



UNIVERSIDAD  
CATÓLICA  
DE CUENCA

**UNIVERSIDAD CATÓLICA DE CUENCA**

*Comunidad Educativa al Servicio del Pueblo*

**UNIDAD ACADÉMICA DE SALUD Y BIENESTAR**

**CARRERA DE ODONTOLOGÍA**

**¿QUÉ HACER CON LAS VARIABLES CATEGÓRICAS?**

**PROYECTO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL  
TÍTULO DE ODONTÓLOGO**

**AUTOR: SOLANGE GABRIELA MENA COYAGO**

**DIRECTOR: ING. AURA DEL CISNE GUERRERO LUZURIAGA**

**CUENCA-ECUADOR**

**2024**

**DIOS, PATRIA, CULTURA Y DESARROLLO**



**UNIVERSIDAD CATÓLICA DE CUENCA**

*Comunidad Educativa al Servicio del Pueblo*

**UNIDAD ACADÉMICA DE SALUD Y BIENESTAR**

**CARRERA DE ODONTOLOGÍA**

**¿QUÉ HACER CON LAS VARIABLES CATEGÓRICAS?**

**PROYECTO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL  
TÍTULO DE ODONTÓLOGO**

**AUTOR: SOLANGE GABRIELA MENA COYAGO**

**DIRECTOR: ING. AURA DEL CISNE GUERRERO LUZURIAGA**

**CUENCA-ECUADOR**

**2024**

**DIOS, PATRIA, CULTURA Y DESARROLLO**

# INDICE

<b>Capitulo 2: ¿Qué hacer con las variables categóricas?.....</b>	<b>4</b>
<b>1 Introducción.....</b>	<b>4</b>
<b>2 Definiciones de variables categóricas .....</b>	<b>5</b>
<b>3 Tipos de variables categóricas .....</b>	<b>7</b>
<b>3.1 Dicotómica .....</b>	<b>9</b>
<b>3.2 Politómica.....</b>	<b>9</b>
<b>4 Técnicas estadísticas que se pueden aplicar en estas variables.....</b>	<b>10</b>
<b>5 Gráficas que pueden aplicarse .....</b>	<b>15</b>
<b>6 Pareto y como funciona.....</b>	<b>19</b>
<b>7 Referencias bibliograficas .....</b>	<b>22</b>

## Capítulo 2: ¿Qué hacer con las variables categóricas?

### 1 Introducción

La ciencia es fundamental en pleno siglo XXI, por esa razón es importante comprender que, para cualquier clase de estudio de investigación es necesaria la recolección o búsqueda de información minuciosa, con el fin de obtener resultados de alta calidad. A medida que pasa el tiempo, se ha vuelto más complicado omitir la importancia de las variables dentro de un proyecto de investigación, ya que muchas de las veces estas suelen ser ignoradas llegando a cometer un gran error dentro del estudio, olvidando el hecho de que son las variables quienes están presentes desde el inicio del proyecto, es decir, desde el planteamiento de una idea hasta preguntarse en qué manera se miden, se proyectan o se relacionan con el resultado esperado (1). De hecho, dentro de la investigación de las variables todavía existen algunos vacíos en las asignaturas relacionadas a la investigación y, por lo tanto, se hace necesario resaltar el papel de función y clasificación de estas dentro de cualquier clase de estudio o investigación. Para ello se debe clasificarlas de acuerdo al comportamiento de sus datos: cualitativas o cuantitativas; y ya dependiendo de su naturaleza se establece el rol a cumplir, apoyándose en técnicas estadísticas categóricas o numéricas (2).

Es necesario conocer que las variables son las que van a generar estabilidad a cualquier proyecto de investigación, sin embargo, es necesario saber qué es la variable y si es correcto conceptualizar o más bien definir a la variable de un estudio (1). Por ello, se ha tomado como referencia al diccionario de la Real Academia Española, donde menciona lo siguiente: un concepto es el criterio u opinión de algo, mientras que la definición hace referencia a la proposición que expone con claridad y exactitud los caracteres genéricos y diferenciales de algo material. Por consiguiente, considerando que no existe una verdad absoluta, “el concepto” es el fruto de varias opiniones las cuales están basadas en sus conocimientos y puede llegar a influir el empirismo, al contrario de “la definición” que hace mención a una descripción exacta del objeto que está bajo observación, razón por la cual se puede decir que, una variable está definida como las características o aspectos de un estudio que pueden ser enumeradas o medidas (cualitativas o cuantitativa), las cuales van a estar sujetas a cambios y que además, van a ser trasladadas a una tabla operativa con el fin de obtener un correcto análisis, recuento y establecer a qué tipo de medición pertenecen (3).

## 2 Definiciones de variables categóricas

Las variables son propiedades que pueden ser medidas, observadas y analizadas en cualquier unidad de estudio de una investigación, por eso, es conveniente recordar que las variables se clasifican en variables numéricas y variables categóricas. En la primera situación, son de naturaleza cuantitativa, es decir, se encargan de la cuantificación, las mismas que pueden dividirse en variables de intervalo o variables de razón y cuya distribución es de tipo escalar, por lo que a cada una se le puede asignar un valor numérico menor o mayor (3). En cambio, al hablar de variables categóricas, el autor Sánchez (2) menciona que pertenecen a una naturaleza cualitativa, donde se observan los atributos o propiedades del objeto de estudio; a esta definición Arias añade que tanto la variable categórica como la numérica, son palabras que se hace notar desde el título del tema de investigación y que también debe encontrarse dentro del objetivo, problema e hipótesis general, por otro lado, Hernández (4) menciona que mediante estas variables se obtienen los datos reales necesarios para cumplir con la investigación de cualquier proyecto.

El autor Cuesta (3) añade a las definiciones anteriores, que las variables cualitativas se refieren a los atributos de los elementos que no pueden ser medidos en valores numéricos y, además, estas se construyen en base a los atributos sin imponer un orden entre ellos, sin embargo, Villasis (5) menciona que es verdad que no existe orden dentro de estos tipos de variables, pero la falta de orden hace referencia específicamente a las variables categóricas que tienen una escala de medición nominal, ya que también existe una de medición ordinal. En el primer caso, hace referencia a que sus datos son similares entre ellos, en otras palabras, no siguen un orden, por lo que enlistarlas de diferente manera no va a alterar el resultado o la relación que existe en ellas, un claro ejemplo en este punto, son los grupos sanguíneos (A-B, Positivo-Negativo), ninguno de estos grupos tiene más o menos valor si son colocados en diferente orden. Al contrario de las variables con escala ordinal, estas presentan elementos que siguen cierto orden, por ejemplo, una enfermedad periodontal puede clasificarse en leve, moderada o severa, sin embargo, dicho orden solo expresa que una es más grave que otra, pero esta característica no es tan específica como establecer grandes diferencias entre una unidad y otra.

Dentro de la escala de medición nominal de las variables categóricas, Espinoza (6) agrega que son atributos que se asocian en categorías ajenas, y que esta escala está dividida en dos grupos: Dicotómicas y Politómicas. También menciona sobre las variables categóricas, que, aunque no tengan una figura basada en números, esa característica hace que tenga relevancia, pues su definición como en su procesamiento estadístico presentan peculiaridades sobre las que es

importante analizar. En sí, las variables cualitativas por su etiología presentan características diferenciales, por ejemplo, al hablar de sexo solo hay dos opciones para clasificar a los individuos (femenino o masculino), pero si hablamos del nivel de enseñanza la situación es distinta ya que existen otros factores que intervienen dentro de la misma.

La autora Cienfuegos (7) menciona también que las variables categóricas nominales tienen una medición de baja categoría y que a pesar de que no es posible que ejecuten datos aritméticos, ni puedan imponer un orden; estas si pueden establecer tablas de frecuencias, de contingencia y porcentajes. Una propiedad fundamental de este tipo de variable es que buscan mantener igualdad o equivalencia, mientras que, en las variables ordinales, sus elementos se pueden corresponder entre ellas con signos de mayor y menor que, por el concepto de orden, en este caso un ejemplo puede ser el nivel de dolor que presente un individuo.

Otro punto de vista es el de Cauas (8), donde indica que las variables categóricas son un elemento fundamental en las investigaciones sociales, por ejemplo, en la investigación participativa, investigación etnográfica, investigación acción participativa, estudio de casos, etc., hacen uso constante de estas variables, ya que estos proyectos de investigación buscan detallar u obtener características exactas del o los objetos de estudio. Por otro lado, autores como Ochoa y Romero (9), mencionan que las variables categóricas, si son variables basadas en el uso de propiedades cualitativas, que se dividen en dos grupos, pero también mencionan, que como son basadas únicamente en atributos del fenómeno estudiado, no van a dar un número real, únicamente brindan al investigador una categoría, la misma que puede tener un conjunto limitado o fijo llevando consigo un orden o no, pero que, si es sumamente necesario tener a disposición un instrumento validado para las mediciones, porque caso contrario solo generará distorsiones en la investigación.

### 3 Tipos de Variables Categóricas

En los libros se pueden encontrar diversas clasificaciones de variables, lo que la mayoría de veces genera desconciertos en los científicos y también en la mayoría de los estudiantes. Para evitar dicho aspecto, se hace uso de la clasificación que fue planteada por Carballo en el año 2014, quien fundamenta esta lista a partir de diversas clasificaciones de numerosos autores, exponiendo entonces que las variables se pueden clasificar según la función, complejidad, naturaleza y su escala de medición. Considerando la función que presente la variable, esta puede ser, dependiente e independiente; una variable es independiente cuando el investigador puede manipularla con el fin de explicar al objeto de estudio, esta variable es la que explica la razón de los cambios de la variable dependiente, entonces se puede decir, que una variable es dependiente cuando se encuentra susceptible a ser modificada y son las responsables de las causas o efectos que van a dar origen a resultados dentro del proyecto de investigación. En relación a la complejidad de las variables, van a ser dos: Simples, cuando no se desintegran en más dimensiones; y complejas cuando si se componen de más de dos dimensiones y luego se determinan indicadores para cada dimensión y, allí es donde entra la clasificación según la naturaleza donde se explican dos tipos de variables, las cuantitativas y las cualitativas. Una variable es cuantitativa o numérica, cuando sus propiedades presentan un carácter numérico, al contrario de una variable de tipo cualitativa o conocida también como categórica, la cual se basan en elementos o atributos del objeto de estudio (5,6).

Tomando en cuenta la clasificación por escala de medición es preciso mencionar que, ambas variables (numéricas o categóricas), presentan escalas las cuales datan desde los años 90. Sin embargo, antes de hablar sobre las escalas de medición es necesario entender la palabra medición, ya que este término hace referencia al proceso de consignar, siguiendo las reglas ya definidas; valores numéricos a propiedades de fenómenos, y al mencionar propiedades hace mención a los aspectos observados o características empíricas, además, la medición incluye la evaluación, proceso por el cual se puede diferenciar las cosas. Por lo tanto, las escalas de medición son un conjunto de valores ordenados consecutivamente adoptando un extremo inicial y final, permitiendo que la variable pueda ocupar dichos valores para poder ser medida. Además, se menciona que el nivel en el que se puede medir a una variable, va a establecer; las cualidades de medición de dicha variable, los tipos de operaciones matemáticas que pueden ser empleadas de manera adecuada en el nivel, formulas y métodos estadísticos útiles para analizar datos y probar hipótesis teóricas. En pocas palabras, podemos decir que las escalas de medición son usadas para medir atributos y se clasifican en cuatro: las de intervalo y de razón, las cuales

pertenecen a variables cuantitativas; y las nominales y ordinales, quienes pertenecen a las variables categóricas y son de mayor interés del capítulo presente (10).

Con respecto a las variables cualitativas de tipo ordinal, esta consiste en ordenar categorías en forma jerárquica, tomando en consideración el grado de diferencia, pero sin expresar los datos sobre el tamaño de la misma, por ejemplo, el puesto de ganadores de alguna competencia, grado de dolor, etc.; en este punto son los símbolos quienes indican la posición (3). Si utilizamos números, la magnitud de estos no es arbitraria, sino que representa el orden del rango del atributo observado. Se supone un continuo subyacente en los números de modo que las relaciones típicas son, en este caso, «más alto que», «mayor que» o «preferible a». Sólo las relaciones «mayor que», «menor que» e «igual a» tienen significado en una escala de medición ordinal. Bajo una escala ordinal es posible clasificar u ordenar algunos objetos o eventos que tengan diversas cantidades de alguna característica, basados en la característica. Mientras que un tipo de variable nominal se basa en dividir los fenómenos del estudio según la categoría, esta clase se aplica en variables categóricas que son independientes, es decir, que no tiene relación entre ellas y que los números asignados solo sirven como etiqueta para identificar su categoría, un ejemplo es el número que se le asigna a un concursante (10,11).

Las variables nominales categóricas constituyen la forma más elemental de escala y representan el método de medición más antiguo. En una escala como esta, las unidades de aprendizaje, ya sean, individuos u objetos, se organizan o se distribuyen en categorías, considerando las características, propiedades o elementos distintivos de la variable que ha sido observada; y por consiguiente asignando un nombre específico a cada categoría. Se debe considerar que, si bien cada categoría presenta un nombre en específico, no necesariamente tiene que ser un nombre alfanumérico o alfabético de manera obligatoria, al contrario, se puede hacer empleo de números para designar las categorías. En una escala nominal, los valores numéricos atribuidos determinan a cada uno de los grupos y sirven de manera específica para etiquetar o identificar, de hecho, los números otorgan mejores diferencias categóricas más que en cantidades, ya que, efectúan una función netamente de clasificación, la cual no está sujeta a manipulaciones aritméticas porque cada número representa una categoría distinta. Por ejemplo, no existiría sentido calcular un promedio final de veinte códigos de estudiantes, ya que la medida de los números no representa un orden de mayor a menor, porque los números simplemente funcionan como códigos de identificación y no reflejan magnitudes o jerarquías.

No es posible organizar las observaciones de manera ascendente o descendente, ya que ninguna de las categorías tiene una jerarquía superior a las demás; simplemente, están

señalando variaciones en la variable. Las preguntas relacionadas con la magnitud de la variable subyacente carecen de relevancia en la medición nominal. No obstante, es válido plantear que la única cuestión comparativa significativa para los datos nominales radica en determinar si dos observaciones son idénticas o no. Dentro de la escala nominal perteneciente a la variable categórica, se pueden presentar dos tipos: Dicotómica y Politómica (10,11).

### **3.1 Dicotómica**

Las variables categóricas de tipo dicotómica solo presentan 2 categorías, por eso también son conocidas como binarias. Un ejemplo puede ser el sexo, en este caso solo nos referimos a femenino o masculino (6).

### **3.2 Politómica**

Al hablar de una variable categórica de tipo politómica hace referencia a cuando las variables se pueden expresar en más de dos categorías, por ejemplo, el color de cabello de alguna persona, en este caso podemos encontrar, rubio, café, negro, entre otros (6).

Es fundamental mencionar que muchas veces la investigación suele recurrir a la conversión de las politómicas en dicotomías, en la investigación social, a fin de mejorar la estrategia de contrastación de la hipótesis, para lograr mayor precisión con sólo dos niveles de variabilidad. Por ejemplo, es frecuente el uso de la medición del rendimiento académico que puede expresarse en rasgos con decimales, vigesimales, centesimales, puede también ser expresada con una dicotomía: aprobados, desaprobados. Por lo tanto, es posible que una politomía se convierta una dicotomía, pero una dicotomía no puede convertirse en politomía(12).

#### **4 Técnicas estadísticas que se pueden aplicar en estas variables**

En la mayoría de los casos y en el transcurso del tiempo, la estadística ha sido detestada por parte de los estudiantes, no obstante, los investigadores y profesionales aprecian y valoran a esta ciencia, pues resulta ser de las más útiles durante la elaboración de un proyecto, ya que la estadística recopila información de ciertos acontecimientos, personas o comunidades, obteniendo análisis, resultados exactos o predicciones futuras, posibilitando tomar buenas decisiones para la investigación, y por esa razón, la estadística es uno de los instrumentos más usados en toda la literatura científica y que, además, es indispensable en varias áreas científicas y cotidianas; se podría decir en pocas palabras, que la estadística es una ciencia transversal. Entonces, la estadística va a ser la ciencia que se encarga de recopilar, ordenar, analizar e interpretar datos con el fin de tomar decisiones acertadas, correctas y efectivas, sin embargo, existen algunos autores que tienen ideas diferentes sobre la definición de estadística; gran parte de investigadores la definen como una ciencia cuyo objetivo es el análisis cualitativo, mientras que otros, la definen como una expresión cuantitativa y cualitativa condicionada para el análisis de los datos (13,14).

Es importante entender que al mencionar el termino estadística, son pocas las personas que tienen clara la definición de la palabra, y por tal razón, muchos individuos al no saber la definición clara de la estadística, interpretan la información de manera incorrecta, causando de tal manera, un falso análisis e interpretación de la información. De todas las definiciones sobre la estadística, la más acertada es la de Mínguez (13) quien la define de la siguiente manera: “La ciencia tiene por objeto aplicar las leyes de la cantidad a los hechos sociales para medir su intensidad, deducir las leyes que los rigen y hacer su predicción próxima”, por lo que quiere decir que, dentro de la investigación, la estadística cumple un rol como instrumento o herramienta fundamental para entender los resultados de un análisis. Se destaca el hecho de que la estadística es una herramienta encargada de recolectar, agrupar, representar, analizar, e interpretar datos, lo cual facilita la percepción de los datos, y que, además, es una técnica científica que tiene como finalidad establecer conclusiones a partir de observaciones realizadas (15).

Se debe considerar que la estadística como ciencia presenta métodos y procedimientos que son utilizados en los diferentes campos del área técnica y científica. Muchos investigadores, sean personas de negocios, empresarios, profesionales, administradores públicos o privados, trabajando en un entorno propio a su actividad requieren información sobre características o comportamientos de dicho entorno, tanto para progresar en el conocimiento del problema, indagar sus causas, como para tomar decisiones. La estadística, es la ciencia de la

experimentación; y, por lo tanto, tener conocimiento acerca de las técnicas estadísticas ayuda a tratar datos experimentales y a programar experimentos de una forma más eficiente (16). Como se mencionó dentro de la estadística se encuentran las técnicas estadísticas, las mismas que se aplican de manera amplia en mercadotecnia, contabilidad, control de calidad, estudios de consumidores, análisis de resultados en deportes, administración de instituciones, en la educación, organismos políticos, en la medicina y en otras muy distintas áreas como un auxiliar en la toma de decisiones. Sin embargo, antes de profundizar sobre el uso y tipos de técnicas estadísticas, es importante conocer con exactitud que son las técnicas estadísticas (13).

La metodología de investigación científica se estructura como un método estándar, respaldado por la experiencia práctica, cuyo propósito principal es adquirir y procesar información relevante para abordar problemas de conocimiento en diversos campos científicos. Por lo tanto, en cualquier proyecto de investigación, se otorga una importancia fundamental a aspectos como los tipos de variables, escalas, análisis estadístico y técnicas estadísticas, dado que estos elementos desempeñan un papel crucial en el proceso de obtención y análisis de datos con el fin de avanzar en la resolución de los problemas planteados (17). Las técnicas estadísticas se utilizan en diversos campos de la ciencia, incluida la medicina, y desempeñan un papel fundamental como herramientas matemáticas para analizar las relaciones de causa y efecto, con el tiempo, han sufrido cambios importantes y han evolucionado con los avances en los campos tecnológico y científico (18), sin embargo, en muchos estudios se han observado problemas repetidos y son causados por el uso inadecuado de las estadísticas de las pruebas; este fenómeno puede deberse a varios factores, entre ellos: falta de conocimiento de estadística descriptiva e inferencial, falta de dominio de los métodos de investigación y falta de conocimiento en el uso de software estadístico (19). Hay dos tipos de técnicas aplicadas a variables. Uno es un método paramétrico que se utiliza a menudo con variables numéricas y el otro es un método no paramétrico que es más compatible con variables categóricas (7).

Las pruebas de tipo no paramétricas son sencillas de aplicar, los datos presentan jerarquía, única opción para muestras pequeñas, para calcular se ordena por rangos o frecuencias. Las estadísticas no paramétricas, como sugiere el nombre, no se refieren a parámetros como la media o la desviación estándar y, por lo tanto, no pueden realizar operaciones aritméticas; por tanto, ésta es la técnica más adecuada para escalas de baja categoría, en este caso se mencionan variables nominales y ordinales, y, además, cuando los supuestos del modelo son débiles o no suelen cumplirse. Las escalas nominales corresponden a la naturaleza de los datos que se han clasificado en categorías o clases, por ejemplo, en el caso de las variables de género

(masculino y femenino), cada una es una categoría o una clase, en cuyo caso no se pueden realizar operaciones aritméticas ni establecer relaciones de orden, sin embargo, la frecuencia, el porcentaje y el recuento se pueden determinar con una tabla de contingencia. Las pruebas estadísticas apropiadas para probar la hipótesis son la prueba de chi-cuadrado y la prueba binomial, que muestra la frecuencia de cada categoría (7,19).

En las pruebas de tipo no paramétricas, la distribución de datos no es conocida con anterioridad y no se encuentra relacionada a ninguna clase de distribución de probabilidad, y por esa razón, algunos autores la denominan como la prueba de dispersión libre. Es más, las ciencias sociales y la pedagogía, hacen gran uso de pruebas no paramétricas para realizar el análisis de su proyecto, lo que hoy en día dentro de esas áreas se considera común o normal, ya que existen diversas variables que no se encuentran ligadas a condiciones de tipo paramétricas (19). Entonces, haciendo una síntesis sobre las pruebas no paramétricas, se puede decir que, presentan pocas restricciones, el análisis estará basado en rangos, la hipótesis se realizará bajo rangos, medianas y el uso de frecuencias y esta prueba de dispersión libre puede ser aplicada con tamaños de muestra pequeños. Existen muchas clases o tipos de pruebas no paramétricas, pero las que mayormente son utilizadas dentro del área de la salud son las siguientes: Prueba de U de Mann-Whitney, Prueba de Wilcoxon, B, Prueba de Kruskal-Wallis, Prueba de Friedman (20).

Empezando por la prueba de U de Mann-Whitney, esta fue propuesta por Mann y Whitney con el fin de comparar 2 muestras independientes y aleatorias de una misma población, así como para probar si dos muestras son iguales. Esta prueba surge como alternativa a la prueba T Student la cual pertenece al ámbito paramétrico ya que compara medias, al contrario de la prueba de U Mann-Whitney que compara las relaciones que existen entre dos valores de la mediana basándose en rangos, entonces, la diferencia entre la prueba de t Student es que la prueba U Mann-Whitney dependiendo de los supuestos que se hagan sobre la distribución de la misma, se van a extraer diferentes conclusiones de los datos. Continuando con la siguiente prueba, es la de Wilcoxon, esta prueba es algo parecida a la anterior, ya que también busca comparar dos muestras relacionadas, sin embargo, la diferencia radica en que la prueba de Wilcoxon busca determinar si existen diferencias entre ellas y no busca ver si hay similitudes. Al tratarse de un tipo de prueba no paramétrica no es obligatorio que los datos se dispongan de manera normal. Pasando a la siguiente prueba que es la Kruskal-Wallis, esta sirve para confirmar si un grupo de datos proviene o no de una misma población, además esta prueba es una ampliación o una opción alterna a la prueba paramétrica ANOVA. Finalmente, la prueba de

Friedman es usada para analizar datos que se presentan repetidamente, el uso de esta prueba es principalmente ejecutada cuando no se cumplen con los ámbitos de normalidad y homogeneidad de las varianzas (19,20).

El uso de pruebas estadísticas facilita el proceso de tabulación de los datos, sin embargo, es preciso recordar que antes de hacer uso de dichas pruebas, se debe de contar con instrumentos los cuales son necesarios para la recolección de datos. La parte del proceso de investigación que requiere más mano de obra, es la selección del instrumento, ya que este servirá como medio para recopilar datos con el fin de cumplir con los objetivos del estudio. Es más, la selección del instrumento puede ser considerada como la parte principal del trabajo, pues los resultados del estudio se basan en la información procesada, la cual es producto de los datos recolectados por dichos instrumentos contruidos. Entre los varios instrumentos que se pueden aplicar dentro las variables categóricas encontramos a la encuesta, entrevista, observación y revisión documental (21).

La encuesta es un método para obtener datos o información de manera directa de los individuos que están involucrados en un tema de investigación; muchas veces se suele confundir a la encuesta con la entrevista, sin embargo, se puede diferenciar de la entrevista por el menor nivel de interacción que existe con los participantes. La recopilación de datos se va a obtener por medio de cuestionarios, pruebas o evaluaciones de conocimientos. Dentro de la encuesta podemos encontramos diversas opciones para obtener los datos, la más famosa es la escala de Likert que es un tipo de cuestionario que consta con una serie de opciones o enunciados presentado a los participantes con el fin de obtener una respuesta por parte de ellos. Por lo general, consta de cinco ítems de respuesta para evitar respuestas extremas por parte del encuestado y dichas respuestas se les otorga una puntuación con el fin de calcular un promedio. Otro tipo de escala es la de Guttman, este escalograma presenta sus ítems de manera ordenada, de tal manera que la declaración positiva de una de ellas significa la afirmación del elemento anterior, en pocas palabras, sustenta que algunos factores son más indicativos de intensidad o fuerza de una actitud. Por último, se encuentra la escala de Thurstone, que consta de dos opciones de respuesta, la cual una suele ser positiva y la otra negativa, y posterior se promedian dichas respuestas y un grupo de expertos seleccionan las más relevantes (21).

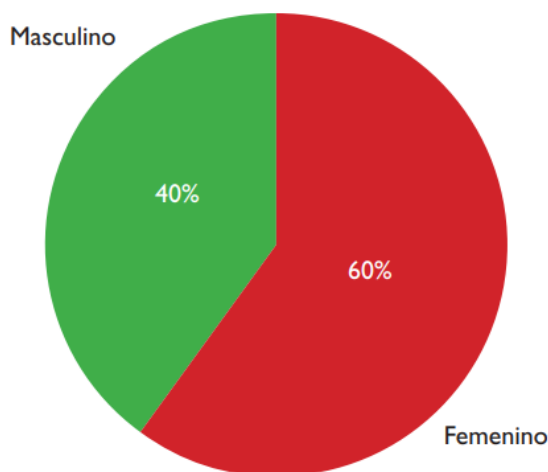
Otra herramienta que se emplea para la toma de recolección de datos es “La entrevista”, la cual es evento que ocurre frente a frente entre dos individuos, en esta actividad el entrevistador obtiene de manera verbal, información del participante, en este caso se habla del entrevistado, además de aquello, también se consigue conocer la opinión del individuo, así como intercambiar

criterios acerca de un tema en específico. Se puede mencionar otra opción de herramienta como la “Técnica de Observación” donde el investigador busca conectar con la realidad y establecer una noción sobre el dilema de su estudio, sin embargo, esta técnica no es tan precisa, ya que los hechos pueden ser interpretados desde la subjetividad del investigador y dar conclusiones erróneas, por esa razón, no es una herramienta que brinde el cien por ciento de veracidad. Para finalizar, se menciona la “Técnica de revisión documental”, la cual consiste en una indagación completa de literaturas y documentos de un asunto en particular, con el fin de escoger distinta información sobre la variable de estudio y de tal forma, mejorar las ideas o conceptos sobre el tema de investigación (21). En la actualidad el uso de diferentes métodos, herramientas y técnicas dentro de un proyecto de investigación, hace que el trabajo del investigador sea más sencillo, ya que, gracias a los distintos sistemas estadísticos encargados del procesamiento de datos, evita que el científico realice las actividades más complejas de cálculo, permitiéndole cumplir con tareas como la elección del diseño, de pruebas y la interpretación y análisis de los resultados (22).

## 5 Gráficas que pueden aplicarse

La estadística descriptiva es una rama de la estadística, la cual comúnmente se aplica con el fin de obtener el primer acercamiento con los resultados de una investigación y es allí donde entran las representaciones gráficas de las variables, ya que estas se usan con frecuencia para tener mayor comprensión de las mismas. Los gráficos son representaciones visuales de un conjunto de datos resultando ser un medio eficaz, porque un gráfico bien diseñado, logra captar la atención de los lectores y además la información será presentada de manera clara, sencilla y concisa evitando de tal manera errores en el procesamiento de información y, por lo tanto, permite una mejor comprensión de las relaciones, diferencias y objetivo principal de la variable de estudio. Al igual que las pruebas y técnicas estadísticas, los gráficos deben ser seleccionados según el objetivo del proyecto de investigación, el tipo de variable (categórica o numérica), y según el análisis que se desea explicar, debido a que un gráfico no adecuado induce a una incorrecta interpretación y los mismos pueden llegar a convencer a los científicos a inferencias erróneas (23,24).

Figura 1. Diagrama de sectores: Distribución de la variable sexo

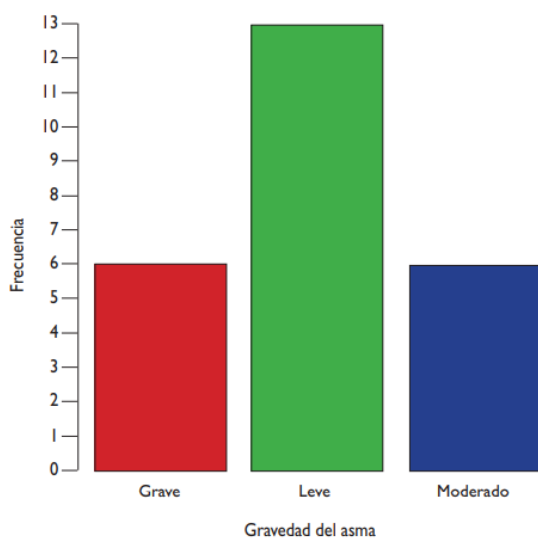


Fuente: Tomado de Ortega Páez E, et al. (2019)

Es importante recordar que las variables cualitativas son aquellas que describen categorías o rasgos de una variable. Cuando las variables no presentan un orden específico, se está frente a una variable cualitativa con escala nominal, mientras que, si una de la variable establece o incluye una jerarquía se estaría hablando de una variable de tipo ordinal. Por ejemplo, la variable "Alcohólico" sería de tipo nominal si las opciones de respuesta se limitan a si o no, de lo contrario,

si la variable presenta más ítems, tales como: “Nunca”, “A veces”, “Casi siempre” o “Siempre”, se establece que la variable es de tipo ordinal, pues ya existe un grado de jerarquía. Existen dos tipos de representaciones gráficas bien diferenciadas, según sean variables cualitativas categóricas nominales u ordinales, la primera se representa con el gráfico de sectores y la segunda con un gráfico de barras (23).

Figura 2. Gráfico de Barras: Distribución de la variable "Gravedad del asma"



Fuente: Tomado de Ortega Páez E, et al. (2019)

Empezando por el gráfico de sectores o también conocido gráfico de pastel o pizza (figura 1), su forma es circular y sus secciones o partes pueden representar un porcentaje que será directamente equivalente al de su frecuencia. La suma total de sus partes equivale al total de las frecuencias. De este modo, cada categoría presentará diferentes tamaños en su área, unas serán de mayor dimensión mientras que otras apenas se harán resaltar, dando a notar la distribución de la frecuencia de cada variable. Los datos idóneos para este tipo de figuras son los cuantitativos continuos, de manera especial los porcentajes. La simplicidad en elaboración lo catapulta como el más usado en análisis de datos descriptivos. Entre las ventajas al momento de su elección, por ejemplo, resulta muy útil cuando existe un número bajo de variables (sexo, estado civil, etnia, etc.), porque cuando existen numerosas variables el diseño de la ilustración se vuelve confuso y difícil de comprender, y eso es gracias a que las divisiones se solapan y pierden su claridad, no obstante, se puede complementar al gráfico con una tabla de frecuencias para mejorar la

interpretación de los datos. Sin embargo, existe una gran desventaja al aplicar este tipo de gráfico, se debe tomar en cuenta que, para hacer empleo del mismo, la información o los datos deben ser estables y no estar expuestos a cambios, ya que, si son datos temporales, es decir, que con el tiempo van evolucionar, este gráfico no es la mejor opción para obtener resultados y análisis exactos (23,25).

El gráfico de sectores sirve de manera principal para representar las variables cualitativas, específicamente las de escala nominal, además, la elaboración de este gráfico es fácil de llevar a cabo, ya que el total de los datos equivale a los 360 grados de la circunferencia, entonces para obtener el grado de cada área, se debe multiplicar la frecuencia relativa de cada una de las variables por el número total de grados de la circunferencia, es decir, los 360 grados. Para comprender mejor, se puede observar que la figura 1, está representando a la variable "Sexo", la cual presenta 25 observaciones totales, de las cuales, 10 son del género masculino y las 15 observaciones restantes son del género femenino, estos valores corresponden a una frecuencia relativa de 0,4 que equivale a 144 grados para el género masculino y 0.6 para el género femenino equivaliendo 215 grados. Estos datos se obtienen ya que se divide el número de observaciones de cada género, sobre las observaciones totales y dicho valor obtenido se multiplica para los 360 grados que corresponde al total de la circunferencia del diagrama. Por esa razón, se puede observar en la figura 1 como la variable del género femenino tiene mayor frecuencia sobre la del género masculino (20,23). Es necesario saber que si se escoge un gráfico de sectores para una variable tipo ordinal, no se vería reflejada la jerarquía de las mismas, por lo cual, resulta más factible hacer empleo de otro tipo de gráfico que permita imponer un orden entre los niveles, en este caso, se habla del "Gráfico de Barras"; este diagrama (figura 2), es el idóneo para interpretar a las variables categóricas ordinales, aunque también puede ser usado en las variables numéricas discretas (25).

Para comprender mejor la aplicación del gráfico de barras o de columnas, es necesario saber que este diagrama es un modo de representar gráficamente la información o los datos recolectados, por medio de formas rectangulares, las mismas que pueden ser horizontales o verticales, generalmente conocidas como barras. El gráfico de barras vertical es el más usado, los valores de carácter categórico se plasman en el eje X, es decir, en la línea horizontal, en tanto que, los datos numéricos estarán dispuestos en todo el eje vertical, es decir el eje Y. Cuando se requiere comparar datos entre las distintas categorías que puede presentar una variable, el gráfico de barras vertical, es el más apto para representar estas diferencias. Mientras que, en el gráfico de barras horizontal, la distribución de los ejes es invertida, es decir, los valores numéricos

se colocaran en el eje horizontal y los valores categóricos en el eje vertical. Generalmente a este tipo de diagrama, se los emplea cuando los nombres de las categorías son extensos o cuando existe numerosos valores numéricos y por eso requieren más amplitud para que exista mejor comprensión lectora (25,26).

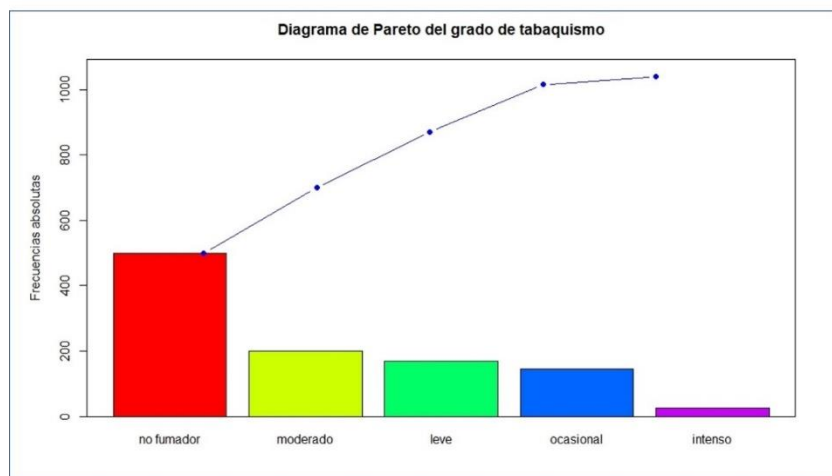
Estas clases de gráficos, permiten comparar visualmente cantidades y frecuencias, lo que hace más fácil y sencillo la interpretación de datos, y cabe mencionar que la dimensión de las barras va a depender totalmente del valor que represente dichas comparaciones. En el lado horizontal del gráfico estarán dispuestas las distintas categorías que presenta la variable en cuestión, mientras que en su parte vertical estarán ubicadas barras o columnas, por ejemplo, en la figura 2, la variable que representa "Asma" consta con tres jerarquías, empezando con leve, luego moderada y al final grave, donde se puede observar que la categoría con mayor frecuencia es de asma leve. Estos diagramas se utilizan generalmente en diversos campos, como las ciencias sociales, en administración de empresas, en el área de la economía, y en las ciencias de la salud. El prestigio de este gráfico se debe a su gran aptitud de mostrar datos de modo simple, conciso y de fácil comprensión, facilitando el análisis y toma de decisiones apoyado en los antecedentes, por esa razón, la elaboración de este diagrama es fácil de realizar, ya que se hace uso de los programas estadísticos, pero la desventaja de este diagrama es su capacidad de reflejar las discrepancias entre cada categoría de las variables, esta aptitud hace que sea sumamente manipulable los ejes de sus categorías; por esa razón, se debe ser cauteloso al realizar los análisis con este tipo de gráficos para evitar transmitir un mensaje erróneo (23,25).

Los gráficos de barras son sumamente polifacéticos y se pueden usar con diferentes fines, por ejemplo, cuando se quiere realizar un análisis de las tendencias actuales, comúnmente se aplica un gráfico de líneas, sin embargo, resulta útil el gráfico de barras, ya que puede existir cambios en los valores, y por lo tanto cuando los datos son discretos y no continuos, el diagrama de barras es el mejor para ser empleado. También es ideal cuando se requiere la comparación de cantidades o frecuencias de diferentes categorías o grupos; y en otro ámbito como la presentación de valores o datos, en esta área es clara la elección ya que el gráfico de barras permite visualizar dicha información de manera fácil por lo que el análisis de la información es sencillo de entender. En pocas palabras, las opciones de gráficos, pruebas y técnicas estadísticas van a depender del tipo de variable y el número de las mismas, frente al que se encuentre el proyecto de investigación. Por lo tanto, como ocurre con cualquier prueba estadística, los gráficos deben elegirse correctamente para evitar diferencias, similitudes y resultados de variables erróneos (26).

## 6 Pareto y su funcionamiento

Pareto fue un italiano nacido en París por mediados del siglo XIX, el cual se hizo popular por diversas aportaciones al mundo científico, tales como la distribución, la eficacia, el índice y el principio de Pareto; y para explicar este último, desarrolló el diagrama de Pareto. Este diagrama también es conocido como “Línea de curva 80-20”, el cual establece una gráfica cuyo objetivo es ordenar los datos de manera que al final queden descendentemente o de izquierda a derecha. Esta disposición permite establecer prioridades a las diferentes categorías, reforzando entonces, el hecho de que en la mayoría de grupos de componentes que se les atribuye un efecto similar, solo unos cuantos son los responsables de causar dichos efectos; es decir, que el 20 % de las causas va a generar el 80% de los efectos. El gráfico de Pareto brinda la oportunidad de encontrar si existen ciertas irregularidades en los datos de algún proyecto de investigación y reconocer las partes donde requiere una retroalimentación con el fin de poner en marcha acciones que corrijan esos errores. Otra característica de este diagrama es que, presenta una técnica sencilla para plasmar cálculos simples y además realiza análisis de los distintos grupos de cada variable y les asigna la prioridad que les corresponde (23).

Figura 3. Diagrama de Pareto: Grado de Tabaquismo



Fuente: Tomado de Blum Rh, et al. (2020)

Este gráfico (figura 3) integra en su diagrama una línea de curva con el fin de representar el nivel de frecuencia acumulada de cada categoría que presente la variable, lo que se busca con esta

línea de curvatura, es exponer que existen diversas problemáticas que no presentan tanta importancia, frente a otros pocos que, si tienen relevancia, por lo que resulta útil aplicar esta grafica para enfocar todo el interés en la búsqueda de soluciones adecuadas que permitan tomar buenas decisiones para el futuro de cualquier proyecto de investigación. Como es comprensible, esta preferencia hace que este diagrama sea muy práctico para representar a las variables categóricas con escala ordinal, por lo que se puede decir que supera al diagrama de barras ya que este grafico brinda información sobre la frecuencia acumulada. Entonces, el objetivo de este diagrama de Pareto es identificar las principales necesidades las cuales requieren de mayor esfuerzos y recursos y tal manera establecer cuáles son las prioridades del proyecto de estudio, facilitando, por lo tanto, la toma de decisiones (26).

Es importante recordar que el diagrama de Pareto al tener como objetivo ordenar de forma descendente los datos, a manera que se establezcan prioridades, va a estar conformado por dos ejes, los cuales son el eje "Y" derecho, en el cual se representan al porcentaje acumulado de una variable y por otro eje "Y" en la parte izquierda que representa la cantidad de frecuencia, mientras que en la parte de abajo se encuentra el eje "X" en la cual se describen las razones, problemas, defectos, entre otros. Es por eso que, gracias a estas características en su diseño, este famoso diagrama puede ser empleado en distintas áreas, tales como, áreas de ventas donde se establece que el 80% de las ventas van estar relacionadas al 20% de los productos expuestos y también se determina que el 20% de los clientes son los que van a generar que exista el 80% de ganancias gracias a las ventas provocadas por estos clientes más fieles. También es aplicado en el área de sugerencias y reclamos, donde se determinará que el 20% de los productos que sufran de rechazos son representados por el 80% de las quejas emitidas por clientes; otra área donde puede ser aplicada esta gráfica es en el control de calidad, igual que con las anteriores áreas se establecerá el 20-80%, donde el 20% son los defectos que van a estar afectando a el 80% de los procesos. Queda claro entonces, que una de las grandes ventajas que brinda este diagrama elaborado por Pareto, es que establece cuales son los puntos más importantes con el fin de buscar soluciones exactas que permita analizar y elegir las opciones mas beneficiosas para el proyecto de investigación que se lleva a cabo (27).

Comprender como se elabora un diagrama de Pareto es de suma importancia, ya que no se puede realizar un análisis correcto si se tiene una estructura mal diseñada, por eso, existen diversos softwares estadísticos que pueden ayudar a la elaboración de este diagrama, por ejemplo, R-Studio, sin embargo, este programa no cuenta con una formula sencilla para la creación rápida del diagrama, pero puede aplicarse las ordenes textualmente para que inicie la

representación del diagrama, tal y como se puede observar en la figura 4. Para empezar la creación del diagrama de Pareto, el primer paso a realizar es establecer la problemática a tratar, una vez definido el problema se deberá realizar la recolección de datos y posterior la agrupación de los mismos, en este punto los datos deberán estar divididos, ya sea por categorías o frecuencias. El tercer paso involucra la organización de los datos, el orden de la frecuencia deber ser de forma descendente y después se debe añadir la frecuencia acumulada; posterior a este paso se debe añadir una columna donde se deben plasmar los valores porcentuales, este valor se obtiene mediante la división de la frecuencia y el total de frecuencias, es decir, la frecuencia acumulada, luego al valor resultante se debe multiplicar por 100 y el valor obtenido será el porcentaje final. Luego se delinea el gráfico utilizando los datos de la tabla anterior, se inserta el eje vertical, se insertan los datos (en orden descendente) y se agregan las barras. Para datos de porcentaje acumulado, cada barra tiene un punto central y está conectada por una línea. Finalmente, se analiza el diagrama y se agregan datos al final del diagrama, como fecha, título, detalles de la investigación y acciones correctivas aplicadas (28).

Figura 4: Script diagrama de Pareto

```
# Construir la tabla con las frecuencias de cada categoría
tabaquismo <- c(500, 145, 170, 200, 25)
grado <- c("no fumador", "ocasional", "leve", "moderado", "intenso")
datos <- data.frame(tabaquismo, grado)
# Reordenar en orden descendente
datos <- datos[order(datos$tabaquismo, decreasing = TRUE), ]
# Convertir los grados de tabaquismo en factor
datos$grado <- factor(datos_$grado, levels = datos_$grado)
# Calcular la frecuencia acumulada total de las variables
datos$acumulada <- cumsum(datos$tabaquismo)
# Construir el diagrama de Pareto
barplot(datos$tabaquismo, names.arg = datos$grado,
        main = "Diagrama de Pareto del grado de tabaquismo", #Título del gráfico
        ylab = "Frecuencias absolutas", # Título del eje y
        ylim = c(0, sum(datos$tabaquismo)*1.05), # Límites del eje y
        col = rainbow(length(datos$grado))) # Lo coloreamos
# Dibujamos la línea de frecuencia acumuladas
lines(1:length(datos$tabaquismo), datos$acumulada, type = 'b', pch=16, col="blue")
# Dibujamos una caja alrededor del gráfico
box()
```

Fuente: Tomado de Blum Rh, et al. (2020)

Concluyendo, es importante recordar que los diagramas de Pareto permiten asignar un conjunto de prioridades y dedicar el esfuerzo y tiempo al 20% de las causas que resuelven el 80% de los problemas. Esto permite ser más proactivo y eficaz al enfocar toda la concentración e interés en completar actividades que brinden resultados excelentes (28).

## 7 Referencias bibliográficas

1. Oyola A. La variable. Revista del Cuerpo Médico Hospital Nacional Almanzor Aguinaga Asenjo. 2021;14(1):90–3.
2. Sánchez Martínez B, Rodríguez Plasencia A, Vilema Vizuete E. Vista de Caracterización de las variables para la investigación en ciencias de la salud. Revista Universidad y Sociedad. 2022;14(3):455–64.
3. Cuestas E. Definición de variables. Revista de Facultad de Ciencias Médicas. 2019;66(3):113–7.
4. Arias J. Guía para elaborar la operacionalización de variables. Revista Espacio I+D Innovación más Desarrollo. 2021;10(28):42–56.
5. Villasís A., Novales M. El protocolo de investigación IV: las variables de estudio. Rev Alerg Mex. 2019;63(3):303–10.
6. Espinoza E. Las variables y su operacionalización en la investigación educativa. Parte I. Revista Conrado. 2019;14(65):39–49.
7. Cienfuegos A, Cienfuegos A. Lo cuantitativo y cualitativo en la investigación. Un apoyo a su enseñanza. Revista Iberoamericana para la Investigación y el Desarrollo Educativo. 2019;7(13).
8. Cauas D. Definición de las variables, enfoque y tipo de investigación. Rev Alerg Mex. 2018;
9. Ochoa J, Romero Y. El estudio descriptivo en la investigación científica. Universidad Autónoma del Perú. 2020.
10. Coronado J. ESCALAS DE MEDICIÓN. Corporación Universitaria Unitec. 2018;2(2):104–25.
11. Espinoza E. Las variables y su operacionalización en la investigación educativa. Segunda parte. Revista Conrado. 2019;15(69):171–80.
12. Núñez M. LAS VARIABLES: ESTRUCTURA Y FUNCIÓN EN LA HIPÓTESIS. Investigación Educativa. 2019;11(20):163–79.
13. Barreto A. El progreso de la Estadística y su utilidad en la evaluación del desarrollo. Consejo de Investigación y Evaluación de la Política Social. 2012.
14. Jacho A, Loaiza I, López J. La importancia de la estadística para el éxito de resultados en una investigación. 2020;3(1):32–41.
15. Arredondo E, Gómez R, Lalama R, Chóez L. Investigación científica y estadística para el análisis de datos. Revista Dilemas Contemporáneos: Educación, Política y Valores. 2020;8(1).
16. Micheli E. Estadística aplicada a la investigación. Pontificia Universidad Católica del Perú. 2019.
17. Rojas IR. ELEMENTOS PARA EL DISEÑO DE TÉCNICAS DE INVESTIGACIÓN: UNA PROPUESTA DE DEFINICIONES Y PROCEDIMIENTOS EN LA INVESTIGACIÓN CIENTÍFICA. Revista Internacional de Investigación Educativa. 2011;12(24):277–97.

18. Ságaro del Campo N, Zamara L. Técnicas estadísticas para identificar posibles relaciones bivariadas. *Revista Cubana de Anestesiología y Reanimación*. 2020;19(2).
19. Ramírez AF, Polack AM. Estadística inferencial. Elección de una prueba estadística no paramétrica en investigación científica. *Horizonte de la Ciencia*. 2020 May 12;10(19).
20. Ortega Páez E, Ochoa Sangrador C, Molina Arias M. Pruebas no paramétricas. *Evid Pediatr*. 2021;17:17–37.
21. Useche M, Artigas W, Queipo B, Perozo E. TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS CUALI-CUANTITATIVOS. In: *Proceso de recolección de datos cuali-cuantitativos*. Universidad Guajira. 2019. p. 11–85.
22. López Tamayo P, Pericj Company A. Uso de métodos y técnicas estadísticas en las investigaciones educativas. *ROCA Revista científico*. 2016;12(2):29–37.
23. Blum RH, Muret Wagstaff S, Cooper J, Petrusa E. Columnas, tartas y un italiano ilustre. Representación gráfica de variables cualitativas. *ReaR*. 2020;12(5).
24. Rivas Ruiz R, Roy García I, Pérez Rodríguez M, Berea R, Moreno Palacios J, Moreno Noguez M, et al. Pertinencia e impertinencia de los gráficos en la investigación clínica. *Rev Alerg Mex*. 2020;67(4):381–96.
25. Ortega Páez E, Ochoa Sangrador C, Molina Arias M. Representación gráfica de variables. *Evid Pediatr*. 2019.
26. Escamilla Martínez P, Álvarez Padón E. Herramientas de control y evaluación de proyectos para la toma de decisiones en el proceso administrativo. *Revista Académica: Contribuciones a la Economía*. 2019;
27. Martínez Alba I, Pérez Vergara I. LAS TÉCNICAS CUANTITATIVAS Y SU UTILIDAD EN EL DIAGNÓSTICO DEL SERVICIO DE ASISTENCIA TÉCNICA. *Dialnet*. 2010;31(2).
28. Bonet Borjas M. LEY DE PARETO APLICADA A LA FIABILIDAD. *Ingeniería Mecánica*. 2005;8(3):1–9.