos categóricos mediante la escala optima genera el procedimiento de los principales factores y componentes que permitió acomodar las variables de mediciones mixtas nominales y ordinales y clasificarlas de tal modo que se puedo establecer los

factores principales de crecimiento y los perfiles de los pacientes con diabetes mellitus tipo 2, se mostró las relaciones entre grupos y factores por medio de la correlación de variables.

El CATPCA permite la cuantificación óptima de las variables sociodemográficas, de comportamiento y hábitos de consumo de pacientes con diabetes mellitus tipo 2 y permitió obtener mapas de dimensiones reducidas para identificar las relaciones de los factores.

CAPÍTULO IV

4. RESULTADOS SEGÚN OBJETIVOS PLANTEADOS

En este capítulo se presentan los resultados obtenidos sobre el ordenamiento, presentación estadística descriptiva de las variables en relación al objeto de estudio de esta investigación, así como cálculo y modelación de los métodos estocásticos.

4.1 Descripción estadística de factores que inciden en el crecimiento mediante análisis exploratorio de datos

Dentro de la caracterización se identificó mediante un mapa georreferenciado los casos de personas con diabetes en la ciudad de Guaranda, también se realizó una descripción exploratoria de datos, donde se identificó estadísticos principales como la media aritmética, mediana, moda, desviación estándar, curtosis y asimetría con y con esto se determinó la normalidad de los datos de las variables divididas en diferentes subgrupos nominales como el lugar donde se realizaron la atención médica, sexo de los pacientes, tipo de localidad, frecuencia en la atención y clasificación del índice de masa corporal.

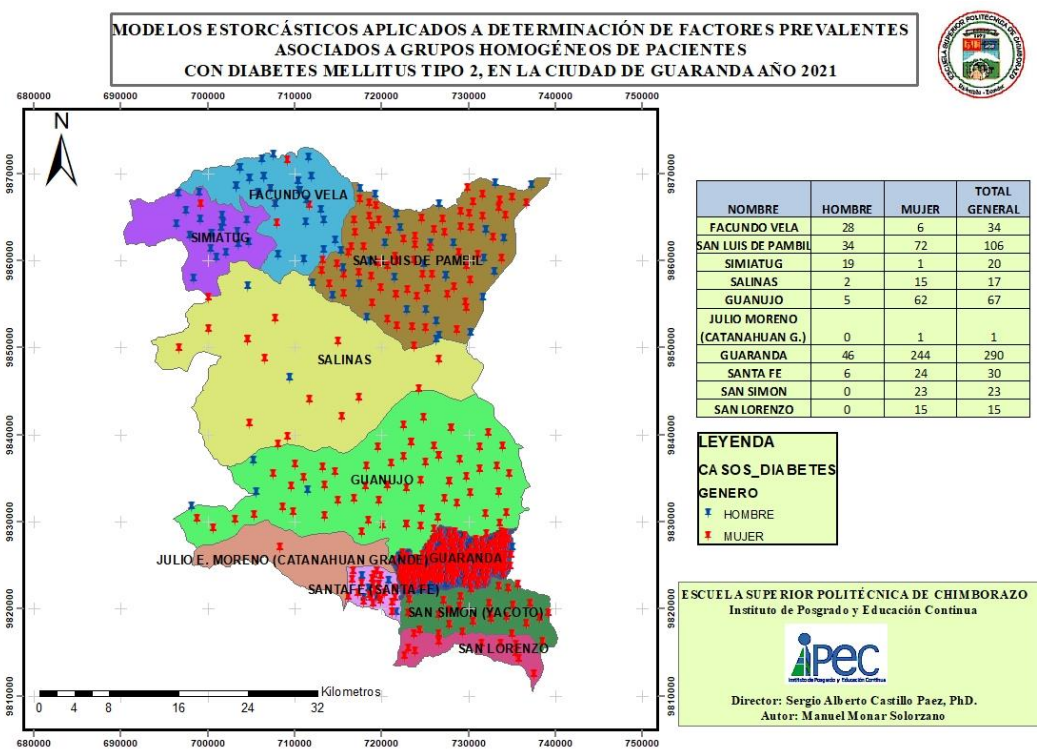


Figura 2-4: Mapa de casos con diabetes en la ciudad de Guaranda

Realizado por: Monar Manuel, 2022

Tabla 2-4: Análisis exploratorio de la variable
 Lugar de atención medica de personas que
 usan la red de salud publica

N	Válido	627
	Perdidos	0
Media		3,70
Mediana		2,00
Moda		2
Desv. Desviación		2,370
Varianza		5,617
Asimetría		,735
Error estándar de asimetría		,098
Curtosis		-,907
Error estándar de curtosis		,195

Realizado por: Monar Manuel, 2022

Tabla 3-4: Lugar de Atención según los Centros y Subcentros de la Red Pública de Salud

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido Facundo Vela	42	6,7	6,7	6,7
Guaranda	334	53,3	53,3	60,0
Salinas	17	2,7	2,7	62,7
San Lorenzo	18	2,9	2,9	65,6
San Luis Pambil	141	22,5	22,5	88,0
San Simón	21	3,3	3,3	91,4
Santa Fé	29	4,6	4,6	96,0
Simiuatug	25	4,0	4,0	100,0
Total	627	100,0	100,0	

Realizado por: Monar Manuel, 2022

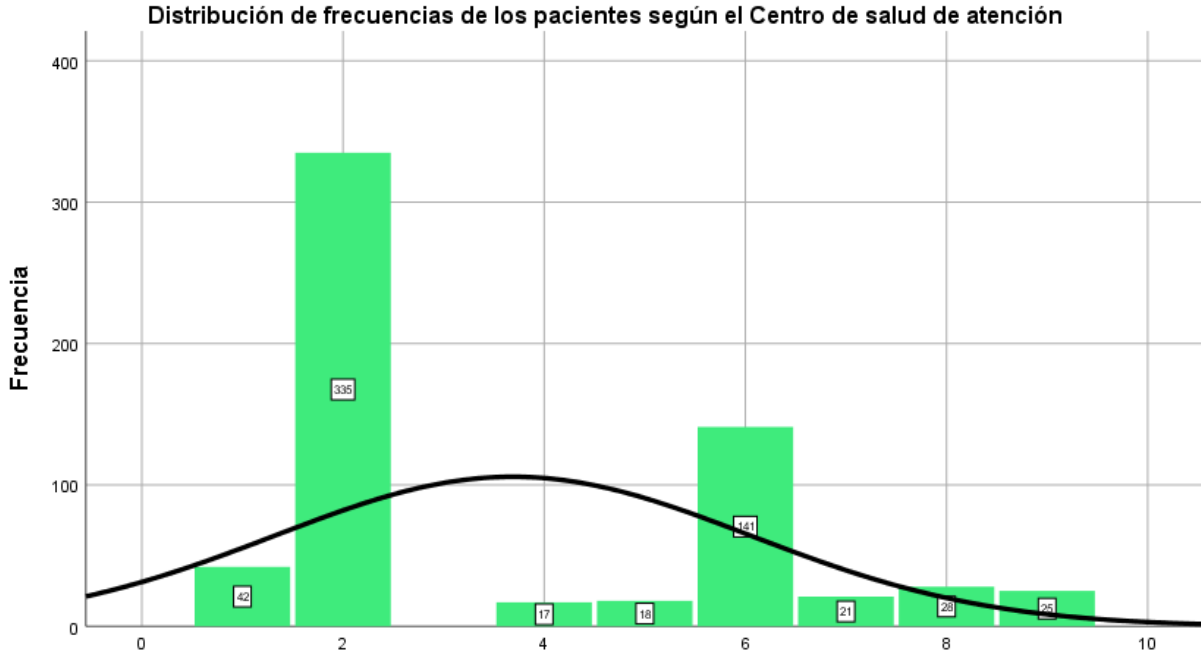


Figura 3-4: Curva de distribución de frecuencias de la atención de pacientes en Centros y Subcentros de la red pública de salud

Realizado por: Monar Manuel, 2022

De acuerdo con la tabla 4 y la gráfica 3 el centro de Salud de Guaranda tiene el mayor porcentaje de atención médica con el 53.3%, seguido del centro de salud ubicado en la parroquia de San Luis de Pambil con 22,5%, y la parroquia con menos atenciones médicas es la parroquia de San Simón con 3,3%. Con una simetría positiva de 0,735 lo que implica una su inclinación a la derecha con respecto a la media.

Tabla 4-4: Clasificación de los usuarios por su localidad de residencia

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Rural	334	53,3	53,3	53,3
	Urbano	293	46,7	46,7	100,0
	Total	627	100,0	100,0	

Realizado por: Monar Manuel, 2022

Tabla 5-4: Análisis exploratorio de datos de la variable localidad de la población

N	Válidos	627
	Perdidos	0
Media		1,47
Mediana		1,00
Moda		1
Desv. Desviación		,499
Varianza		,249
Asimetría		,131
Error estándar de asimetría		,098
Curtosis		-1,989
Error estándar de curtosis		,195

Realizado por: Monar Manuel, 2022

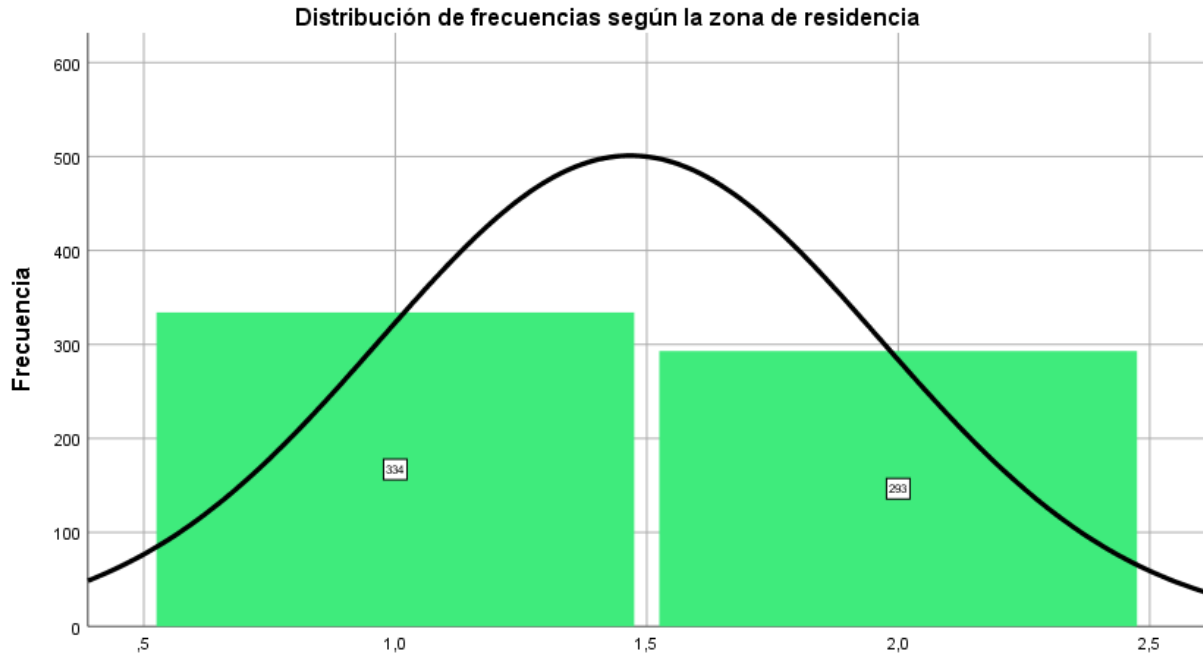


Figura 4: Curva de distribución de la variable localidad de la población

Realizado por: Monar Manuel, 2022

La tabla 5 muestra el porcentaje según la zonificación de los usuarios donde el 53,3% corresponde a usuarios del sector rural, y el 46,7 representa al sector urbano, la figura 4 muestra una distribución de los datos con una simetría de 0,131 lo que implica una asimetría positiva con inclinación a la derecha cercana a la media aritmética.

Tabla 6-4: Análisis exploratorio de datos de la variable sexo del paciente

N	Válido	627
	Perdidos	0
Media		1,36
Mediana		1,00
Moda		1
Desv. Desviación		,480
Varianza		,230
Asimetría		,590
Error estándar de asimetría		,098
Curtosis		-1,657
Error estándar de curtosis		,195

Realizado por: Monar Manuel, 2022

Tabla 7-4: distribución de frecuencias de la variable sexo del paciente

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Hombre	402	64,1	64,1	64,1
	Mujer	225	35,9	35,9	100,0
	Total	627	100,0	100,0	

Realizado por: Monar Manuel, 2022

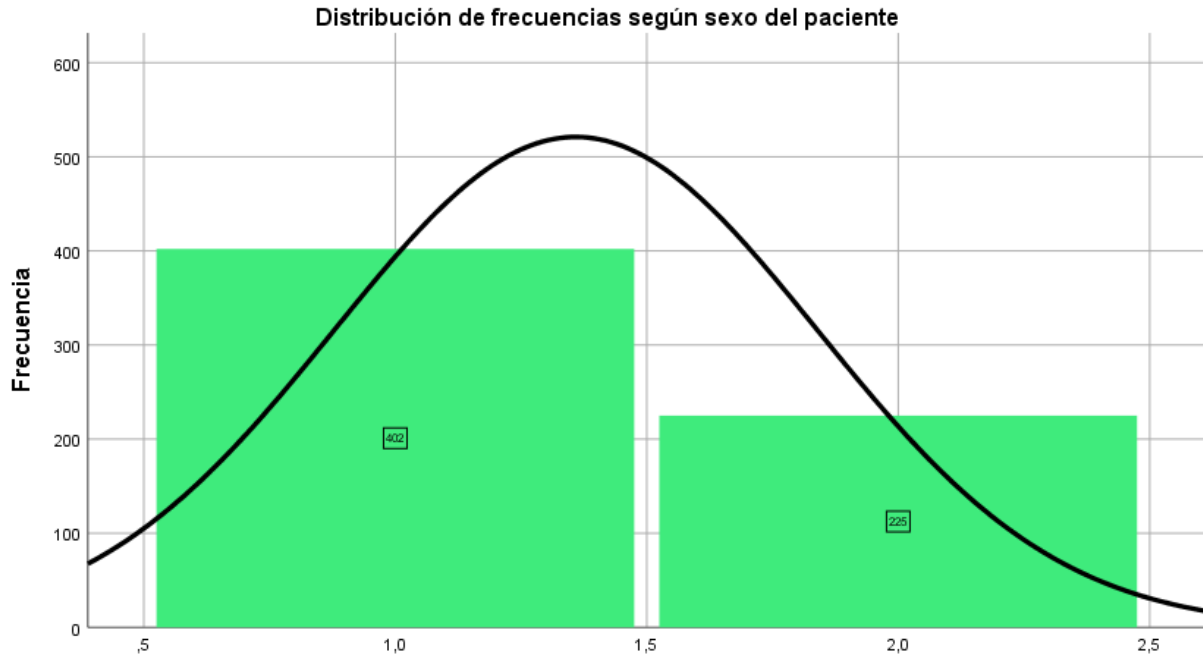


Figura 5-4: curva de distribución de la variable sexo

Realizado por: Monar Manuel, 2022

De acuerdo con la tabla 8 y la gráfica 5 la distribución de los datos de la variable sexo, donde los hombres tiene un 65% y mujeres un 35% en relación con los pacientes que se hacen atender en los Centros de Saludo de la ciudad de Guaranda, en la tabla 7 muestra una simetría positiva de 0,590 lo que implica una inclinación a la derecha con respecto a la media.

Tabla 8-4: Análisis exploratorio de datos de la variable Autoidentificación del paciente

N	Válido	627
	Perdidos	0
Media		1,20
Mediana		1,00
Moda		1
Desv. Desviación		,542
Varianza		,294
Asimetría		,403
Error estándar de asimetría		,098
Curtosis		,542
Error estándar de curtosis		,195

Realizado por: Monar Manuel, 2022

Tabla 9-4: Distribución de frecuencia de la variable autoidentificación

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	MESTIZO/A	432	68,9	68,9	68,9
	INDÍGENA	154	24,6	24,6	93,5
	MONTUBIO/A	5	,8	,8	94,3
	NO SE IDENTIFICA	36	5,7	5,7	100,0
	TOTAL	627	100,0	100,0	

Realizado por: Monar Manuel, 2022

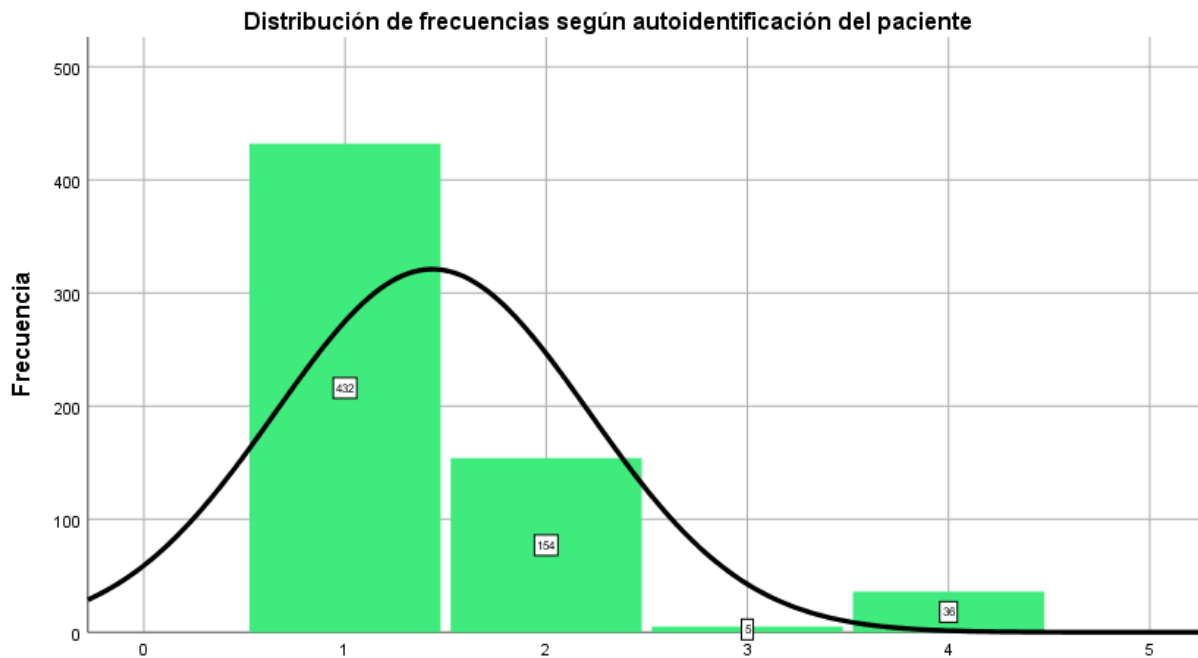


Figura 6-4: Curva de distribución de la variable tipo de diagnóstico

Realizado por: Monar Manuel, 2022

De acuerdo con la tabla 10 y la figura 6 la distribución de los datos de la variable autoidentificación, donde el 68,9% corresponde a los pacientes que se autoidentifican con mestizos, el 24,6% se identifican como indígenas y el 5,75 no se autoidentifica. En la tabla 9 muestra una simetría positiva de 0,403 lo que implica una inclinación de la curva a la derecha con respecto a la media.

Tabla 10-4: Análisis exploratorio de datos de la variable complejidad del diagnóstico

N	Válido	627
	Perdidos	0
Media		1,94
Mediana		2,00
Moda		2
Desv. Desviación		,242
Varianza		,058
Asimetría		-3,634
Error estándar de asimetría		,098
Curtosis		11,242
Error estándar de curtosis		,195

Realizado por: Monar Manuel, 2022

Tabla 11-4: Distribución de frecuencias de la variable complejidad del diagnóstico

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Primera	39	6,2	6,2	6,2
	Subsecuente	588	93,8	93,8	100,0
	Total	627	100,0	100,0	

Realizado por: Monar Manuel, 2022

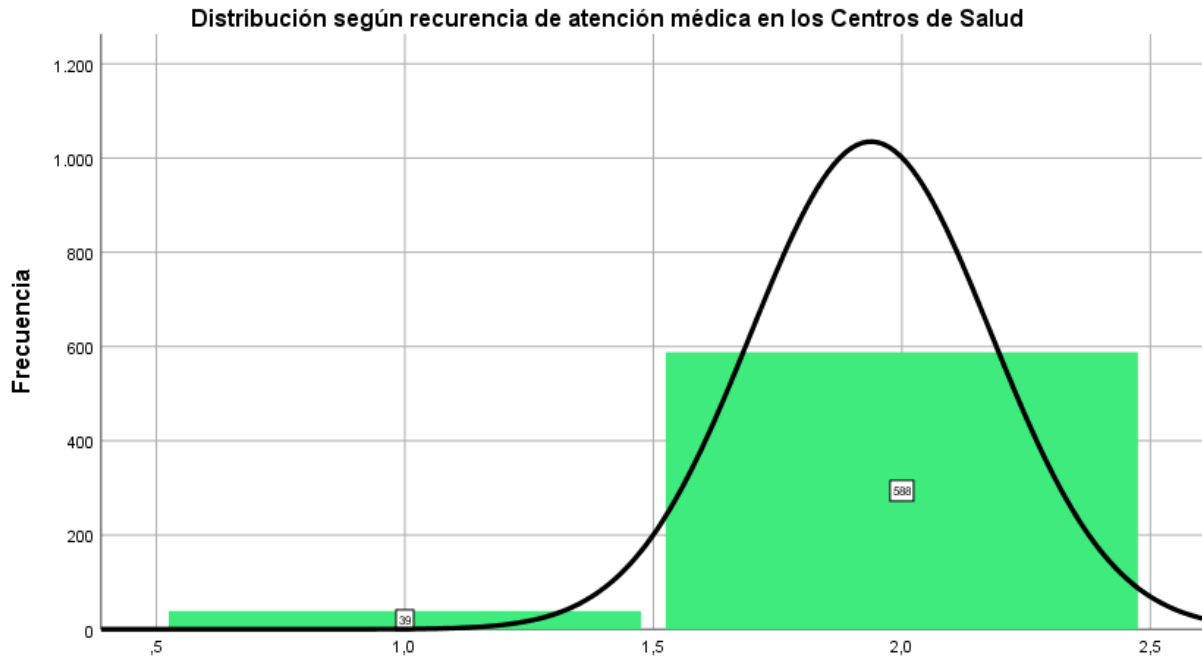


Figura 7-4: Curva de distribución de la variable complejidad del diagnóstico

Realizado por: Monar Manuel, 2022

La tabla 12 y la figura 7 muestra la distribución de los datos de la variable complejidad del diagnóstico en referencia a las visitas que hacen los pacientes a los centros de salud del cantón Guaranda, donde el 93,8% visita subsecuentemente y el 6,2% lo hace por primera vez. La tabla

Tabla 12-4: Análisis exploratorio de la variable clasificación de la IMC

N	Válido	627
	Perdidos	0
Media		6,50
Mediana		7,00
Moda		8
Desv. Desviación		1,684
Varianza		2,835
Asimetría		-,556
Error estándar de asimetría		,098
Curtosis		-1,404
Error estándar de curtosis		,195

Realizado por: Monar Manuel, 2022

Tabla 13-4: Distribución de frecuencias de la variable clasificación del IMC

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Delgadez Moderada	1	,2	,2	,2
	Obesidad Tipo I	149	23,8	23,8	23,9
	Obesidad Tipo II	64	10,2	10,2	34,1
	Obesidad Tipo III	21	3,3	3,3	37,5
	Peso Normal	104	16,6	16,6	54,1
	Sobrepeso	288	45,9	45,9	100,0
	Total	627	100,0	100,0	

Realizado por: Monar Manuel, 2022

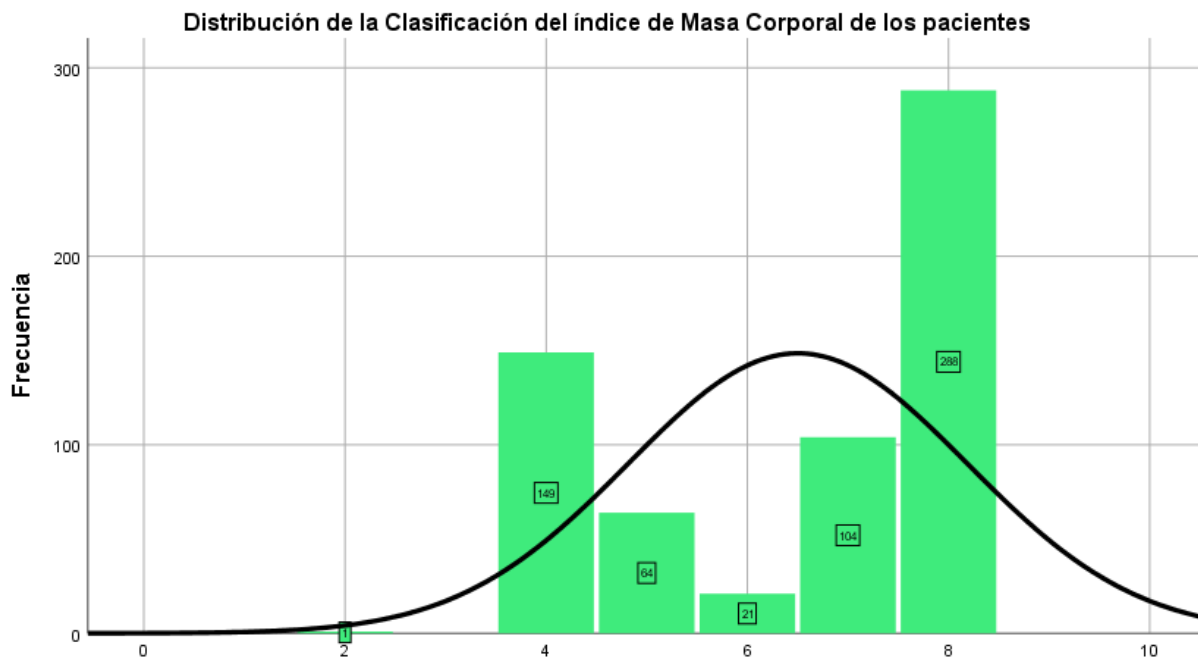


Figura 8-4: Curva de distribución normal de la variable clasificación del IMC

Realizado por: Monar Manuel, 2022

En la tabla 13 y 14 y la figura 8 muestra el análisis exploratorio de datos de la variable clasificación del índice de masa corporal, y donde se observa una desviación media estándar de

1,68 y una media de 6,6 el factor de sobre peso es el porcentaje de mayor frecuencia con un 45,9 % y una distribución normal de datos con una simetría de -0,556.

4.2 Identificación de factores mediante tabla de contingencias que inciden en el crecimiento con personas con diabetes.

Para la identificación de los factores de crecimiento de personas con diabetes se procedió a ser una contratación de tablas de contingencia y mediante la comparación de los valores de Chi cuadrado realizar la asociación de variables

4.2.1. Anova para determinar chi cuadrado de las variables sociodemográficas.

Tabla 14-4: Tabla de contingencia entre las variables el nivel de ingreso vs destino de la población (Rural, Urbano).

			Destino de la Población		
			Rural	Urbano	Total
Nivel de ingresos familiares	Menos de 425	Recuento	89	67	156
		% dentro de Destino de la Población	74,2%	72,0%	73,2%
	De 425 a 650	Recuento	23	21	44
		% dentro de Destino de la Población	19,2%	22,6%	20,7%
	De 700 a 1000	Recuento	8	4	12
		% dentro de Destino de la Población	6,7%	4,3%	5,6%
	De 1100 a 1600	Recuento	0	1	1
		% dentro de Destino de la Población	0,0%	1,1%	0,5%
Total	Recuento	120	93	213	
	% dentro de Destino de la Población	100,0%	100,0%	100,0%	

Realizado por: Monar Manuel, 2022

La tabla 15 de contingencia entre la variable columnas destino de la población y filas destino de la población para identificar se observa que el mayor porcentaje de frecuencia de datos corresponde

a la categoría de al nivel de ingresos menos de 450 dólares con la variable rural y urbano, y esto corresponde a los pacientes que usan los servicios de la red pública de salud.

Tabla 15-4: Pruebas de chi-cuadrado

	Valor	df	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	2,139 ^a	3	,544
Razón de verosimilitud	2,518	3	,472
Asociación lineal por lineal	,052	1	,819
N de casos válidos	213		

a. 2 casillas (25,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es ,44.

Realizado por: Monar Manuel, 2022

La tabla 16 muestra la prueba de chi cuadrado entre las variables destino de la población y nivel de ingresos, mostrando una dependencia entre los dos factores de 2,13 con una razón de verosimilitud de 2,518 esto implica una baja dependencia entre las dos variables.

Tabla 16-4: Tabla de contingencia destino de la población (Rural, Urbano) y formación académica.

			Destino de la Población		
			Rural	Urbano	Total
Nivel de formación Académica	Ninguna	Recuento	12	11	23
		% dentro de Destino de la Población	10,0%	11,8%	10,8%
	Primaria	Recuento	44	41	85
		% dentro de Destino de la Población	36,7%	44,1%	39,9%
	Bachiller	Recuento	54	33	87
		% dentro de Destino de la Población	45,0%	35,5%	40,8%
	Técnico - Superior	Recuento	4	2	6
		% dentro de Destino de la Población	3,3%	2,2%	2,8%
	Tercer Nivel	Recuento	6	6	12
		% dentro de Destino de la Población	5,0%	6,5%	5,6%
Total		Recuento	120	93	213

% dentro de Destino de la Población	100,0%	100,0%	100,0%
-------------------------------------	--------	--------	--------

Realizado por: Monar Manuel, 2022

La tabla 17 se observa el arreglo matricial de la variable nivel de formación académica y destino de la población, donde se observa que el mayor porcentaje entre la columna de destino de la población y nivel de formación académica es bachiller con un 54% y urbano es el 41%

Tabla 17-4: Chi cuadrado de la tabla de contingencia nivel de formación y destino de la población.

	Valor	df	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	2,503 ^a	4	,644
Razón de verosimilitud	2,517	4	,642
Asociación lineal por lineal	,531	1	,466
N de casos válidos	213		

a. 2 casillas (20,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 2,62.

Realizado por: Monar Manuel, 2022

4.2.1. Anava para determinar chi cuadrado de los factores modificables.

Para el análisis de las tablas de contingencia y determinar dependencia o independencia entre dos variables de los factores modificables como son la actividad física, índice de masa corporal, hábito de fumar y consumo de tabacos y alcohol. de pacientes con diabetes mellitus tipo dos de la ciudad de Guaranda.

Tabla 18-4: Tabla de contingencia de las variables actividad física y Destino de la Población

			Destino de la Población		
			Rural	Urbano	Total
Realiza actividad física	No realiza actividad física	Recuento	50	32	82
		% dentro de Realiza actividad física	61,0%	39,0%	100,0%
		Recuento	64	50	114

	De 1 a 3 horas por semana	% dentro de Realiza actividad física	56,1%	43,9%	100,0%
	De 4 a 10 horas por semana	Recuento	6	9	15
		% dentro de Realiza actividad física	40,0%	60,0%	100,0%
	Más de 10 horas por semana	Recuento	0	2	2
		% dentro de Realiza actividad física	0,0%	100,0%	100,0%
Total		Recuento	120	93	213
		% dentro de Realiza actividad física	56,3%	43,7%	100,0%

Realizado por: Monar Manuel, 2022

La tabla de contingencia de la variable realiza actividad física vs la variable destino de la población donde los mayores porcentajes de conteo de casos se da en la población rural y el factor de 1 a 3 horas semanales es el más significativo en cuanto a la mayor frecuencia.

Tabla 19-4: Pruebas de chi-cuadrado de la variable actividad física y destino de la población

	Valor	df	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	4,927 ^a	3	,177
Razón de verosimilitud	5,652	3	,130
Asociación lineal por lineal	3,391	1	,066
N de casos válidos	213		

a. 2 casillas (25,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es ,87.

Realizado por: Monar Manuel, 2022

Las variables realizan actividad física y destino de la población son dependientes con un valor de 0,177 de significancia según la prueba de chi cuadrado de la tabla 20.

Tabla 20-4: Tabla cruzada Tabla cruzada Clasificación del IMC*Destino de la Población

		Destino de la Población		Total	
		Rural	Urbano		
Clasificación del IMC	Delgadez Moderada	Recuento	0	1	1
		% dentro de Destino de la Población	0,0%	0,3%	0,2%
	Obesidad Tipo I	Recuento	77	72	149
		% dentro de Destino de la Población	23,1%	24,6%	23,8%
	Obesidad Tipo II	Recuento	29	35	64
		% dentro de Destino de la Población	8,7%	11,9%	10,2%
	Obesidad Tipo III	Recuento	6	15	21
		% dentro de Destino de la Población	1,8%	5,1%	3,3%
	Peso Normal	Recuento	43	61	104
		% dentro de Destino de la Población	12,9%	20,8%	16,6%
	Sobrepeso	Recuento	179	109	288
		% dentro de Destino de la Población	53,6%	37,2%	45,9%
Total		Recuento	334	293	627
		% dentro de Destino de la Población	100,0%	100,0%	100,0%

Realizado por: Monar Manuel, 2022

Tabla cruzada de la variable clasificación del índice de masa corporal con la variable destino de la población, donde se observa que el mayor porcentaje de recuento de los casos está en el factor sobre peso con un 53,6% con el factor rural, también se observa un valor significativo del factor obesidad tipo I con un 23 % y 24,6% dentro de la variable destino de la población.

Tabla 21-4: Pruebas de chi-cuadrado de las variables

	Valor	df	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	23,135 ^a	5	,000
Razón de verosimilitud	23,736	5	,000
Asociación lineal por lineal	5,817	1	,016
N de casos válidos	627		

a. 2 casillas (16,7%) han esperado un recuento menor que 5. El recuento mínimo esperado es ,47.

No son dependientes las variables clasificación del índice de masa corporal y destino de la población por el valor de significancia 0,000.

Tabla 22-4: Tabla cruzada Consume por lo menos dos tabacos al día*Destino de la Población.

			Destino de la Población		Total
			Rural	Urbano	
Consume por lo menos dos tabacos al día	No consume	Recuento	90	79	169
		% dentro de Destino de la Población	75,0%	84,9%	79,3%
	De 1 a 5 diarios	Recuento	24	12	36
		% dentro de Destino de la Población	20,0%	12,9%	16,9%
	De 6 a 20 diarios	Recuento	5	2	7
		% dentro de Destino de la Población	4,2%	2,2%	3,3%
	Más de 20 diarios	Recuento	1	0	1
		% dentro de Destino de la Población	0,8%	0,0%	0,5%
Total	Recuento	120	93	213	
	% dentro de Destino de la Población	100,0%	100,0%	100,0%	

Realizado por: Monar Manuel, 2022

Tabla cruzada de la variable consumo de tabacos por lo menos dos tabacos al día con la variable destino de la población, donde se observa que el mayor porcentaje de recuento de los casos está en el factor No consume con el 75% y el 84,95 en el factor urbano.

Tabla 23-4: Pruebas de chi-cuadrado variables consumo de tabacos oblación

	Valor	df	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	3,638 ^a	3	,303
Razón de verosimilitud	4,077	3	,253
Asociación lineal por lineal	3,450	1	,063
N de casos válidos	213		

a. 4 casillas (50,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es ,44.

Realizado por: Monar Manuel, 2022

Existe dependencia entre la variable consumo de tabacos y destino de la población.

4.2.1. Anova para determinar chi cuadrado de los factores no modificables.

Tabla 24-4: Tabla de contingencia edad del paciente vs destino de la población (Rural, Urbano).

		Destino de la Población		Total	
		Rural	Urbano		
Edad del paciente	<= 40	Recuento	20	13	33
		% dentro de Destino de la Población	6,0%	4,4%	5,3%
	41 - 55	Recuento	114	52	166
		% dentro de Destino de la Población	34,1%	17,7%	26,5%
	56 - 71	Recuento	150	116	266
		% dentro de Destino de la Población	44,9%	39,6%	42,4%
	72 - 86	Recuento	49	94	143
		% dentro de Destino de la Población	14,7%	32,1%	22,8%
	87+	Recuento	1	18	19
		% dentro de Destino de la Población	0,3%	6,1%	3,0%
Total		Recuento	334	293	627
		% dentro de Destino de la Población	100,0%	100,0%	100,0%

Realizado por: Monar Manuel, 2022

Tabla cruzada de la variable edad del paciente con la variable destino de la población, donde se observa que el mayor porcentaje de recuento de los casos está en el factor de 56 a 71 años con un 44,9% y el 39,6 en el factor rural y urbano.

Tabla 25-4: Chi cuadrado de la variable Edad y destino de la población

	Valor	df	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	55,917 ^a	4	,000
Razón de verosimilitud	59,807	4	,000
Asociación lineal por lineal	44,973	1	,000
N de casos válidos	627		

a. 0 casillas (0,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 8,88.

Realizado por: Monar Manuel, 2022

No existe dependencia entre la variable edad y destino de la población.

Tabla 26-4: Tabla cruzada de la variable antecedentes familiares vs destino de la población

			Destino de la Población		
			Rural	Urbano	Total
Familiar diagnosticado con diabetes	SI	Recuento	92	66	158
		% dentro de Destino de la Población	76,7%	71,0%	74,2%
	NO	Recuento	28	27	55
		% dentro de Destino de la Población	23,3%	29,0%	25,8%
Total		Recuento	120	93	213
		% dentro de Destino de la Población	100,0%	100,0%	100,0%

Realizado por: Monar Manuel, 2022

La tabla 27 muestra la ponderación en relación entre la variable destino de la población y antecedentes familiares sobre diagnóstico de diabetes mellitus tipo 2 y el mayor porcentaje 76,7 en la zona rural con la opción de si tuvo antecedentes familiares.

Tabla 27-4: Pruebas de chi-cuadrado de la variable destino de la población y antecedentes familiares

	Valor	df	Significación		
			asintótica (bilateral)	Significación exacta (bilateral)	Significación exacta (unilateral)
Chi-cuadrado de Pearson	,888 ^a	1	,346		
Corrección de continuidad ^b	,616	1	,433		
Razón de verosimilitud	,884	1	,347		
Prueba exacta de Fisher				,349	,216
Asociación lineal por lineal	,884	1	,347		
N de casos válidos	213				

a. 0 casillas (0,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 24,01.

b. Sólo se ha calculado para una tabla 2x2

Realizado por: Monar Manuel, 2022

Existe dependencia entre las variables antecedentes familiares y destino de la población

Tabla 28-4: Tabla de contingencia entre la variable autoidentificación y destino de la población

			Destino de la Población		
			Rural	Urbano	Total
Autoidentificación	No Se Identifica	Recuento	21	15	36
		% dentro de Destino de la Población	6,3%	5,1%	5,7%
	Mestizo/A	Recuento	228	204	432
		% dentro de Destino de la Población	68,3%	69,6%	68,9%
	Indígena	Recuento	82	72	154
		% dentro de Destino de la Población	24,6%	24,6%	24,6%
	Montubio/A	Recuento	3	2	5
		% dentro de Destino de la Población	0,9%	0,7%	0,8%
	Total	Recuento	334	293	627
		% dentro de Destino de la Población	100,0%	100,0%	100,0%

Realizado por: Monar Manuel, 2022

La tabla 29 muestra la ponderación en relación entre la variable destino de la población y autoidentificación y el mayor porcentaje 68,3% en la zona rural y urbana con la opción de la autoidentificación mestizo.

Tabla 29-4: Pruebas de chi-cuadrado entre la variable autoidentificación y destino de la población

	Valor	df	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	,504 ^a	3	,918
Razón de verosimilitud	,507	3	,917
Asociación lineal por lineal	,031	1	,861
N de casos válidos	627		

a. 2 casillas (25,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 2,34.

Realizado por: Monar Manuel, 2022

Existe dependencia entre las variables autoidentificación y destino de la población.

4.3 Modelos estadísticos mediante Análisis de componentes principales para datos categóricos de factores de crecimiento de los pacientes de diabetes mellitus tipo 2.

Tabla 30-4: Dimensiones del Modelo (CAPTCA)

Dimensión	Alfa de Cronbach	Varianza contabilizada para Total (autovalor)
1	,752	3,269
2	,688	2,743
Total	,903 ^a	6,011

Realizado por: Monar Manuel, 2022

La tabla 31 se observa el resumen del modelo según el número de interacciones de las variables categóricas, donde la dimensión uno con una Alfa de Cronbach de 0.752 y un auto valor de 3,269 es la dimensión que más se ajusta al modelo de componentes principales categóricos.

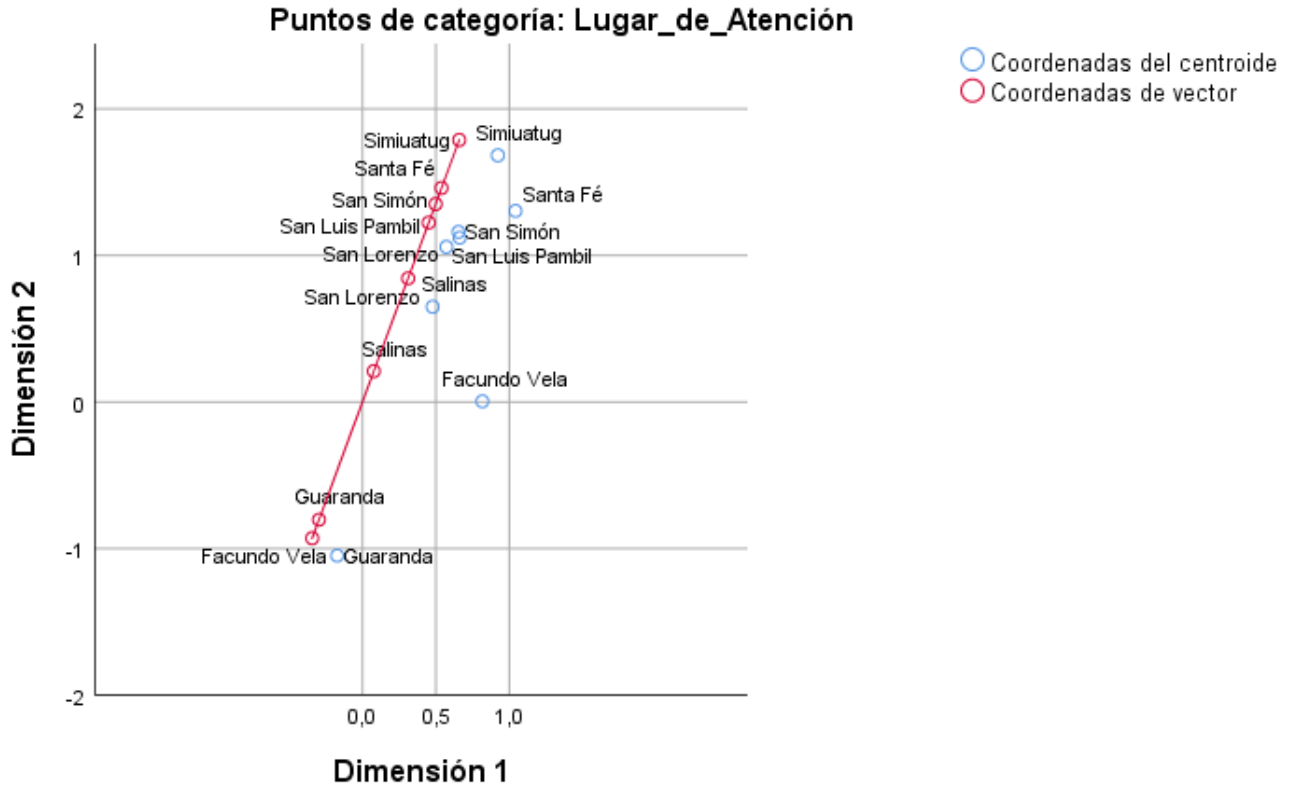


Figura 9-4: Cuantificación de puntos variable lugar de atención de pacientes con diabetes mellitus tipo 2

Realizado por: Monar Manuel, 2022

La figura 9 muestra a los centros de atención de Guaranda y Facundo Vela como los factores que se acercan al origen y los más representativos de la variable lugar de atención del paciente con diabetes mellitus tipo 2, así como el vector lineal de las diferentes parroquias por dimensión de atención.

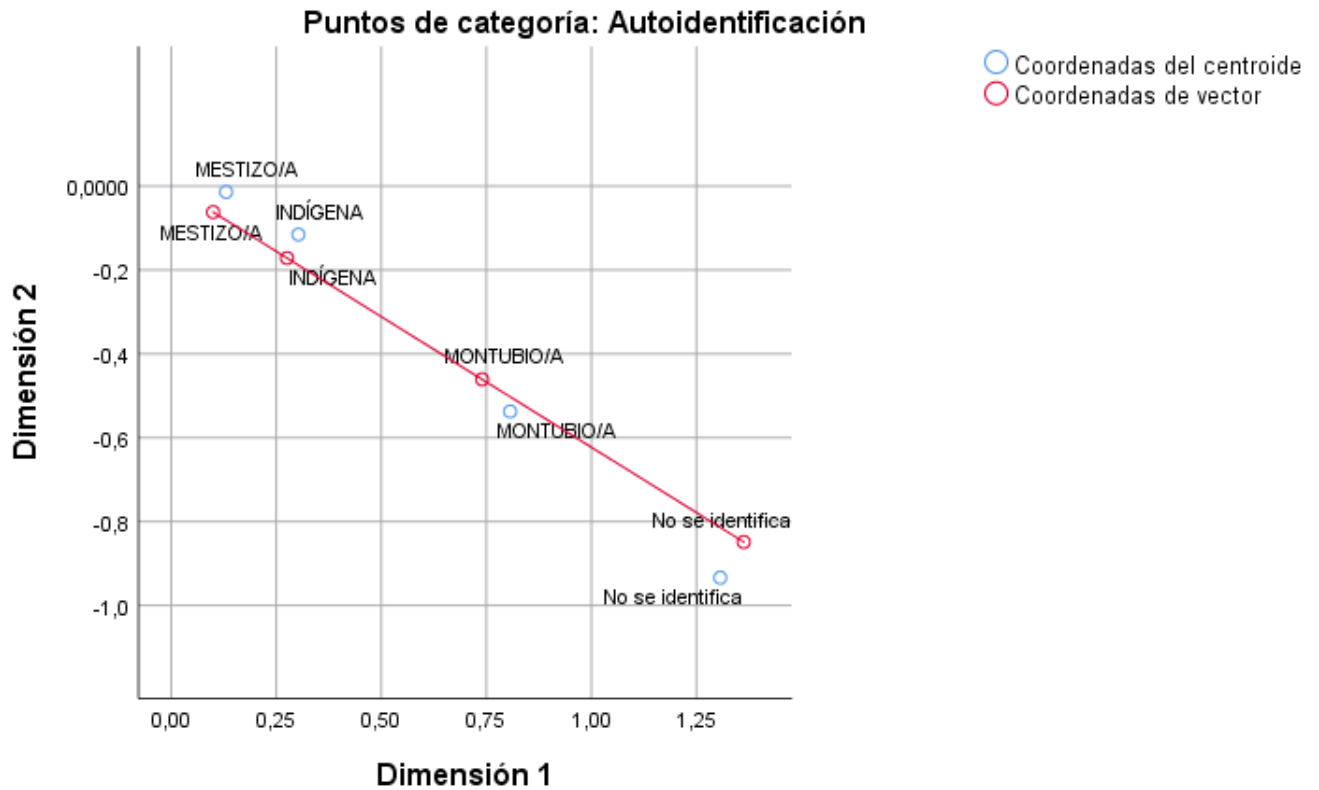


Figura 10-4: Cuantificación de puntos de categoría variable autoidentificación

Realizado por: Monar Manuel, 2022

La figura 10 identifica el factor mestizo cercanos al origen que es más representativo de la variable autoidentificación de los pacientes con diabetes mellitus tipo 2, así como el vector lineal de factores de la variable.

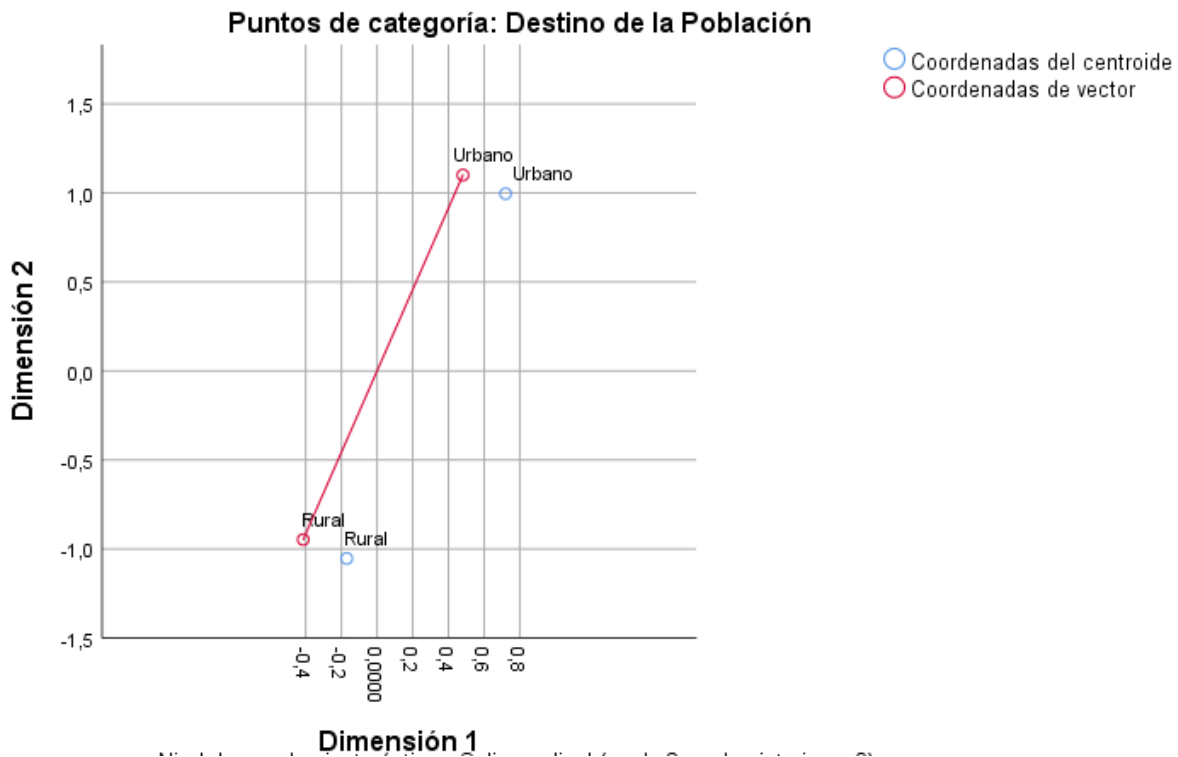


Figura 11-4: Cuantificación de categoría variable Destino de la población

Realizado por: Monar Manuel, 2022

La figura 11 muestra el factor población rural que se acercan al origen como los más representativo de la variable destino de la población, así como el vector lineal de los diferentes factores de la variable destino de la población.

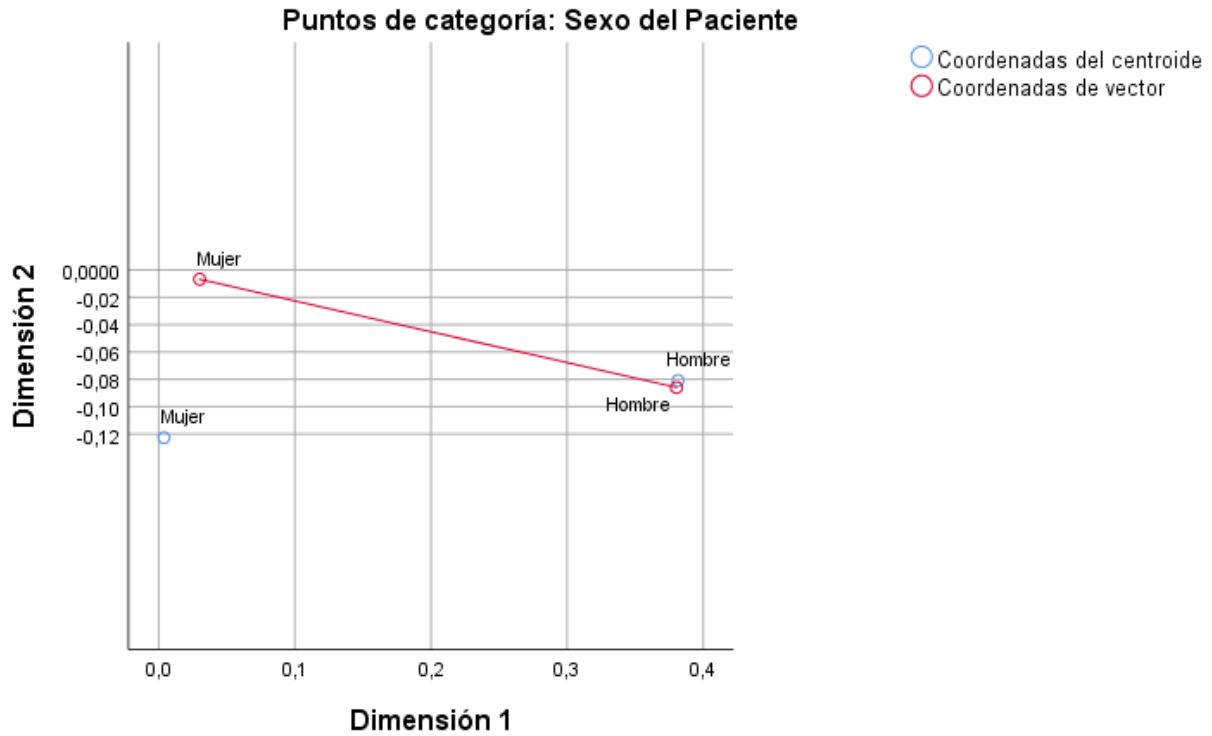


Figura 12-4: Cuantificación de puntos de categoría variable sexo del paciente

Realizado por: Monar Manuel, 2022

La figura 12 identifica al factor mujer que se acercan al origen como el más representativo de la variable sexo del paciente.

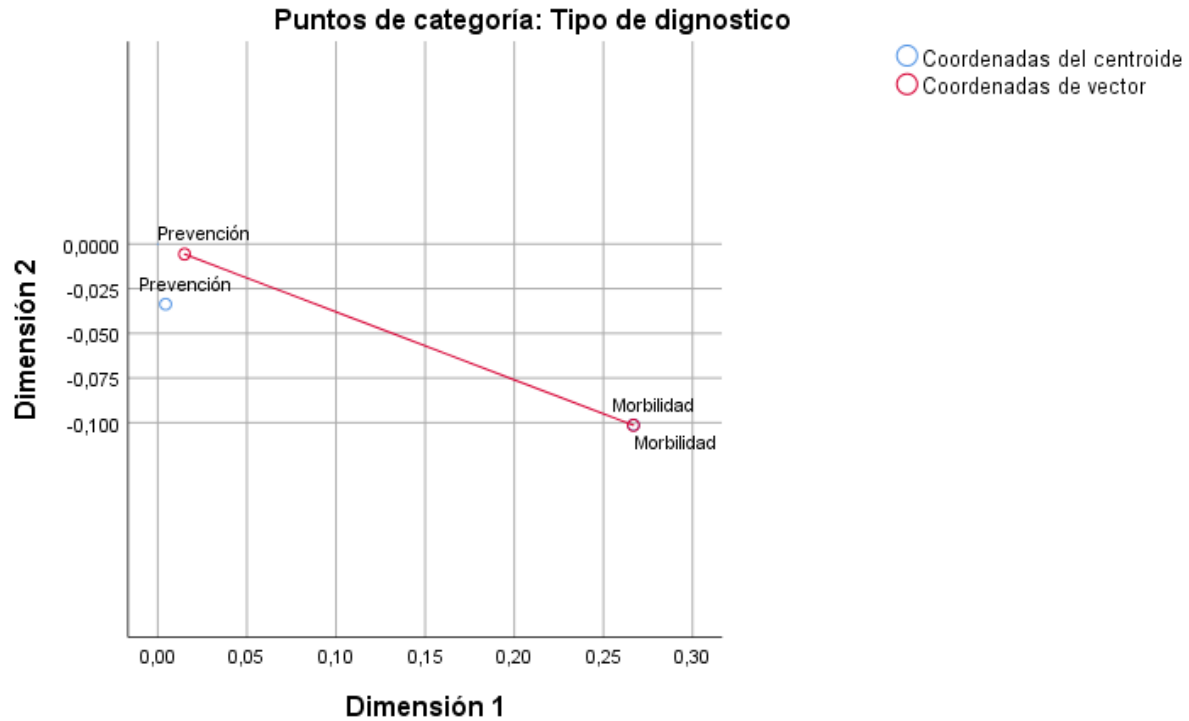


Figura 13-4: Cuantificación de puntos de categoría variable tipo de diagnóstico

Realizado por: Monar Manuel, 2022

La figura 13 se identificó el factor prevención que se refiere a pacientes que acudieron por primera para hacerse atender en los centros de salud que se acercan al origen como el más representativo de la variable tipo de diagnóstico, así como el vector lineal de los diferentes factores de la variable tipo de diagnóstico.

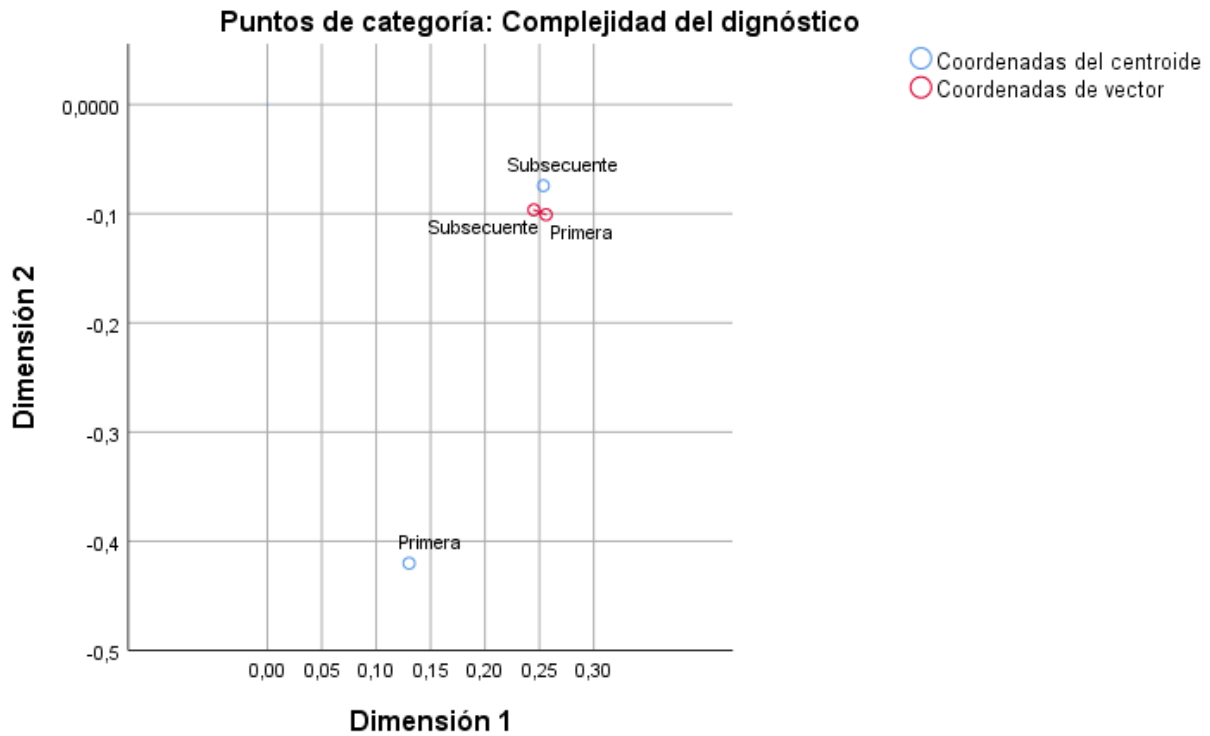


Figura 14-4: Cuantificación de puntos de categorías variable complejidad del diagnóstico

Realizado por: Monar Manuel, 2022

La figura 14 se identificó el factor subsecuente que se refiere a pacientes que acudieron por primera para hacerse atender en los centros de salud que se acercan al origen como el más representativo de la variable tipo de diagnóstico, así como el vector lineal de los diferentes factores de la variable complejidad.

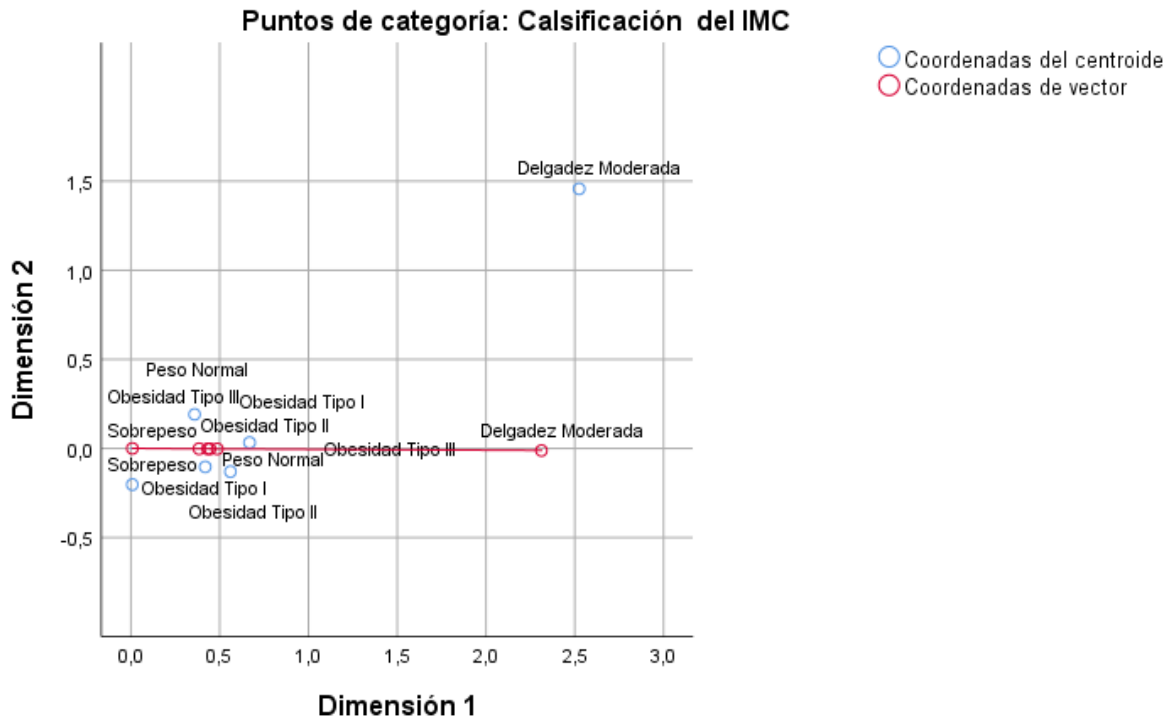


Figura 15-4: Cuantificación de puntos de categoría de la variable Clasificación del IMC

Realizado por: Monar Manuel, 2022

La figura 15 se identificó a los factores sobrepeso obesidad tipo II y II que se acercan al origen como más representativo de la variable clasificación del IMC, así como el vector lineal de los diferentes factores de la variable.

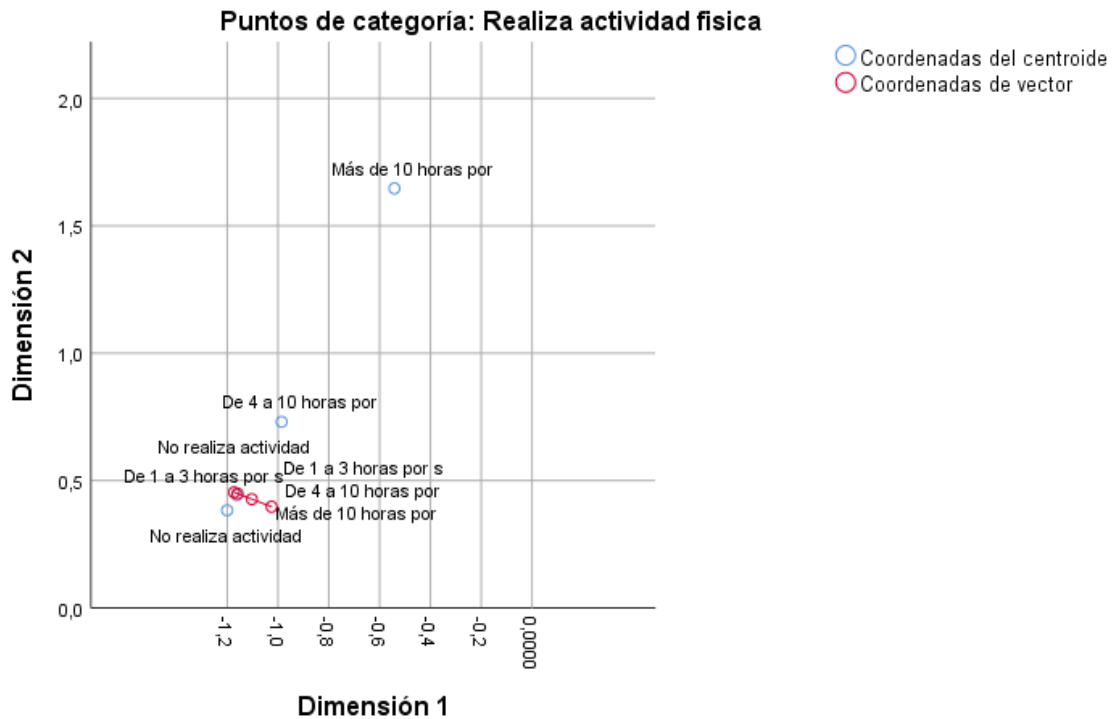


Figura 16-4: Cuantificación categórica de puntos de la variable realiza actividad física

Realizado por: Monar Manuel, 2022

La figura 16 grafica al factor No realiza actividad física como más cercano al origen y por ende al más representativo de la variable realiza actividad física.

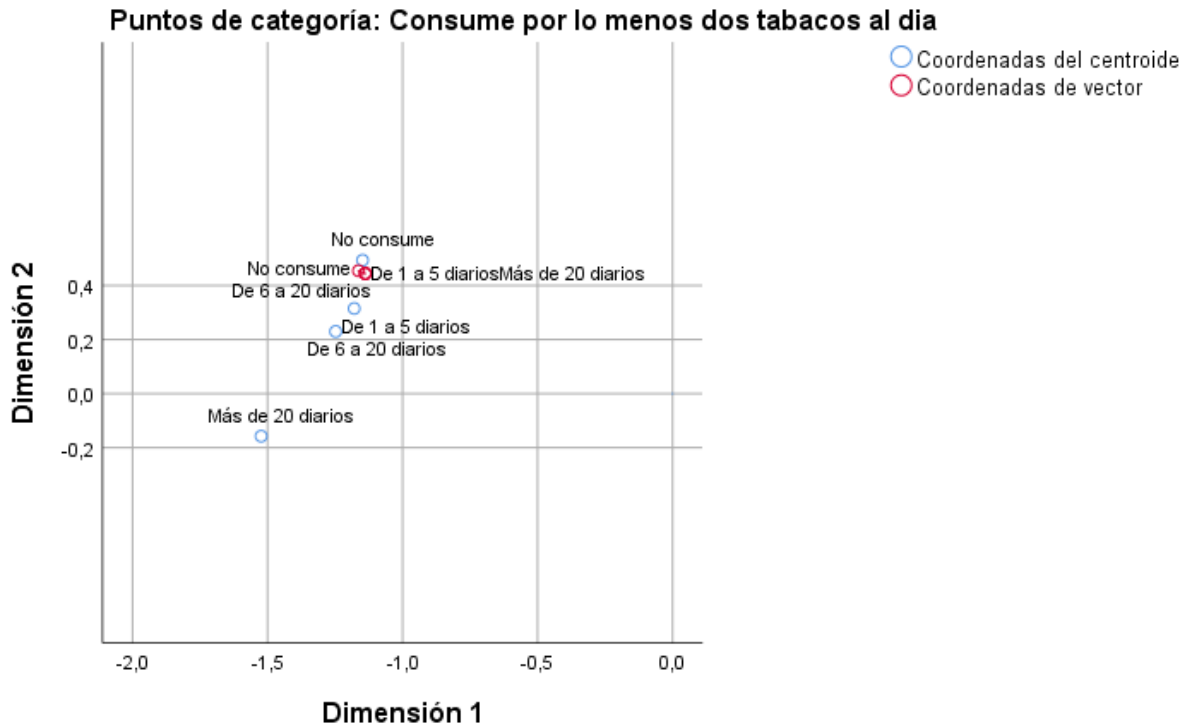


Figura 17-4: Cuantificación de puntos de categoría de la variable consumo de tabacos

Realizado por: Monar Manuel, 2022

La figura 17 identifica al factor más de 20 cigarrillos diarios dentro de la variable consumo de tabacos, como el factor más cerca al origen por lo tanto el más significativo.

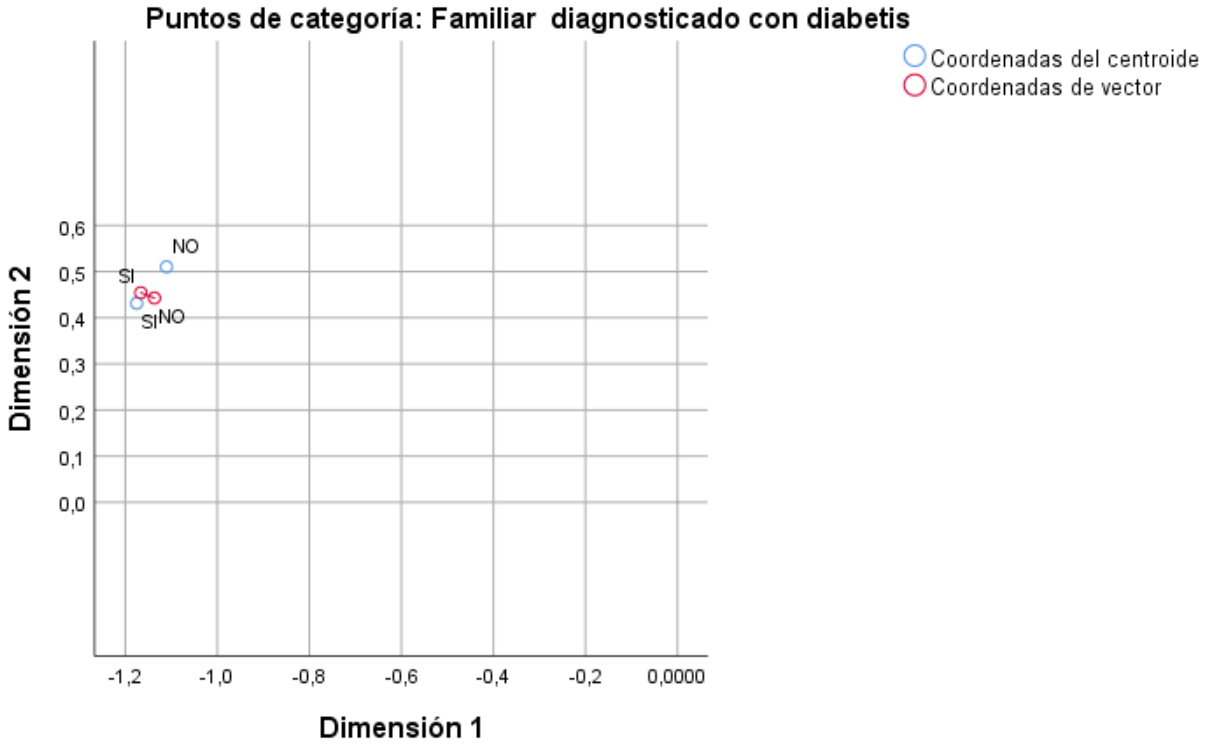
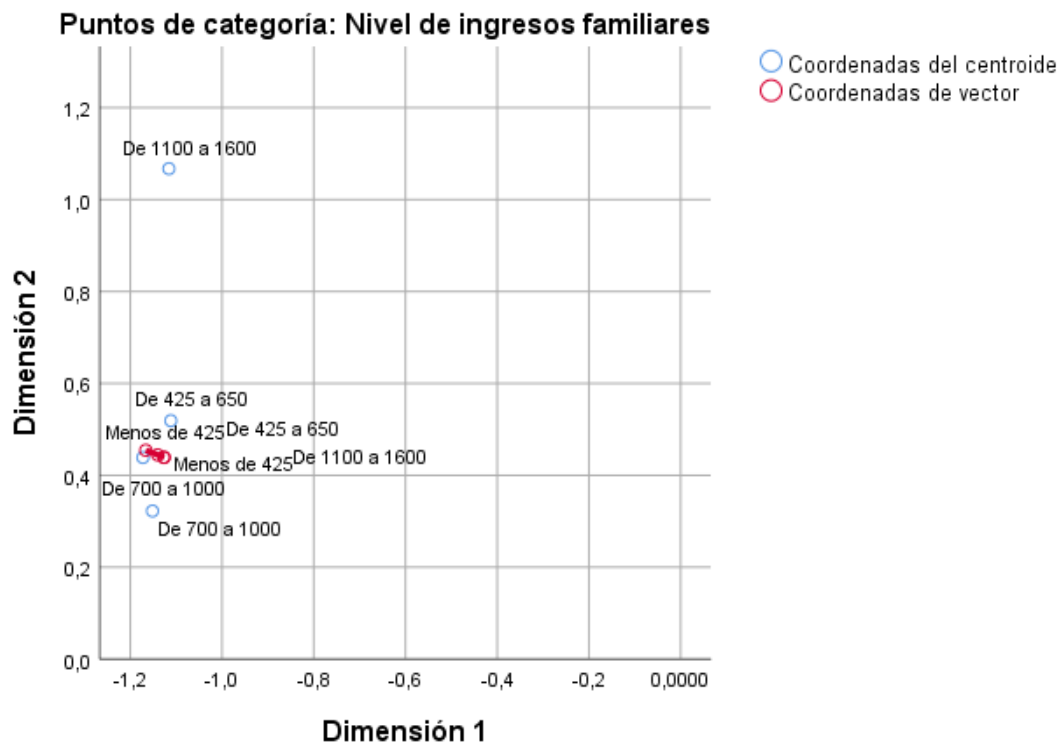


Figura 18-4: Cuantificación de puntos de categoría de la variable Familiar diagnosticado con diabetes

Realizado por: Monar Manuel, 2022



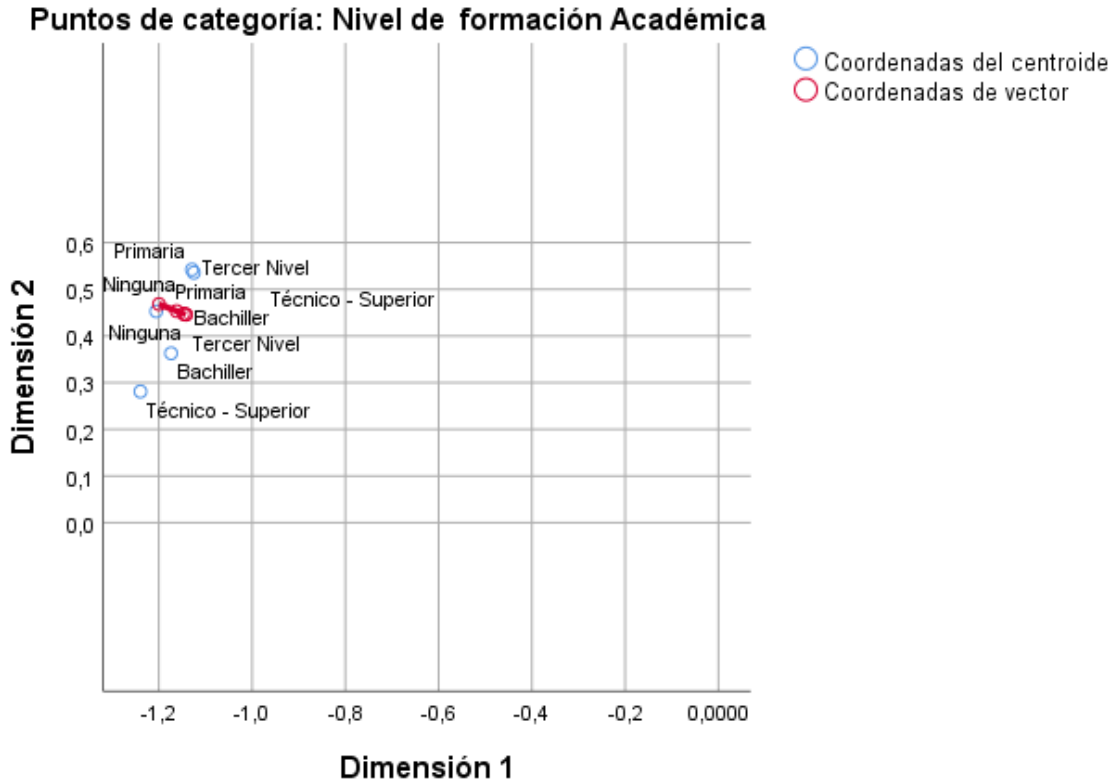
Nivel de escalamiento óptimo: Spline ordinal (grado 2, nudos interiores 2).

Normalización de principal de variable.

Figura 19-4: Cuantificación de puntos de categoría de la variable nivel de ingresos familiares

Realizado por: Monar Manuel, 2022

La figura muestra al factor de 700 a 1000 dólares como la categoría más cercana al origen y por lo tanto el más representativo de la variable nivel de ingresos familiares



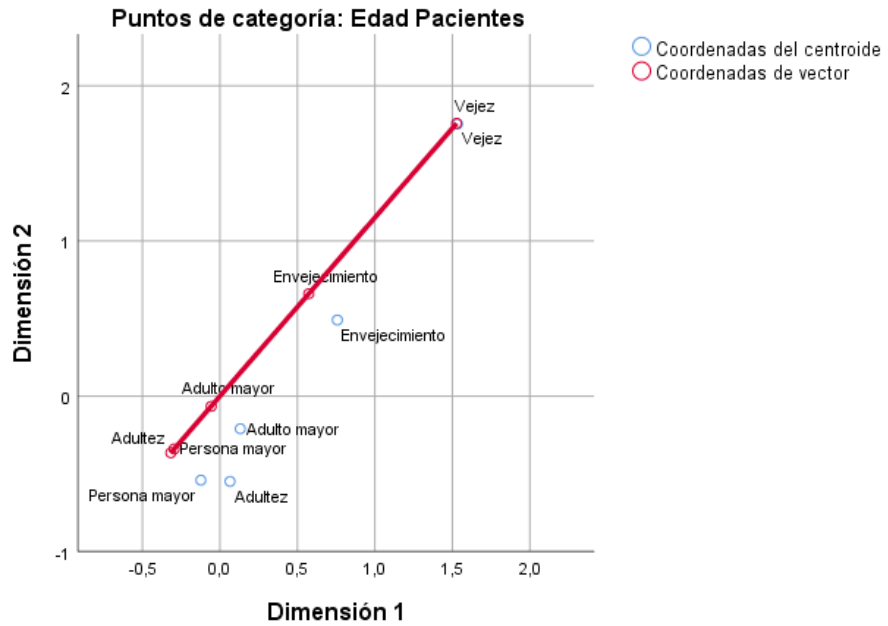
Nivel de escalamiento óptimo: Spline ordinal (grado 2, nudos interiores 2).

Normalización de principal de variable.

Figura 20-4: Cuantificación de puntos de categoría de la variable formación académica

Realizado por: Monar Manuel, 2022

La figura 20 muestra el factor bachiller como la categoría más significativa dentro de la variable formación academia



Nivel de escalamiento óptimo: Spline ordinal (grado 2, nudos interiores 2).
 Normalización de principal de variable.

Figura 21-4: Puntos de categoría de la viable edad paciente

Realizado por: Monar Manuel, 2022

La figura 21 grafica a los factores personas mayores y adultos como loas categorías más significativas dentro de la variable edad de pacientes.

Tabla 31-4: Variables transformadas de correlaciones

	Edad	Autoidentificación	Destino de la Población	Sexo	Tipo de diagnóstico	Complejidad del diagnóstico	Clasificación del IMC	Realiza actividad física	Consumo de tabacos al día	Familiar diagnosticado con diabetes	Ingresos familiares	Formación Académica	Lugar de Atención
Edad	1,000	-,038	,295	-,105	-,038	,083	,002	,117	-,040	-,062	-,047	,049	,227
Autoidentificación	-,038	1,000	-,027	-,019	,004	,017	-,039	,001	-,034	-,059	-,053	,098	-,034
Destino de la Población	,295	-,027	1,000	-,081	-,004	,069	-,156	,087	-,082	,015	-,011	-,010	,834
Sexo del Paciente	-,105	-,019	-,081	1,000	-,012	-,014	,167	-,088	,055	-,067	,033	-,036	-,013
Tipo de diagnóstico	-,038	,004	-,004	-,012	1,000	,051	,020	-,050	-,035	,013	-,046	-,039	-,003
Complejidad del diagnóstico	,083	,017	,069	-,014	,051	1,000	,064	,014	,045	-,037	,019	-,016	,047
Clasificación del IMC	,002	-,039	-,156	,167	,020	,064	1,000	,018	,046	,006	,015	-,055	-,112
Realiza actividad física	,117	,001	,087	-,088	-,050	,014	,018	1,000	,002	-,018	,153	,138	,085
Consumo por lo menos dos tabacos al día ^a	-,040	-,034	-,082	,055	-,035	,045	,046	,002	1,000	,069	,555	-,072	-,048
Familiar diagnosticado con diabetes	-,062	-,059	,015	-,067	,013	-,037	,006	-,018	,069	1,000	,128	-,257	,027
Nivel de ingresos familiares	-,047	-,053	-,011	,033	-,046	,019	,015	,153	,555	,128	1,000	,017	,020
Nivel de formación Académica ^a	,049	,098	-,010	-,036	-,039	-,016	-,055	,138	-,072	-,257	,017	1,000	,000
Lugar de Atención	,227	-,034	,834	-,013	-,003	,047	-,112	,085	-,048	,027	,020	,000	1,000
Dimensión	1	2	3	4	5	6	7	8	9	10	11	12	13
Autovalor	2,072	1,638	1,340	1,172	1,065	1,043	,944	,927	,819	,724	,689	,409	,159

a. Los valores perdidos se han imputado con la modalidad de la variable cuantificada.

Realizado por: Monar Manuel, 2022

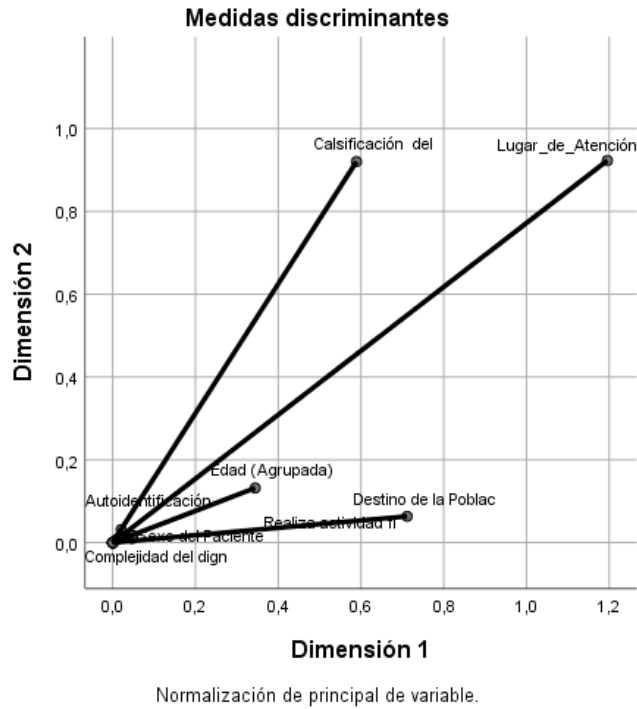


Figura 22-4: Grafica de componentes principales

Mediante la gráfica de componentes principales se puede determinar la formación de tres grupos:

Grupo 1 conformado por las variables clasificación del índice de masa corporal, lugar de atención de pacientes.

Grupo 2 aquí podemos ubicar las variables destino de la población, edad del paciente, realiza actividad física.

Grupo 3 complejidad del diagnóstico autoidentificación del paciente, sexo del paciente.

Para la conformación de estos grupos podemos identificar los autovalores que permite la formación de los grupos

$Y_1 = 0,002$ Edad $+ -0,039$ Autoidentificación $+ -0,156$ Destino de la Población $+0,167$ Sexo $+0,02$ Tipo de diagnóstico $+0,064$ Complejidad del diagnóstico $+1$ Clasificación del IMC $+0,018$ Realiza actividad física $+0,046$ Consumo de tabacos al día $+0,006$ Familiar diagnosticado con diabetes $+0,015$ Ingresos familiares $+ -0,055$ Formación Académica $+ -0,112$ Lugar de Atención.

$Y_2 = 0,227$ Edad $+ -0,034$ Autoidentificación $+0,834$ Destino de la Población $+ -0,013$ Sexo $+ -0,003$ Tipo de diagnóstico $+0,047$ Complejidad del diagnóstico $+ -0,112$ Clasificación del IMC $+0,085$ Realiza actividad física $+ -0,048$ Consumo de tabacos al día $+0,027$ Familiar diagnosticado con diabetes $+0,02$ Ingresos familiares $+0$ Formación Académica $+1$ Lugar de Atención.

$Y_3 = 0,083$ Edad $+0,017$ Autoidentificación $+0,069$ Destino de la Población $+ -0,014$ Sexo $+0,051$ Tipo de diagnóstico $+1$ Complejidad del diagnóstico $+0,064$ Clasificación del IMC $+0,014$ Realiza actividad física $+0,045$ Consumo de tabacos al día $+ -0,037$ Familiar diagnosticado con diabetes $+0,019$ Ingresos familiares $+ -0,016$ Formación Académica $+0,047$ Lugar de Atención

CONCLUSIONES

Mediante la revisión de la teoría se describió los factores de crecimiento de personas con diabetes millitos tipo II, donde se realizó la clasificación de los mismos como factores modificables y factores no modificables, además mediante la realización de un mapa geográfico se localizó los casos de personas con diabetes dentro de las diferentes parroquias urbanas y rurales que conforman el cantón Guaranda.

De la base de datos se obtuvo que PRAS 2020 y mediante el análisis exploratorio de datos se identificó las variables y factores con mayor frecuencia, dando como resultado 53% de los pacientes son de la zona rural del cantón, el 64% corresponde al sexo masculino, el 68% de la población se autoidentifica como mestizo, el 93% de los pacientes son recurrentes esto quiere decir que su visita a los centros de salud son frecuentes, de estos pacientes el 45% están dentro de la clasificación del índice de masa corporal con sobre peso.

Mediante la aplicación de las tablas cruzadas se determinó la dependencia entre las variables que destino de la población y nivel de ingreso, nivel de formación académica, realiza actividad física, consumo de tabos por día, antecedentes familiares, autoidentificación, al dar como resultado un valor de significancia mayor al 0,05

Mientras que las variables que no son dependencia entre las variables que destino de la población y la clasificación del índice de masa corporal, edad del paciente, al dar como resultado valores de significancia menores al 0,05.

El método de componentes principales categóricos es apropiado para determinar perfiles de pacientes con diabetes millitos tipo dos con las variables aplicadas, donde se identificó tres grupos bien diferenciados de las demás variables como son las variables índices de masa corporal, el lugar de atención, la edad del paciente realiza actividad física, complejidad del diagnóstico, autoidentificación del paciente y el sexo como factores de incremento en los auto vectores.

RECOMENDACIONES

Se recomienda a las instancias necesarias como son el ministerio de salud pública permita una mayor accesibilidad a la información y a una base de datos más concisa que permita realizar más análisis y modelos estadísticos, que permita generar patrones y de esta manera poder frenar la enfermedad.

Seguir utilizando más simulaciones en base al análisis de componentes principales categóricos, porque permiten tomar decisiones adecuadas, razonadas, y en el tiempo que se requieran.

Al tener una base de datos en su gran mayoría con variables cualitativas se hace difícil la relación entre ellas, y no poder aplicar herramientas y modelos más estadísticos y así poder identificar las que tiene mayor relación que otras, como por ejemplo hacer regresiones lineales, múltiples y así descartar variables que no aportan al modelo.

GLOSARIO

Diabetes mellitus.- Es considerado un problema de salud pública, catalogada como una causa primaria de muertes para quienes los padecen según, (Carabetti & Aníbal, 2017). La prevalencia de la diabetes en el mundo

Estilo de vida. - Es el conjunto de actitudes y comportamientos que adoptan y desarrollan las personas de forma individual o colectiva para satisfacer sus necesidades

Salud pública. - El cuidado y la promoción de la salud aplicados a toda la población o grupo preciso de la población.

Prevalencia. - la tasa de prevalencia es el número de personas que padecen una enfermedad determinada en un punto determinado de tiempo por cada 1.000 habitantes.

Enfermedad metabólica. - Es el trastorno metabólico ocurre cuando hay reacciones químicas anormales en el cuerpo que interrumpe este proceso. Cuando esto pasa, es posible que tenga demasiadas o muy pocas sustancias que su cuerpo necesita para mantenerse saludable.

Riesgo cardiovascular. - La probabilidad que tiene un individuo de sufrir una de estas enfermedades dentro de un determinado plazo de tiempo y esto va a depender fundamentalmente del número de factores de riesgo que estén presente en un individuo.

Prevalencia. - la prevalencia mide la proporción de personas que se encuentren enfermas al momento de evaluar el paciente en la población, por lo tanto, no hay tiempo de seguimiento.

Desórdenes metabólicos. - Un trastorno metabólico ocurre cuando hay reacciones químicas anormales en el cuerpo que interrumpen este proceso. Cuando esto pasa, es posible que tengan demasiadas o muy pocas sustancias que el cuerpo necesita para mantenerse saludable.

Enfermedades crónicas no transmisibles. - son las principales causas de muerte y discapacidad en el mundo, se refiere a un grupo de enfermedades no transmisibles que no son causadas principalmente por una infección aguda, dan como resultado consecuencias para la salud a largo plazo.

Hábitos alimenticios. - Son comportamientos conscientes, colectivos y repetitivos, que conducen a las personas a seleccionar, consumir y utilizar determinados alimentos o dietas, en respuesta a una influencia social o cultural.

Modelo estocástico. - Es aquel cuyo comportamiento es no determinista, en la medida que el subsiguiente estado del sistema está determinado tanto por las acciones predecibles del proceso como por elementos aleatorios.

Modelación matemática. - Es la actividad que consiste en representar, manipular y comunicar objetos del mundo real con fórmulas y contenidos matemáticos y que, en alguna forma, permite la simulación de los procesos complejos, generen hipótesis y sugieran experimentos o métodos de validación.

Grupos homogéneos. - Es un adjetivo que indica que es igual para los diversos elementos que componen un determinado grupo o conjunto.

Variables Endógenas. - son las que vienen explicadas por el modelo macroeconómico.

Variables exógenas. - son las que vienen explicadas por factores externos al modelo macroeconómico.

Técnicas multivariantes. - son herramientas que permiten extraer información de datos disponibles, las mismas que son complejas y requieren utilización de un conocimiento profundo de sus fundamentos.

Linealidad. - la linealidad es la representación de los datos mediante una línea recta que puede ser creciente o decreciente, cuando algo no puede representarse mediante línea recta, entonces es una relación no lineal.

Homocedasticidad. - Es la condición en la que la varianza residual del término, o error, en un modelo de regresión es constante, es decir el término error no cambia mucho a medida que cambia el valor de la variable del predecesor.

BIBLIOGRAFÍA

- Arnau Gras, J. (1996). Métodos y técnicas avanzadas de análisis de datos en ciencias del comportamiento (Edicions Universitat Barcelona, Ed.; 1st ed., Vol. 1).
- Bohórquez, E., Borreto, M., Muvdi, Y., Rodríguez, A., Badillo, M., Martínez, W., & Mendoza, X. (2020). Modifiable factors and risk of type 2 diabetes mellitus in young adults : a cross-sectional study fatores modificáveis e risco de diabetes mellitus tipo 2 em adultos jovens : um estudo de corte transversal. *ciencia y enfermería*, 26(14). <https://doi.org/10.29393/CE26-7FMCB70007>
- Calderón, A. (2007). *D I A B E T O L O G Í A Epidemiología , genética y mecanismos patogénicos de la diabetes mellitus*. 3–11. file:///C:/Users/mmona/Downloads/S1131358707752688.pdf
- Carabetti, M., & Aníbal, J. (2017). *Cardiomiopatía diabética*.
- Columbié, Y. L., Miguel-soca, P. E., Rivas, D., & Borrego, Y. (2016). Risk factors associated to occurrence of type 2 diabetes mellitus in adults. *Revista Cubana de Endocrinología*, 27(2), 123–133.
- Ezkurra, P. (2016). *Guía de actualización en diabetes mellitus tipo 2*.
- Falconi, S., Añazco, D., Santos, J., Pereira, M., Floreano, L., Almache, V., & Maldonado, K. (2017). Factor de riesgo modificable y no modificable de diabetes mellitus II en una población urbana. *Ciencia y Tecnología*, 1(1), 918–928. <http://investigacion.utmachala.edu.ec/proceedings/index.php/utmach/article/view/155>
- Federation, I. D. (2015). *Atlas Diabetes de la FID*. www.eatlas.idf.org
- Frenk, P., & Márquez, E. (2006). Diabetes mellitus tipo 2 en niños y adolescentes. *Medwave*, 6(8), 36–47. <https://doi.org/10.5867/medwave.2006.08.3421>
- Gómez, W., Almirall, A., & Gutiérrez, M. (2019). Risk factors for type 2 diabetes mellitus in adolescents. *Medisur*, 17(3), 8–10.
- Guevara, E., & Carrasco, M. (2001). Modelación estocástica de DBO y OD. Caso estudio Río

Cabrales. *Revista INGENIERÍA UC*. <https://www.redalyc.org/articulo.oa?id=70780203>

Leiva, A. M., Martínez, M. A., Petermann, F., Garrido, A., Poblete, F., Díaz, X., & Celis, C. (2018). Factores asociados al desarrollo de diabetes mellitus tipo 2 en Chile. *Nutrición Hospitalaria -Epidemiología y Dietética*, 35, 400–407.

Montesinos López, A. O., & Hernández Suárez, C. M. (2007). Modelos matemáticos para enfermedades infecciosas. *Salud Pública de Mexico*, 49(3), 218–226. <https://doi.org/10.1590/s0036-36342007000300007>

Neciosup Vera, C. S. (2018). Modelo lineal mixto conjunto de clases latentes aplicado a un conjunto de datos longitudinales del sector salud. In *Pontificia Universidad Católica del Perú*. <http://hdl.handle.net/20.500.12404/12998>

Palacios, A., Durán, M., & Obregón, O. (2012). factores de riesgo para el desarrollo de diabetes tipo 2 y síndrome metabólico. *Revista Venezolana de Endocrinología y Metabolismo*, 10(1).

Quisiguiña, D. (2011). Factores que inciden en el incremento de las complicaciones crónicas de la diabetes mellitus tipo 2 en el servicio de medicina interna del hospital Alfredo Noboa Montenegro cantón Guaranda provincia Bolívar, periodo febrero 2009 - febrero 2010. In *Biomédica* (Vol. 31, Issue sup3.2). <https://doi.org/10.7705/biomedica.v31i0.530>

Rivera, S., & Valencia, T. (2019). Simulación Estocástica para determinar el valor presente neto y el costo de incertidumbre en una Planta Eólica. *Ciencia En Desarrollo*, 10(2). <https://doi.org/10.19053/01217488.v10.n2.2019.7678>

Rizo, S. M., & Sandoval, R. K. (2016). *Tesis Diabetes Mellitus Abril 2016*.

Sánchez R, M. Á. (2005). Uso metodológico de las tablas de contingencia en la Ciencia Política. *Espacios Públicos*, 8(16), 60–84.

Támara, L. G. (2019). Análisis exploratoria de datos. *Scampus.Com, Estadística*. <https://doi.org/10.2307/j.ctvc5pc9g>

Triola, M. (2013). *Estadística* (p. 888).

Villalba Mata, D. (2011). *Construcción y utilización de un modelo estocástico para la simulación de estrategias de manejo invernal en rebaños de vacas nodrizas.*

ANEXOS

Anexo A: captura de la base de datos de análisis.

Numer_id	ENT_NOM	ENT_DES_PARR	ENT_DES	PCTE_AUT	PROF_SEX	PCTE_ANI	PCTE_ULT	PCTE_ULT	PCTE_ULT	ATEMED_1	ATEMED_2	ATEMED_3	PCTE_ULT
83	CENTRO DE SALUD CORDERO CRESPO	GUARANDA	Urbano	MESTIZO/A	Mujer	39	58	149	26,12	DIABETES ME	Morbilidad	Subsecuente	Sobrepeso
84	CENTRO DE SALUD CORDERO CRESPO	GUARANDA	Urbano	MESTIZO/A	Mujer	59	50,5	138	26,52	DIABETES ME	Morbilidad	Subsecuente	Sobrepeso
114	SAN LUIS DE PAMBIL	SAN LUIS DE PAMBIL	Rural	MESTIZO/A	Mujer	70	71	149	31,98	DIABETES ME	Morbilidad	Subsecuente	Obesidad Tif
131	SANTA FE	SANTA FE (SANTA FE)	Rural	MESTIZO/A	Mujer	48	100	165	36,73	DIABETES ME	Morbilidad	Subsecuente	Obesidad Tif
179	CENTRO DE SALUD CORDERO CRESPO	GUARANDA	Urbano	MESTIZO/A	Mujer	59	61,3	146	28,76	DIABETES ME	Morbilidad	Subsecuente	Sobrepeso
240	SAN LUIS DE PAMBIL	SAN LUIS DE PAMBIL	Rural	MESTIZO/A	Mujer	62	58,6	138	30,77	DIABETES ME	Morbilidad	Subsecuente	Obesidad Tif
287	CENTRO DE SALUD CORDERO CRESPO	GUARANDA	Urbano	MESTIZO/A	Mujer	53	101	153	43,15	DIABETES ME	Morbilidad	Subsecuente	Obesidad Tif
300	SAN LUIS DE PAMBIL	SAN LUIS DE PAMBIL	Rural	MESTIZO/A	Mujer	62	84	160	32,81	DIABETES ME	Morbilidad	Subsecuente	Obesidad Tif
327	CENTRO DE SALUD CORDERO CRESPO	GUARANDA	Urbano	INDÍGENA	Mujer	60	67,7	152	29,3	DIABETES ME	Morbilidad	Subsecuente	Sobrepeso
420	FACUNDO VELA	FACUNDO VELA	Rural	INDÍGENA	Hombre	43	43	143	21,03	DIABETES ME	Morbilidad	Subsecuente	Peso normal
489	SAN SIMON	SAN SIMON (YACOTO)	Rural	MESTIZO/A	Mujer	72	70	140	35,71	DIABETES ME	Morbilidad	Subsecuente	Obesidad Tif
502	CENTRO DE SALUD CORDERO CRESPO	GUARANDA	Urbano	MESTIZO/A	Mujer	58	85	176	27,44	DIABETES ME	Morbilidad	Subsecuente	Sobrepeso
515	GUANUJO	GUARANDA	Urbano	MESTIZO/A	Mujer	46	74	163	27,85	DIABETES ME	Morbilidad	Subsecuente	Sobrepeso
517	SIMIATUG	SIMIATUG	Rural	MESTIZO/A	Hombre	45	51,8	146	24,3	DIABETES ME	Morbilidad	Subsecuente	Peso normal
590	CENTRO DE SALUD CORDERO CRESPO	GUARANDA	Urbano	INDÍGENA	Mujer	44	61	145,7	28,73	DIABETES ME	Morbilidad	Subsecuente	Sobrepeso
724	SIMIATUG	SIMIATUG	Rural	INDÍGENA	Hombre	59	60	158	24,03	DIABETES ME	Morbilidad	Subsecuente	Peso normal
805	CENTRO DE SALUD CORDERO CRESPO	GUARANDA	Urbano	MESTIZO/A	Mujer	59	61	146	28,62	DIABETES ME	Morbilidad	Subsecuente	Sobrepeso
838	GUANUJO	GUARANDA	Urbano	MESTIZO/A	Mujer	54	50	149	22,52	DIABETES ME	Morbilidad	Subsecuente	Peso normal
895	SAN LUIS DE PAMBIL	SAN LUIS DE PAMBIL	Rural	INDÍGENA	Mujer	74	61	144	29,42	DIABETES ME	Morbilidad	Subsecuente	Sobrepeso
944	CENTRO DE SALUD CORDERO CRESPO	GUARANDA	Urbano	MESTIZO/A	Mujer	37	84,2	160	32,89	DIABETES ME	Morbilidad	Subsecuente	Obesidad Tif
996	SAN LUIS DE PAMBIL	SAN LUIS DE PAMBIL	Rural	MESTIZO/A	Mujer	82	58,2	150	25,87	DIABETES ME	Morbilidad	Subsecuente	Sobrepeso
1062	CENTRO DE SALUD CORDERO CRESPO	GUARANDA	Urbano	MESTIZO/A	Hombre	64	68,6	147,8	31,4	DIABETES ME	Morbilidad	Subsecuente	Obesidad Tif
1112	GUANUJO	GUARANDA	Urbano	MESTIZO/A	Mujer	66	85,9	170	29,72	DIABETES ME	Morbilidad	Subsecuente	Sobrepeso
1373	FACUNDO VELA	FACUNDO VELA	Rural	MESTIZO/A	Hombre	59	74	159	29,27	DIABETES ME	Morbilidad	Subsecuente	Sobrepeso
1388	CENTRO DE SALUD CORDERO CRESPO	GUARANDA	Urbano	MESTIZO/A	Hombre	63	76	151	33,33	DIABETES ME	Morbilidad	Primera	Obesidad Tif
1401	CENTRO DE SALUD CORDERO CRESPO	GUARANDA	Urbano	MESTIZO/A	Hombre	53	60,9	145	28,97	DIABETES ME	Morbilidad	Subsecuente	Sobrepeso

Anexo B: Captura de la encuesta realizada a los pacientes

ENCUESTA PARA DETERMINAR CRECIMIENTO DE PACIENTES CON DIABETES MILLITUS TIPO 2			
1	Realiza usted actividad física como correr, caminar ir al gimnasio ¿Cuántas horas realiza a la semana?	<input type="text"/>	GRABAR
2	¿Tiene el habito de fumar consumo usted tabacos?	<input type="text"/>	
3	¿algún familiar suyo fue diagnosticado con diabetes millitus tipo 2 o alguna enfermedad relacionada?	<input type="text"/>	LIMPIAR
4	¿Cual es el nivel de ingresos familiares ?	<input type="text"/>	
5	¿Cual es el nivel de formación académica ?	<input type="text"/>	

Anexo C: Código del estudio de variables en el software SPSS

```
GET
  FILE='C:\Users\mmona\Desktop\MAESTRIA DE MATEMATICAS -
ESPOCH\TITULACION\EDA_PRAS_2020\Base_Datos_PRAS2020.sav'.
DATASET NAME ConjuntoDatos1 WINDOW=FRONT.
CATPCA VARIABLES=Edad_Agrupada Raza Lugar_Destino Sexo_del_Paciente
Tipo_Diagnostico Cron_Diag
  IMC_Categoria Sedentarismo Consumo_Tabaco Antecedetes_familiaares
Nivel_Ingresos Grado_Academico
  lugar_residencia
  /ANALYSIS=Edad_Agrupada (WEIGHT=1, LEVEL=SPORD, DEGREE=2, INKNOT=2)
  Raza (WEIGHT=1, LEVEL=SPORD, DEGREE=2, INKNOT=2)
Lugar_Destino (WEIGHT=1, LEVEL=SPORD, DEGREE=2, INKNOT=2)
Sexo_del_Paciente (WEIGHT=1, LEVEL=SPORD, DEGREE=2, INKNOT=2)
Tipo_Diagnostico (WEIGHT=1, LEVEL=SPORD, DEGREE=2, INKNOT=2)
Cron_Diag (WEIGHT=1, LEVEL=SPORD, DEGREE=2, INKNOT=2)
IMC_Categoria (WEIGHT=1, LEVEL=SPORD, DEGREE=2, INKNOT=2)
Sedentarismo (WEIGHT=1, LEVEL=SPORD, DEGREE=2, INKNOT=2)
Consumo_Tabaco (WEIGHT=1, LEVEL=SPORD, DEGREE=2, INKNOT=2)
Antecedetes_familiaares (WEIGHT=1, LEVEL=SPORD, DEGREE=2, INKNOT=2)
Nivel_Ingresos (WEIGHT=1, LEVEL=SPORD, DEGREE=2, INKNOT=2)
Grado_Academico (WEIGHT=1, LEVEL=SPORD, DEGREE=2, INKNOT=2)
  lugar_residencia (WEIGHT=1, LEVEL=SPORD, DEGREE=2, INKNOT=2)
  /MISSING=Edad_Agrupada (PASSIVE, MODEIMPU) Raza (PASSIVE, MODEIMPU)
Lugar_Destino (PASSIVE, MODEIMPU)
  Sexo_del_Paciente (PASSIVE, MODEIMPU) Tipo_Diagnostico (PASSIVE, MODEIMPU)
Cron_Diag (PASSIVE, MODEIMPU)
  IMC_Categoria (PASSIVE, MODEIMPU) Sedentarismo (PASSIVE, MODEIMPU)
Consumo_Tabaco (PASSIVE, MODEIMPU)
  Antecedetes_familiaares (PASSIVE, MODEIMPU)
Nivel_Ingresos (PASSIVE, MODEIMPU)
  Grado_Academico (PASSIVE, MODEIMPU) lugar_residencia (PASSIVE, MODEIMPU)
  /DIMENSION=2
  /NORMALIZATION=VPRINCIPAL
  /MAXITER=100
  /CRITITER=.00001
  /RESAMPLE=NONE
  /PRINT=CORR LOADING
  /PLOT=OBJECT (20) CATEGORY(lugar_residencia Raza Lugar_Destino
Sexo_del_Paciente Tipo_Diagnostico
  Cron_Diag IMC_Categoria Sedentarismo Consumo_Tabaco
Antecedetes_familiaares Nivel_Ingresos
  Grado_Academico Edad_Agrupada ) (20) JOINTCAT(lugar_residencia Raza
Lugar_Destino Sexo_del_Paciente
  Tipo_Diagnostico Cron_Diag IMC_Categoria Sedentarismo Consumo_Tabaco
Antecedetes_familiaares
  Nivel_Ingresos Grado_Academico Edad_Agrupada) (20).
```

Anexo D: Análisis de componentes principales para datos categóricos

CATPCA - Análisis de componentes principales para datos categóricos

[ConjuntoDatos1] C:\Users\mmona\Desktop\MAESTRIA DE MATEMATICAS - ESPOCH\TITULACION\EDA_PRAS_2020\Base_Datos_PRAS2020.sav

Advertencias

Para la variable Destino de la Población se han especificado demasiados nudos interiores en la palabra clave INKNOT en el subcomando ANALYSIS. Se ha establecido en el número máximo de nudos interiores (0) para la variable.

Para la variable Sexo del Paciente se han especificado demasiados nudos interiores en la palabra clave INKNOT en el subcomando ANALYSIS. Se ha establecido en el número máximo de nudos interiores (0) para la variable.

Para la variable Tipo de diagnóstico se han especificado demasiados nudos interiores en la palabra clave INKNOT en el subcomando ANALYSIS. Se ha establecido en el número máximo de nudos interiores (0) para la variable.

Para la variable Complejidad del diagnóstico se han especificado demasiados nudos interiores en la palabra clave INKNOT en el subcomando ANALYSIS. Se ha establecido en el número máximo de nudos interiores (0) para la variable.

Para la variable Familiar diagnosticado con diabetes se han especificado demasiados nudos interiores en la palabra clave INKNOT en el subcomando ANALYSIS. Se ha establecido en el número máximo de nudos interiores (0) para la variable.

Crédito

CATPCA

Version 2.0

by

Leiden SPSS Group

Leiden University

The Netherlands

Anexo E: Resumen del modelo

Resumen de procesamiento de casos

Casos activos válidos	213
Casos activos con valores perdidos	414
Casos complementarios	0
Total	627
Casos utilizados en análisis	627

Anexo F: historial de interacciones.

Número de iteración	Varianza contabilizada para		Total	Pérdidas	
	Total	Aumentar		Coordenadas del centroide	Restricción del centroide en coordenadas de vector
0 ^a	5,866421	,000006	20,133579	19,630081	,503498
36 ^b	6,011410	,000010	19,988590	19,550030	,438560

a. La iteración 0 muestra las estadísticas de la solución con todas las variables, excepto las variables con un nivel de escalamiento óptimo de nominal múltiple, tratadas como numéricas.

b. El proceso de iteración se ha detenido porque se ha alcanzado el valor de prueba de convergencia.