Ricardo Michel MALLQUI BAÑOS

Estadística

PRIMVN VIVERE DE INDE PHILOSOPHARI

universitas + guamangensis + sancti + christophori

+ 1677 +

Índice general

Índice de cuadros

Índice de figuras

Resumen

Este libro sobre la estadistica descriptiva. cuyo objetivo es demostrar resultados basicos muy útiles en el desarrollo de investigaciones.

$$\sum_{1}^{2}$$

Introducción

$$\sum_{1}^{2}$$

$$\vec{u} = (1,1) - \rho \int_{2}^{3}$$

Debido a la poca información estructurada de estadistica descriptiva se propone escribir este libro con un enfoque demostrativo.

$$x^2 + y^2$$

1

Variables

1.1. Variables cualitativas

Denotan cualidades de objetos personas o animales tales como caracteristicas inherentes que no son medibles por números, tenemos dos casos de esta variable.

1.1.1. Nominales

Son caracteristicas que simplemente nominan y estan propensos a ser jerarquizados u ordenados tales como: El estado civil (soltero, casado, divorciado, viudo), Religion (católic, evangelico, judio, etc).

1.1.2. Ordinales

Son caracteristicas que que si están propensos a ser jerarquizados tales como: Nivel de instrucción (primaria, secundaria, superior).

1.2. Variables cuantitativas

Son aqueellas variables que están propensos a ser medidas mediante números ya sean números enteros o reales.

1.2.1. Discretas

Aquellas que solo son medidos mediante numeros enteros por ejemplo: Número de hijos, número de habitaciones.

1.2.2. Continuas

Aquellas que solo son medidos mediante numeros reales es decir este incluye a los numeros racionales e irracionales. Estatura, volumen, peso.

2

Tabulación

La tabulación es un proceso en el cual los datos son ordenados en grupos llamados *clases* para un análisis más eficaz de estos, los datos podrían estar clasificados mediante una variable cualitativa o cuantitativa en el caso de las variables cualitativas Y_i, se considera la siguiente Tabla 2.1

Cuadro 2.1: Caption

Y_i	f_i	F_i	F_i^*	h_i	H_i	H_i^*	h_i %	H_i %	H_i^* %
Y_1	f_1	F_1	F_1^*	$\frac{f_1}{n}$	$\frac{F_1}{n}$	$\frac{F_1^*}{n}$	h_1	H_1	H_1^*
Y_2	f_2	F_2	F_2^*	$\frac{f_2}{n}$	$\frac{F_2}{n}$	$\frac{F_2^*}{n}$	h_2	H_2	H_1^*
Y_3	f_3	F_3	F_3^*	$\frac{f_3}{n}$	$\frac{F_3}{n}$	$\frac{F_3^*}{n}$	h_3	H_3	H_1^*
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Y_r	f_r	F_r	F_r^*	$\frac{f_r}{n}$	$\frac{F_r}{n}$	$\frac{F_r^*}{n}$	h_r	H_r	H_1^*

En el caso de variables cuantitativas ademas si los datos son muy variados, que para se clasificados adecuadamente, necesitan generarse particiones de longitudes semejantes entonces se utiliza el siguiente proceso; el **número de las particiones** r se consideran de acuerdo a **tres criterios**

- Criterio del investigador r no puede ser más de 20 ni menos de 5
- $r = \sqrt{n}$ donde n es el número de datos
- La regla de Starges que consiste en considerar la fórmula $r = 3{,}322 \cdot \log_{10} n$ Una vez establecido el número de particiones se procede a generar los límites laterales de cada una de las particiones, sea L la longitud de todo el conjunto es decir $L = x_{\max} - x_{\min}$ entonces la longitud de las particiones o amplitud interválica se obtiene con $l = \frac{L}{r}$

Clase	Clase	f_i	F_i	F_i^*	h_i	H_i	H_i^*	$\ldots$
$[y_1 - y_2 >$	y_1	f_1	F_1	F_1^*	$\frac{f_1}{n}$	$\frac{F_1}{n}$	$\frac{F_1^*}{n}$	$\ldots$
$< y_1 - y_2 >$	y_2	f_2	F_2	F_2^*	$\frac{f_2}{n}$	$\frac{F_2}{n}$	$\frac{F_2^*}{n}$	$\ldots$

Clase	Clase	f_i	F_i	F_i^*	h_i	H_i	H_i^*	...
$< y_r - y_r >$	y_3	f_3	F_3	F_3^*	$\frac{f_3}{n}$	$\frac{F_3}{n}$	$\frac{F_3^*}{n}$	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$< y_{r-1} - y_r]$	y_r	f_r	F_r	F_r^*	$\frac{f_r}{n}$	$\frac{F_r}{n}$	$\frac{F_r^*}{n}$	...

Tenga en cuenta que n es el número de datos, es decir $n = f_1 + f_2 + \ldots + f_r = \sum_{i=1}^{r}$ donde f_i es número de datos en la partición X_i, una de las r particiones del conjunto total de datos.

- Las **frecuencias absolutas** f_i indican el número de datos con la característica X_i.
- Las **frecuencias absolutas acumuladas menor que** F_i obedecen a la fórmula
$$F_m = f_1 + f_2 + \ldots + f_m = \sum_{i=1}^{m} f_i$$
- Las **frecuencias absolutas acumuladas mayor que** F_i^* obedecen a la fórmula
$$F_m^* = f_m + f_{m+1} + \ldots + f_r = \sum_{i=m}^{r} f_i = n - \sum_{i=1}^{m-1} f_i = n - (f_1 + f_2 + \ldots + f_{m-1})$$
- Las **frecuencias absolutas relativas** obedecen a la fórmula
$$h_m = \frac{f_m}{n}$$
- Las **frecuencias absolutas relativas menor que** obedecen a la fórmula
$$H_m = \frac{f_m}{n}$$
- Las **frecuencias absolutas relativas mayor que** obedecen a la fórmula
$$H_m^* = \frac{F_m}{n}$$
- Las **frecuencias absolutas relativas porcentuales** obedecen a la fórmula $h_i \,\% = 100 \cdot h_i$
- Las **frecuencias absolutas relativas menor que porcentuales** obedecen a la fórmula $H_i \,\% = 100 \cdot H_i$
- Las **frecuencias absolutas relativas mayor que porcentuales** obedecen a la fórmula $H_i^* \,\% = 100 \cdot H_i^*$

Ejercicio 2.1. Sean los datos

Solución. Entonces

3

Medidas de tendencia central

Son aquellas medidas que buscan un dato representtivo central de un conjunto de datos tales como la media, la moda y la mediana.

3.1. La media ($\overline{x}$)

A veces llamada *promedio aritmético*, es la medida de tendencia central que pondera los datos.

3.1.1. Media de datos no agrupados

Los datos no están agrupados cuando no están ordenados sobre una tabla de distribución de frecuencias. Sean los n datos $x_1, x_2, \dots, x_n$ entonces la media o promedio aritmético se define como

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.1}$$

$$\frac{d\,[P; F_1]}{d\,[P; \mathcal{L}_1]} = e = \frac{d\,[P; F_2]}{d\,[P; \mathcal{L}_2]} \tag{3.2}$$

1. $\overline{x} = \frac{x_1+x_2+\cdots+x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. $\overline{x} = \frac{x_1+x_2+\cdots+x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$

3.1.2. Media de datos agrupados

Considérese la siguiente tabla de distribucion de frecuencias entonces el promedio es

$$\overline{x} = \frac{y_1 f_1 + y_2 f_2 + \cdots + y_n f_n}{n} = \frac{1}{n}\sum_{i=1}^{n} y_i f_i$$

Clase	Clase	f_i	F_i	F_i^*	h_i	H_i	H_i^*	$\ldots$
$< y_1 - y_2 >$	y_1	f_1	F_1	F_1^*	$\frac{f_1}{n}$	$\frac{F_1}{n}$	$\frac{F_1^*}{n}$	$\ldots$
$< y_2 - y_3 >$	y_2	f_2	F_2	F_2^*	$\frac{f_2}{n}$	$\frac{F_2}{n}$	$\frac{F_2^*}{n}$	$\ldots$
$< y_3 - y_4 >$	y_3	f_3	F_3	F_3^*	$\frac{f_3}{n}$	$\frac{F_3}{n}$	$\frac{F_3^*}{n}$	$\ldots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$< y_{r-1} - y_r]$	y_r	f_r	F_r	F_r^*	$\frac{f_r}{n}$	$\frac{F_r}{n}$	$\frac{F_r^*}{n}$	...

Ejercicio 3.1. Si el promedio de n datos es $\overline{x}$ entonces el promedio del conjunto inicial más un dato adicional x_{n+1} es

$$\overline{x}' = \frac{n\overline{x} + x_{n+1}}{n+1}$$

en general si se adicionan r datos $y_1, y_2, \ldots y_r$ entonces el nuevo promedio será

$$\overline{x}' = \frac{n\overline{x} + y_1 + y_2 + \ldots + y_r}{n+r}$$

Solución. En efecto sea el promedio

$$\begin{aligned}\overline{x}' &= \frac{x_1 + x_2 + \cdots + x_{n+1}}{n+1}\\ &= \frac{n\frac{x_1+x_2+\cdots x_n}{n} + x_{n+1}}{n+1}\\ &= \frac{n\overline{x} + x_{n+1}}{n+1}\end{aligned}$$

3.2. La moda (Mo)

3.2.1. Moda de datos no tabulados

En este caso es dato que más repite en un conjunto de datos dados.

La moda es el dato que más se repite por ejemplo sea el conjunto de datos x_1, x_2, x_2, x_2, x_3 entonces la moda Mo $= x_2$

3.2.2. Moda de datos tabulados

La moda es el dato que más se repite por ejemplo sea el conjunto de datos x_1, x_2, x_2, x_2, x_3 entonces la moda Mo $= Li + \frac{Li-Ls}{Li+Ls}r$

Clase	Clase	f_i	F_i	F_i^*	h_i	H_i	H_i^*	...
$[y_1 - y_2 >$	y_1	f_1	F_1	F_1^*	$\frac{f_1}{n}$	$\frac{F_1}{n}$	$\frac{F_1^*}{n}$	...
$< y_1 - y_2 >$	y_2	f_2	F_2	F_2^*	$\frac{f_2}{n}$	$\frac{F_2}{n}$	$\frac{F_2^*}{n}$	...
$< y_r - y_r >$	y_3	f_3	F_3	F_3^*	$\frac{f_3}{n}$	$\frac{F_3}{n}$	$\frac{F_3^*}{n}$	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$< y_{r-1} - y_r]$	y_r	f_r	F_r	F_r^*	$\frac{f_r}{n}$	$\frac{F_r}{n}$	$\frac{F_r^*}{n}$	...

3.3. la mediana (Me)

3.3.1. Mediana de datos no tabulados

Obtener la mediana consiste en ordenar los datos de menor a mayor y considerar dos casos: El prmero si el numero de datos s impar entonces el dato $x_{\frac{n+1}{2}}$ del conjunto ordenado será la mediana es decir Me $= x_{\frac{n+1}{2}}$ de otro lado si el número de datos es par entonces la mediana es la semisuma de los dos datos intermedios es decir Me $= \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$

Ejercicio 3.2. Sean los conjuntos de datos 5, 6, 8, 2, 1, 5, 6, 7, 10, 0, 14 y 20, 25, 6, 5, 19, 5 obtener la mediana de estos conjuntos de datos.

Solución. Al ordenarlos se obtiene el siguiente arreglo 0, 1, 2, 5, 5, 6, 6, 7, 8, 10, 14 y considerando que $x_1 = 0, x_2 = 1, \ldots, x_{11} = 14$ en este caso el número de datos es impar entonces el dato $x_{\frac{11+1}{2}} = x_6 = 6$ el la mediana. De otro lado el segundo conjunto de datos al ser ordenados 5, 5, 6, 19, 20, 25 ademas considerando que $x_1 = 5, x_2 = 5, \ldots, x_6 = 25$ conducen a obtener la mediana Me $= \frac{x_{\frac{6}{2}} + x_{\frac{6}{2}+1}}{2} = \frac{6+19}{2} = 12{,}5$.

3.3.2. Mediana de datos tabulados

Clase	Clase	f_i	F_i	F_i^*	h_i	H_i	H_i^*	...
$[y_1 - y_2 >$	y_1	f_1	F_1	F_1^*	$\frac{f_1}{n}$	$\frac{F_1}{n}$	$\frac{F_1^*}{n}$	...
$< y_1 - y_2 >$	y_2	f_2	F_2	F_2^*	$\frac{f_2}{n}$	$\frac{F_2}{n}$	$\frac{F_2^*}{n}$	...
$< y_r - y_r >$	y_3	f_3	F_3	F_3^*	$\frac{f_3}{n}$	$\frac{F_3}{n}$	$\frac{F_3^*}{n}$	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Clase	Clase	f_i	F_i	F_i^*	h_i	H_i	H_i^*	...
$< y_{r-1} - y_r]$	y_r	f_r	F_r	F_r^*	$\frac{f_r}{n}$	$\frac{F_r}{n}$	$\frac{F_r^*}{n}$	...

Los pasos son:

- Se halla $\frac{n}{2}$ luego
- x_n
-

4

traductor

5

traducto2

6

estadistica

Teorema 6.1. *En la elipse se verifican las siguientes igualdades*

1. $d\,[B_1; F_i] = d\,[B_2; F_i] = a$
2. $d\,[V_1; C] = d\,[V_2; C] = a$
3. $d\,[C; \mathcal{L}_1] = d\,[C; \mathcal{L}_2] = \frac{c}{e}$
4. $c = d\,[P; F_1] = d\,[P; F_2]$ *entonces* $c = ae$

Demostración. 1. Ya que $d\,[B_1; F_1] + d\,[B_1; F_2] = 2a = d\,[B_2; F_1] + d\,[B_2; F_2]$ es decir $2d\,[B_1; F_i] = 2a = 2d\,[B_2; F_i]$ entonces $d\,[B_1; F_i] = a = d\,[B_2; F_i]$ $i = 1, 2$.

2. Por la definición (**??**) de la elipse se tiene

$$d\,[V_1; F_2] + d\,[V_1; F_1] = 2a \tag{6.1}$$

además la diferencia

$$d\,[V_1; F_2] - d\,[V_1; F_1] = 2c \tag{6.2}$$

restando las ecuaciones (6.1) y (6.2) se tiene

$$d\,[V_1; F_1] = a - c \tag{6.3}$$

entonces haciendo uso de (6.3) en $d\,[V_1; C] = d\,[V_1; F_1] + d\,[F_1; C] = (a - c) + c = a$; de manera similar para el vértice V_2.

3. En efecto

$$\frac{d\,[B; F_i]}{d\,[B; \mathcal{L}_i]} = e \Longleftrightarrow \frac{a}{d\,[B; \mathcal{L}_i]} = e$$

además $d\,[B_i; \mathcal{L}_i] = d\,[C; \mathcal{L}_i]$ por lo tanto $\frac{a}{d[C;\mathcal{L}_i]} = e$.

4. Pues

$$\frac{d\,[P; F_1]}{d\,[P; \mathcal{L}_1]} = e$$

implica $\frac{a-c}{\frac{a}{e}-a} = e$ es decir $c = ae$.

Por lo tanto □

7

eee

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

can be found on the Pandoc website http://pandoc.org.

$$\sum$$

> I thoroughly disapprove of duels. If a man should challenge me, I would take him kindly and forgivingly by the hand and lead him to a quiet place and kill him.

In this section, we give a very brief introduction to Pandoc's Markdown. Readers who are familiar with Markdown can skip this section. The comprehensive syntax of Pandoc's Markdown can be found on the Pandoc website http://pandoc.org. $\sum_1^2$

> I thoroughly disapprove of duels. If a man should challenge me, I would take him kindly and forgivingly by the hand and lead him to a quiet place and kill him.
>
> – Mark Twain

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} - \frac{2}{3} \begin{pmatrix} \alpha_1 & \beta_2 \\ \gamma & \delta \end{pmatrix}$$

* La suma de dos matrices $A_{n \times m}$ y $B_{r \times s}$

$$A_{n \times m} \pm B_{n \times m} = [a_{ij} + b_{ij}]$$

* El producto de dos matrices $A_{n \times m}$ y $B_{r \times s}$

$$A_{n \times m} \cdot B_{n \times m} = [a_{ij} + b_{ij}]$$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix}$$

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$\begin{matrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{matrix}$$

A

Sumatorias

Una suma de números representados por $x_1, x_2, \ldots, x_n$ se simboliza en forma compacta mediante el simbolo $\sum$ (sigma) es decir la suma de los números anteriores se puede escribir del siguiente modo

$$x_1 + x_2 + \cdots + x_n = \sum_{i=1}^{n} x_i .$$

Algunas propiedades son

1. $k \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} k x_i$
2. $\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$
3. $\sum_{i=1}^{n} x_i$

   $$\int_1^3 = \lim_{n \to \infty} \sum_{i=0}^{n} f^i(x)$$

   citado por (Xie, 2015)

A.1. ee

A.2. eeeee

B

Matrices

Una matriz es un arreglo de números distribuidos en filas y columnas por ejemplo la siguiente matriz

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{11} & a_{11} & \dots & a_{nm} \end{pmatrix}_{n\times n}$$

de **orden** $n\times m$ tiene **entradas** a_{ij} donde el primer subindice indica la fila y el segundo la columna; es usual representar por simplicidad una matriz por $A = [a_{ij}]_{n\times m}$. Si en el orden $n = m$ entonces la matriz recibe el nombre de **matriz cuadrada** la suma de los elementos de la diagonal de una matriz cuadrada $\sum_{i=1}^{n} a_{ii}$ se llama **traza**. Si todas las a_{ij} son cero entonces la matriz $A = 0$ recibe el nombre matriz **nula**.

Dos matrices son iguales si tienen el **mismo orden** y cada una de las entradas respectivas son iguales es decir $A = [a_{ij}]_{n\times m}$ y $B = [b_{ij}]_{n\times m}$ son iguales si $a_{ij} = b_{ij}$, $i = 1, 2, \dots n$ y $j = 1, 2, \dots m$

B.1. Algebra de matrices

Sean las matrices $A = [a_{ij}]_{n\times m}$ y $B = [b_{ij}]_{p\times q}$ entonces la suma y producto de matrices se definen

1. Sea k un escalar entonces se verifica que $kA = [ka_{ij}]$, $i = 1, 2, \dots n$ y $j = 1, 2, \dots m$ es decir el escalar k multiplica a cada una de las entradas de la matriz.
2. La suma o diferencia es posible si $n = p$ y $m = q$ es decir los ordenes de A y B son iguales, entonces la suma o diferencia resulta $A \pm B = [a_{ij} + b_{ij}]_{n\times m}$, $i = 1, 2, \dots n$ y $j = 1, 2, \dots m$
3. El producto es posible si $m = p$ es decir el número columnas de la primera matriz es igual al número de filas de la segunda matriz, el orden de la

matriz resultante es $n \times q$ además

$$
\begin{aligned}
A \cdot B &= \left[\sum_{k=1}^{p} a_{ik} b_{kj}\right]_{n\times q} \\
&= \begin{pmatrix}
\sum_{k=1}^{m} a_{1k} b_{k1} & \sum_{k=1}^{m} a_{1k} b_{k2} & \dots & \sum_{k=1}^{m} a_{1k} b_{kq} \\
\sum_{k=1}^{m} a_{2k} b_{k1} & \sum_{k=1}^{m} a_{2k} b_{k2} & \dots & \sum_{k=1}^{m} a_{2k} b_{kq} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{k=1}^{m} a_{nk} b_{k1} & \sum_{k=1}^{m} a_{nk} b_{k2} & \dots & \sum_{k=1}^{m} a_{nk} b_{kq}
\end{pmatrix}_{n\times q}
\end{aligned}
$$

donde $i = 1, 2, \dots n$ y $j = 1, 2, \dots m$

Ejemplo B.1. Sean $\begin{pmatrix} 3 & -1 & 2 \\ 2 & -1 & 2 \\ 1 & -1 & 0 \\ 5 & 0 & 0 \end{pmatrix}_{4\times 3}$ y $\begin{pmatrix} 0 & -1 & 2 & 2 & 0 \\ 1 & -1 & -2 & 1 & 1 \\ 3 & -1 & -3 & 5 & 2 \end{pmatrix}_{3\times 5}$ entonces $A \cdot B =$ $\begin{pmatrix} 5 & -4 & 2 & 15 & 3 \\ 5 & -3 & 0 & 13 & 3 \\ -1 & 0 & 4 & 1 & -1 \\ 0 & -5 & 10 & 10 & 0 \end{pmatrix}_{4\times 5}$

En caso de ser posible la multiplicación entre A, B y C entonces se verfican las siguientes propiedades

- $A(B + C) = AB + AC$
- $(A + B)C$
- $A(BC) = (AB)C$

Bibliografía

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Índice alfabético