



Mi Universidad

Cuadro Sinóptico

Nombre del Alumno: Jesus Alexander Gómez Morales

Nombre del tema: Estadística Descriptiva

Parcial: I

Nombre de la Materia: Bioestadística

Nombre del profesor: Andrés Alejandro Reyes Molina

Nombre de la Licenciatura: Licenciatura en enfermería Grupo B

Cuatrimestre: Cuarto Cuatrimestre

UNIDAD 1: ESTADÍSTICA DESCRIPTIVA

1.8 Distribuciones marginales y condicionadas.

En teoría de probabilidades, la distribución marginal 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

El término variable marginal

se usa para referirse a una variable del subconjunto de retenido y cuyos valores pueden ser conocidos. La distribución de las variables marginales, la distribución marginal, se obtiene marginalizando sobre la distribución de variables descartadas y las variables descartadas se llaman a veces variables marginalizadas.

Partiendo de una distribución bidimensional de frecuencias (xi, yj; nij), se denomina distribución condicionada de la variable X a un valor dado yj de la variable Y a la distribución unidimensional definida por el conjunto de valores tomados por X y de las frecuencias condicionadas de dichos valores de X a qué Y tome el valor yj.

La función de probabilidad marginal es usada para hallar las diferentes distribuciones de probabilidades estadísticas de las variables individuales, con esta función podemos asignar diferentes valores a las variables conjuntas sin tener que relacionarlas

La distribución marginal de X

X es simplemente la función de probabilidad de x, pero la palabra marginal sirve para distinguirla de la distribución conjunta de X e Y.
La distribución marginal de dos variables aleatorias se puede obtener a partir de su distribución conjunta

1.9 Independencia e incorrelación

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

es decir

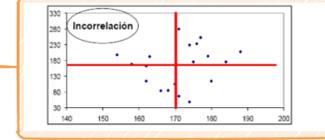
$$\frac{n_{ij}}{n} = \frac{n_{i.}}{n} \cdot \frac{n_{.j}}{n} \text{ Para todo } i, j$$

Ejemplo

el suceso estatura de los alumnos de una clase y el color del pelo son independientes: el que un alumno sea más o menos alto no va a influir en el color de su cabello, ni viceversa. Para que dos sucesos sean independientes tienen que verificar al menos una de las siguientes condiciones

$P(B|A) = P(B)$ es decir, que la probabilidad de que se dé el suceso B, condicionada a que previamente se haya dado el suceso A, es exactamente igual a la probabilidad de B.

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.

1.10 Regresión y correlación

El análisis de correlación generalmente resulta útil para un trabajo de exploración cuando un investigador o analista trata de determinar que variables son potenciales importantes, el interés radica básicamente en la fuerza de la relación. La correlación mide la fuerza de una entre variables; la regresión da lugar a una ecuación que describe dicha relación en términos matemáticos

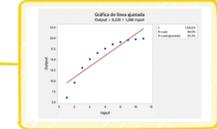
Definiciones

En estadística, el análisis de la regresión es un proceso estadístico para estimar las relaciones entre variables. Incluye muchas técnicas para el modelado y análisis de diversas variables, cuando la atención se centra en la relación entre una variable dependiente y una o más variables independientes (o predictoras)

el análisis de regresión estima la esperanza condicional de la variable dependiente dadas las variables independientes - es decir, el valor promedio de la variable dependiente cuando se fijan las variables independientes. Con menor frecuencia, la atención se centra en un cuantil, u otro parámetro de localización de la distribución condicional de la variable dependiente dadas las variables independientes.

Curva de regresión y coeficiente de determinación

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.



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 es la proporción de la varianza total de la variable explicada por la regresión. El coeficiente de determinación, también llamado R cuadrado, refleja la bondad del ajuste de un modelo a la variable que pretende explicar

Tema principal 1.1

1.11 Otros tipos de regresión

Regresión Múltiple: Este tipo se presenta cuando dos o más variables independientes influyen sobre una variable dependiente. Ejemplo: $Y = f(x, w, z)$.
Análisis de Regresión Múltiple Dispone de una ecuación con dos variables independientes adicionales:

$$Y' = a + b_1x_1 + b_2x_2$$

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

$$\begin{aligned} \sum y &= na + b_1 \sum x_1 + b_2 \sum x_2 \\ \sum x_1 y &= a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 \\ \sum x_2 y &= a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 \end{aligned}$$

El error estándar de la 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

$$S_e = \sqrt{\frac{\sum (Y - Y')^2}{n - m - 1}}$$

Y' = Valores observados en la muestra
Y = Valores estimados a partir de la ecuación de regresión
n = Número de datos
m = Número de variables independientes

El coeficiente de determinación múltiple

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

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

1.12 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. Muchas veces, la búsqueda de una idea global, salvadora, que mejore el todo, impide descubrir la característica específica que, por sí sola, podría producir el resultado deseado.

- 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 pueden establecer para una característica de calidad o para muchas

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