



Cuadro sinóptico.

Nombre del Alumno: José Luis de la Cruz Villamil.

Nombre del tema: Estadística descriptiva.

Parcial: Único.

Nombre de la Materia: Bioestadística.

Nombre del profesor: Rosario Gómez Lujano.

Nombre de la Licenciatura: Lic. Enfermería.

Cuatrimestre: 4to cuatrimestre.

ESTADÍSTICA DESCRIPTIVA.

LA ESTADÍSTICA EN ENFERMERÍA.

Importancia.

Nos ayuda a conocer las problemáticas presentes en una comunidad, los factores de riesgo o predisposición a ciertas patologías y puede ser muy útil a la hora de buscar una respuesta a esta o al tratar de educar para evitarlas en futuras ocasiones.

Su principal ventaja es que...

Comprende una compleja metodología para dar respuesta a las hipótesis, además de agilizar la cuestión de organización del sistema de investigación, desde el diseño general, el de muestreo, el control de la calidad de información y la presentación de los resultados.

En salud pública...

La estadística permite analizar situaciones en las que los componentes aleatorios contribuyen de forma importante en la variabilidad de los datos obtenidos

INTRODUCCIÓN HISTÓRICA.

-El francés Pierre Charles-Alexandre Louis (1787-1872) fue el primer médico que utilizó métodos matemáticos para cuantificar variables de pacientes y sus enfermedades. La primera aplicación del Método numérico en su clásico estudio de la tuberculosis.

-En Francia Louis René Villerme (1782-1863) y en Inglaterra William Farr (1807 1883) que había estudiado estadística médica con Louis hicieron los primeros mapas epidemiológicos usando métodos cuantitativos y análisis epidemiológicos.

- Francis Galton (1822-1911), basado en el darwinismo social, fundó la biometría estadística.

- Pierre Simón Laplace (1749-1827), astrónomo y matemático francés, publicó en 1812 un tratado sobre la teoría analítica de las probabilidades, *Théorie analytique des probabilités*, sugiriendo que tal análisis podría ser una herramienta valiosa para resolver problemas médicos.

- Los primeros intentos de hacer coincidir las matemáticas de la teoría estadística con los conceptos emergentes de la infección bacteriana tuvieron lugar a comienzos del siglo XX.

- William Heaton Hamer (1862-1936) propuso un modelo temporal discreto en un intento de explicar la ocurrencia regular de las epidemias de sarampión.

- John Brownlee (1868 1927), primer director del British Research Council, luchó durante veinte años con problemas de cuantificación de la infectividad epidemiológica.

- Ronald Ross (1857-1932) exploró la aplicación matemática de la teoría de las probabilidades con la finalidad de determinar la relación entre el número de mosquitos y la incidencia de malaria en situaciones endémicas y epidémicas.

- Austin Bradford Hill (1897-1991) con el ensayo clínico aleatorizado y, en colaboración con Richard Doll (n. 1912), el épico trabajo que correlacionó el tabaco y el cáncer de pulmón.

- Los primeros trabajos bioestadísticos en enfermería los realizó, a mediados del siglo XIX la enfermera inglesa Florence Nightingale. Durante la guerra de Crimea, Florence Nightingale observó que eran mucho más numerosas las bajas producidas en el hospital que en el frente. Gracias a sus análisis estadísticos, se comenzó a tomar conciencia de la importancia y la necesidad de unas buenas condiciones higiénicas en los hospitales.

LA ESTADÍSTICA COMO HERRAMIENTA DE TRABAJO EN ENFERMERÍA.

El análisis y técnicas estadísticas son de gran relevancia debido a que en su objeto de estudio consideran...

En gran medida del tipo de estudio, del objetivo que se pretende abordar y del tamaño de la muestra, así como del grado de conocimiento por parte de los investigadores de las técnicas estadísticas y del software para su implementación.

Esta disciplina es usada en diversos campos de la medicina y la salud pública, como la epidemiología, nutrición y salud ambiental. Asimismo, sus métodos son aplicados en estudios relacionados con la ecología y la genómica.

Gracias a los datos arrojados por esta disciplina...

Se ha logrado un mejor entendimiento de la propagación de ciertas enfermedades y las características de males crónicos como el cáncer y el sida. Además, ha contribuido enormemente al desarrollo de nuevos fármacos.

ESTADÍSTICA DESCRIPTIVA.

DESCRIPCIÓN DE UNA VARIABLE ESTADÍSTICA.

Concepto.

Es una cualidad que generalmente adopta una forma numérica. Por ejemplo:

Juan mide 180 cm. La variable estadística sería la altura y está medida en centímetros.

Sin embargo, no todas las variables se pueden expresar en números, por ejemplo, el color de ojos de una persona.

Juan tiene ojos color verde. La variable sería el color de ojos y se trata de una variable cualitativa.

Existen dos tipos de variables.

1. Variable cuantitativa: Son variables que se expresan numéricamente.

-Variable continua: Toman un valor infinito de valores entre un intervalo de datos.

Por ejemplo, el tiempo que tarda un corredor en completar los 100 metros lisos.

-Variable discreta: Toman un valor finito de valores entre un intervalo de datos.

Ejemplo: Número de helados vendidos.

2. Variable cualitativa: Son variables que se expresan, por norma general, en palabras.

-Variable ordinal: Expresa diferentes niveles y orden. Por ejemplo, primero, segundo, tercero, etc.

-Variable nominal: Expresa un nombre claramente diferenciado. Por ejemplo, el color de ojos puede ser azul, negro, castaño, verde, etc.

REPRESENTACIONES GRÁFICAS.

Concepto.

Una gráfica o una representación gráfica o un gráfico, es un tipo de representación de datos, generalmente cuantitativos, mediante recursos visuales (líneas, vectores, superficies o símbolos), para que se manifieste visualmente la relación matemática o correlación estadística que guardan entre sí.

Los tipos de representaciones que existen son...

-Diagramas de barras: muestran los valores de las frecuencias absolutas sobre un sistema de ejes cartesianos, cuando la variable es discreta o cualitativa.

-Histogramas: formas especiales de diagramas de barras para distribuciones cuantitativas continuas.

-Polígonos de frecuencias: formados por líneas poligonales abiertas sobre un sistema de ejes cartesianos.

-Gráficos de sectores: circulares o de tarta, dividen un círculo en porciones proporcionales según el valor de las frecuencias relativas.

-Pictogramas: o representaciones visuales figurativas. En realidad, son diagramas de barras en los que las barras se sustituyen con dibujos alusivos a la variable.

-Cartogramas: expresiones gráficas a modo de mapa.

-Pirámides de población: para clasificaciones de grupos de población por sexo y edad.

REPRESENTACIÓN NUMÉRICA.

Concepto.

La tabla de frecuencias (o distribución de frecuencias) es una tabla que muestra la distribución de los datos mediante sus frecuencias. Se utiliza para variables cuantitativas o cualitativas ordinales. La tabla de frecuencias es una herramienta que permite ordenar los datos de manera que se presenten numéricamente las características de la distribución de un conjunto de datos o muestra.

Existen 4 tipos de frecuencia...

-Frecuencia absoluta. La frecuencia absoluta (n_i) de un valor X_i es el número de veces que el valor está en el conjunto (X_1, X_2, \dots, X_N) . La suma de las frecuencias absolutas de todos los elementos diferentes del conjunto debe ser el número total de sujetos N .

$$\sum_{i=1}^k n_i = n_1 + n_2 + \dots + n_k = N$$

-Frecuencia absoluta acumulada. La frecuencia absoluta acumulada (N_i) de un valor X_i del conjunto (X_1, X_2, \dots, X_N) es la suma de las frecuencias absolutas de los valores menores o iguales a X_i .

$$N_i = n_1 + n_2 + \dots + n_i$$

-Frecuencia relativa. La frecuencia relativa (f_i) de un valor X_i es la proporción de valores iguales a X_i en el conjunto de datos (X_1, X_2, \dots, X_N) .

$$f_i = \frac{n_i}{N}$$

siendo (X_1, X_2, \dots, X_N) el conjunto de datos y n_i el total de valores igual a X_i

-Frecuencia relativa acumulada. Definimos la frecuencia relativa acumulada (F_i) de un valor X_i como la proporción de valores iguales o menores a X_i en el conjunto de datos (X_1, X_2, \dots, X_N) .

$$F_i = \frac{N_i}{N}$$

ESTADÍSTICA DESCRIPTIVA.

CARACTERÍSTICAS DE POSICIÓN, DISPERSIÓN Y FORMA.

Medidas de posición.

Concepto.

Son indicadores estadísticos que permiten resumir los datos en uno solo, o dividir su distribución en intervalos del mismo tamaño. Se clasifican en medidas de posición central y no central.

No centrales.

Estos realizan una serie de divisiones iguales en la distribución ordenada de los datos. De esta forma, reflejan los valores superiores, medios e inferiores.

- El **cuartil**. Divide la distribución en cuatro partes iguales.
- El **quintil**. Divide la distribución en cinco partes.
- El **decil**. Estamos ante un cuartil que divide los datos en diez partes iguales.
- El **percentil**. Este cuartil divide la distribución en cien partes.

Centrales.

Estas nos permiten resumir la distribución de los datos en un solo valor central, alrededor del cual se sitúan; mientras que las segundas dividen la distribución en partes iguales.

- Media aritmética**. Son tres medidas centrales que nos indican un promedio ponderado de los datos.
- Mediana**. Divide la distribución en dos partes iguales. De esta forma, expresa el valor mediano, que no medio.
- Moda**. Estamos ante una medida central de los valores más frecuentes. Por tanto, la moda nos informa sobre aquellos que se repiten en más ocasiones.

Medidas de dispersión.

Concepto.

Expresan cómo se distribuyen los datos en torno a alguna de las medidas de centralización definidas antes, y son un complemento a estas últimas para describir más fielmente un conjunto de datos.

Varianza.

Es una medida de dispersión que se utiliza para representar la variabilidad de un conjunto de datos respecto de la media aritmética de los mismo. Así, se calcula como la suma de los residuos elevados al cuadrado y divididos entre el total de observaciones.

Desviación estándar.

Es una medida que ofrece información sobre la dispersión media de una variable. La desviación estándar es siempre mayor o igual que cero. Se obtiene al sacar la raíz cuadrada a la varianza.

Medidas de forma.

Concepto.

Son aquellas que nos muestran si una distribución de frecuencia tiene características especiales como simetría, asimetría, nivel de concentración de datos y nivel de apuntamiento que la clasifiquen en un tipo particular de distribución.

Coefficiente de asimetría de Fisher.

Una distribución es simétrica cuando al trazar una vertical, en el diagrama de barras o histograma de una variable, según sea esta discreta o continua, por el valor de la media, esta vertical se transforma en eje de simetría y entonces decimos que la distribución es simétrica. En caso contrario, dicha distribución será asimétrica o diremos que presenta asimetría. La asimetría puede ser de dos tipos: Asimétrica por la derecha y Asimétrica por la izquierda.

Coefficiente de curtosis o apuntamiento de Fisher.

La otra medida de forma que vamos a considerar es el apuntamiento, al igual que con la simetría hemos de tomar una referencia para ver si la distribución de los datos es apuntada o no. La referencia citada es la distribución normal, y distinguiremos tres casos:

- Leptocúrtica, si la distribución es más picuda que la normal.
- Mesocúrtica, si la distribución es igual a la normal.
- Platicúrtica, si la distribución es más aplastada que la normal.

ESTADÍSTICA DESCRIPTIVA.

Descripción numérica de una variable estadística bidimensional.

Concepto.

Es el conjunto de pares de valores de dos caracteres o variables estadísticas unidimensionales X e Y sobre una misma población. La variable estadística bidimensional se representa por el símbolo (X, Y) y cada uno de los individuos de la población viene caracterizado por la pareja (xi , yi), en el cual xi representa los datos, valores o marcas de clase x1, x2, ..., xn de la variable X; e yi representa los datos, valores o marcas de clase y1, y2, ..., ym de la variable Y.

Distribuciones bidimensionales.

Son las tablas estadísticas bidimensionales formadas por todas las frecuencias absolutas de todos los posibles valores de la variable estadística bidimensional (X, Y). Las tablas estadísticas bidimensionales pueden ser: Simples y de doble entrada

Distribuciones marginales y condicionadas.

Concepto.

Es la distribución de probabilidad de un subconjunto de variables aleatorias de un conjunto de variables aleatorias. La distribución marginal proporciona la probabilidad de un subconjunto de valores del conjunto sin necesidad de conocer los valores de las otras variables.

La variable marginal...

Se usa para referirse a una variable del subconjunto de retenido y cuyos valores pueden ser conocidos.

Se obtiene marginalizando sobre la distribución de variables descartadas y las variables descartadas se llaman a veces variables marginalizadas.

Se denomina distribución de la variable y condicionada...

A un valor dado xi de la variable X a la distribución unidimensional definida por el conjunto de valores tomados por Y y de las frecuencias de dichos valores de Y condicionadas a que X tome el valor xi.

INDEPENDENCIA E INCORRELACIÓN.

Concepto.

Dos variables estadísticas son estadísticamente independientes cuando el comportamiento estadístico de una de ellas no se ve afectado por los valores que toma la otra; esto es cuando las relativas de las distribuciones condicionadas no se ven afectadas por la condición, y coinciden en todos los casos con las frecuencias relativas marginales.

Se dice que dos variables X e Y son independientes estadísticamente cuando la frecuencia relativa conjunta es igual al producto de las frecuencias relativas marginales en todos los casos.

$$\frac{n_{ij}}{n} = \frac{n_{i.}}{n} \cdot \frac{n_{.j}}{n}$$

La incorrelación...

Es el grado de dispersión entre los puntos de una variable, es decir, el cuándo los puntos no marchan en una misma dirección si no que están dispersos por todos lados, a diferencia de la correlación que es todo lo contrario.

Características numéricas.

Los sistemas de numeración son conjuntos de dígitos usados para representar cantidades, así se tienen los sistemas de numeración decimal, binario, octal, hexadecimal, romano, etc. Los cuatro primeros se caracterizan por tener una base (número de dígitos diferentes: diez, dos, ocho, dieciséis respectivamente) mientras que el sistema romano no posee base y resulta más complicado su manejo tanto con números, así como en las operaciones básicas.

ESTADÍSTICA DESCRIPTIVA.

REGRESIÓN Y CORRELACIÓN.

Concepto.

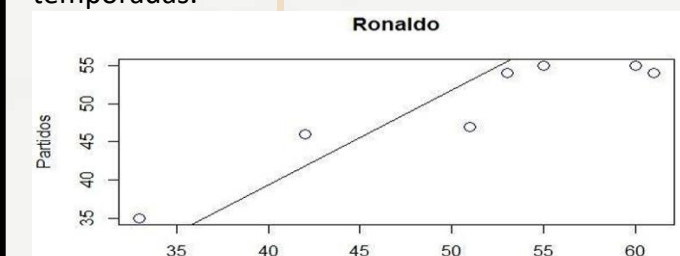
Comprende el análisis de los datos muestrales para saber qué es y cómo se relacionan entre si dos o más variables en una población. El análisis de correlación produce un número que resume el grado de la correlación entre dos variables; y el análisis de regresión da lugar a una ecuación matemática que describe dicha relación.

La curva de regresión de Y sobre X visualiza como cambia la media de la variable Y de aquellos grupos de observaciones caracterizados por tener un mismo valor en la otra variable X. Es decir, como varía, por término medio, la variable Y en función de los valores de X. Por eso la variable Y recibe el nombre de variable dependiente y la variable X el de variable independiente.

El coeficiente de determinación es la proporción de la varianza total de la variable explicada por la regresión. Es también denominado R cuadrado y sirve para reflejar la bondad del ajuste de un modelo a la variable que se pretende explicar. El coeficiente de determinación puede adquirir resultados que oscilan entre 0 y 1. Así, cuando adquiere resultados más cercanos a 1, mayor resultará el ajuste del modelo a la variable que se pretende aplicar para el caso en concreto. Por el contrario, cuando adquiere resultados que se acercan al valor 0, menor será el ajuste del modelo a la variable que se pretende aplicar y, justo por eso, resultará dicho modelo menos fiable.

Interpretación del coeficiente de determinación...

Supongamos que queremos explicar la cantidad de goles que anota Cristiano Ronaldo según la cantidad de partidos que juega. Suponemos que, a mayor cantidad de partidos jugados, más goles meterá. Los datos pertenecen a las últimas 8 temporadas.



REGRESIÓN MÚLTIPLE.

Concepto.

Este tipo se presenta cuando dos o más variables independientes influyen sobre una variable dependiente. Ejemplo: $Y = f(x, w, z)$.

Se puede ampliar para cualquier número "m" de variables independientes:

$$Y' = a' + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_mx_m$$

Para poder resolver y obtener a, b1 y b2 en una ecuación de regresión múltiple el cálculo se presenta muy tediosa porque se tiene atender 3 ecuaciones que se generan por el método de mínimo de cuadrados:

$$\sum y = na + b_1 \sum x_1 + b_2 \sum x_2$$

$$\sum x_1y = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1x_2$$

$$\sum x_2y = a \sum x_2 + b_1 \sum x_1x_2 + b_2 \sum x_2^2$$

El error estándar de regresión múltiple.

Es una medida de dispersión la estimación se hace más precisa conforme el grado de dispersión alrededor del plano de regresión se hace más pequeño. Para medirla se utiliza la fórmula:

$$S_{xy} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n - m - 1}}$$

El coeficiente de determinación múltiple.

Mide la tasa porcentual de los cambios de Y que pueden ser explicados por x1, x2 y x3 simultáneamente.

$$r^2 = \frac{SC_{regresión}}{SC_{Total}}$$

ANÁLISIS DE ATRIBUTOS.

Su principal objetivo es el de evitar un error muy común consistente en tratar de encontrar la forma de mejorar un producto, servicio o proceso analizándolo como un todo.

Características para las Gráficas de Control de Atributos.

- Están basadas en decisiones de pasa/no pasa.
- Se pueden aplicar en casi cualquier operación donde se recolectan datos.
- Se utilizan en características de calidad que no pueden ser medidas o que son costosas o difíciles de medir. A diferencia de las gráficas de control de datos variables, las gráficas de datos atributos se puede establecer para una

Tipos de Gráficas de Atributos:

- Defectivos
 - (np) número de unidades no conformes
 - (p) proporción de unidades no conformes
- Defectos
 - (c) número de defectos
 - (u) proporción de defectos

1.- Obtener: media aritmética, mediana, moda, rango, varianza y desviación estándar de los siguientes datos no agrupados: 6, 7, 8, 9, 10, 8

-Media aritmética: $6+7+8+9+10+8= 48/6 = 8$

-Mediana: 6, 7, 8, 8, 9, 10= $8+8/2 = 8$

-Moda: 8

-Rango: 6, 7, 8, 8, 9, 10= $10-6 = 4$

-Varianza: $(6-8)^2 + (7-8)^2 + (8-8)^2 + (8-8)^2 + (9-8)^2 + (10-8)^2 / 6-1$

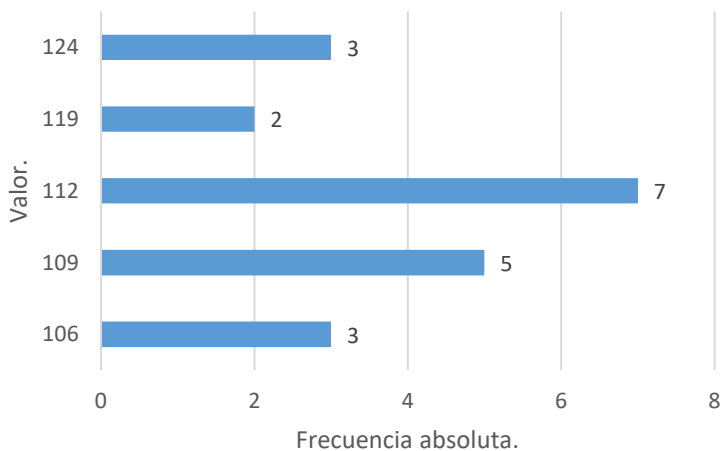
$4+1+0+0+1+4/5= 10/5= 2$

-Desviación estándar: $\sqrt{2}= 1.41421$

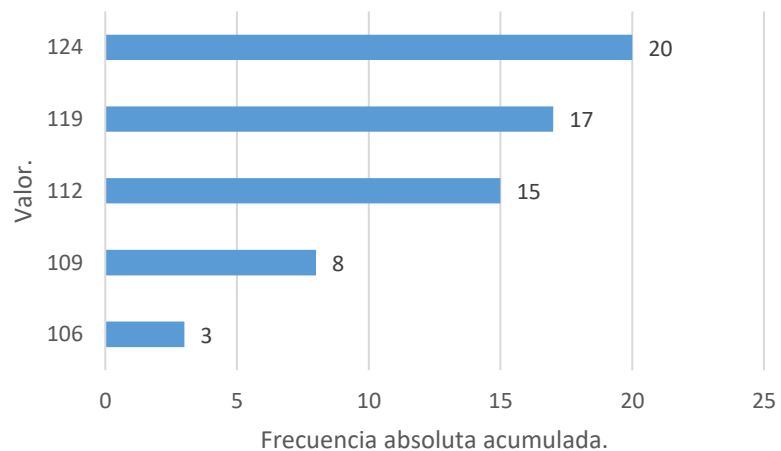
2.- Cierta universidad realizó un experimento sobre el coeficiente intelectual (C.I.) de sus alumnos, para lo cual aplicó un examen de C.I. a un grupo de 20 alumnos escogidos al azar, obteniendo los siguientes resultados: 119, 109, 124, 119, 106, 112, 112, 112, 112, 109, 112, 124, 109, 109, 109, 106, 124, 112, 112, 106.

Valor.	Frecuencia absoluta.	Frecuencia absoluta acumulada.	Frecuencia relativa.	Frecuencia relativa acumulada.
106	3	3	0.15	0.15
109	5	8	0.25	0.4
112	7	15	0.35	0.75
119	2	17	0.1	0.85
124	3	20	0.15	1

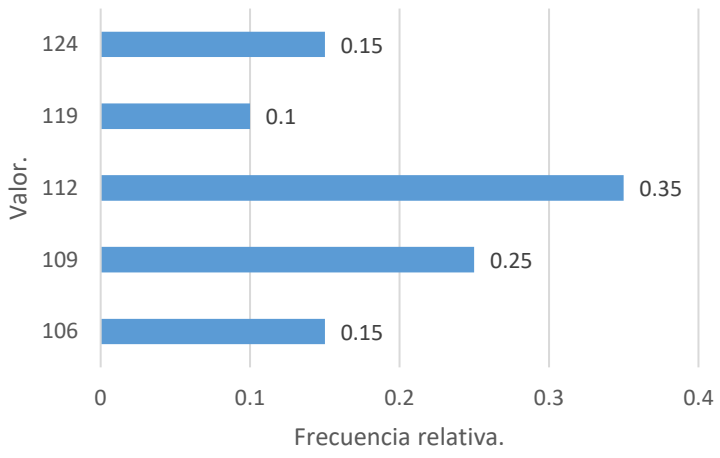
Frecuencia absoluta de valores.



Frecuencia absoluta acumulada.



Frecuencia relativa.



Frecuencia relativa acumulada.

