

Ricardo Michel MALLQUI BAÑOS

Elementos de la estadística



Índice general

Índice de tablas	v
Índice de figuras	vii
Resumen	ix
Introducción	xi
I Estadística descriptiva	1
1. Prerrequisitos	3
2. Variables	5
2.0.1. Nominales	5
2.0.2. Ordinales	5
2.1. Variables cuantitativas	5
2.1.1. Discretas	5
2.1.2. Continuas	6
3. Característica de los datos	7
4. Población, muestra y estadística	9
5. Organización de datos en tablas de frecuencias	11
6. Matriz y rango	17
7. Distribución de frecuencias	19
8. Gráficos estadísticos	21
9. Medidas de tendencia central	23
9.1. La media (\bar{x})	23
9.1.1. Media de datos no agrupados	23
9.1.2. Media de datos agrupados	23
9.2. La moda (M_o)	24
9.2.1. Moda de datos no tabulados	24
9.2.2. Moda de datos tabulados	24

9.3. la mediana (Me)	25
9.3.1. Mediana de datos no tabulados	25
9.3.2. Mediana de datos tabulados	25
10. Medidas de dispersión	27
11. Medidas de asimetría	29
II Probabilidades	31
1. Frecuencia relativa clásica	33
2. Calculando	35
3. Variable aleatoria	37
4. Variable acumulativa	39
5. ww	41
Apéndice	41
A. Sumatorias	43
A.1. ee	43
A.2. eeee	43
B. Matrices	45
B.1. Algebra de matrices	45
Bibliografía	47
Índice alfabético	49

Índice de tablas

5.1. Here is a nice table!	11
5.2. Here!	12
5.3. Figures and tables with captions will be placed in ‘figure’	13
7.1. Caption	19



--

--



Resumen

Este libro sobre la estadística descriptiva, cuyo objetivo es demostrar resultados básicos 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 estadística descriptiva se propone escribir este libro con un enfoque demostrativo.

$$x^2 + y^2$$



Parte I

Estadística descriptiva



1

Prerrequisitos



2

Variables

Es una característica de personas cosas u objetos que son propensos a ser medidas ##
Variables cualitativas

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

2.0.1. Nominales

Son características que simplemente nominan y están propensos a ser jerarquizados u ordenados tales como: El estado civil (soltero, casado, divorciado, viudo), Religión (católico, evangélico, judío, etc).

2.0.2. Ordinales

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

2.1. Variables cuantitativas

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

2.1.1. Discretas

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

2.1.2. Continuas

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

3

Característica de los datos



4

Población, muestra y estadística



5

Organización de datos en tablas de frecuencias

Sea la Tabla 5.1 Figures and tables with captions will be placed in figure and table environments, respectively. Figures and tables with captions will be placed in figure and table environments, respectively. Figures and tables with captions will be placed in figure and table environments, respectively.

Tabla 5.1: Here is a nice table!

País	z	Porcentaje
Dinamarca	8.0	8.00\ %
Islandia	7.7	7.70\ %
Corea del Sur	7.6	7.60\ %
Noruega	7.6	7.60\ %
Israel	7.4	7.40\ %
Nueva Zelanda	7.3	7.30\ %
Estados Unidos	7.3	7.30\ %
Bélgica	6.6	6.60\ %
Canada	6.6	6.60\ %
Finlandia	6.5	6.50\ %
Suecia	6.5	6.50\ %
Reino Unido	6.5	6.50\ %
Chile	6.4	6.40\ %
Irlanda	6.4	6.40\ %
Francia	6.3	6.30\ %
Holanda	6.3	6.30\ %
Mexico	6.2	6.20\ %
Australia	6.1	6.10\ %
Estonia	6.0	6.00\ %
Eslovenia	5.9	5.90\ %
Austria	5.8	5.80\ %
Polonia	5.8	5.80\ %
Portugal	5.8	5.80\ %
España	5.6	5.60\ %
Suiza	5.6	5.60\ %

Japón	5.1	5.10\%
Alemania	5.0	5.00\%
República Checa	4.7	4.70\%
Italia	4.7	4.70\%
Hungría	4.6	4.60\%
Eslovaquia	4.6	4.60\%
Grecia	4.3	4.30\%

Tabla 5.2: Here!

ID	Motivador	X3	\bar{x}	Formativa	X6	X7
	IT1	IT2	O1	IT3	IT4	IT5
1	2	2	2	2	1	1
2	3	2	3	3	3	3
3	3	3	3	3	2	2
4	3	3	3	2	2	2
5	3	2	3	3	3	3
6	1	1	1	1	2	2
7	2	3	3	3	3	2
8	2	2	2	1	1	1
9	1	2	2	1	1	1
10	1	2	2	2	1	1
11	2	2	2	2	2	2
12	2	3	3	3	2	2
13	3	2	3	3	2	2
14	2	3	3	2	2	3
15	2	2	2	2	2	1
16	2	2	2	3	2	3
17	2	2	2	2	2	2
18	1	2	2	1	1	2
19	3	2	3	3	3	3
20	3	3	3	3	2	2
21	1	1	1	1	2	2
22	3	3	3	3	3	2
23	3	2	3	3	3	3
24	3	2	3	3	3	3
25	2	3	3	3	3	2
26	2	2	2	1	1	1
27	1	2	2	1	2	2
28	3	2	3	3	2	2

29	3	2	3	3	3	3
30	1	2	2	2	1	1
31	3	3	3	3	3	2
32	3	3	3	2	3	3
33	1	1	1	1	1	1
34	1	1	1	1	1	1
35	3	2	3	2	2	2

$$\sum_{i=1}^n x_i$$

Tabla 5.3: Figures and tables with captions will be placed in ‘figure’

	X1	α	$\sum_{i=1}^n x_i$	X4	X5	X6	X7
1	ID	Motivador		\bar{x}	Formativa		
2		IT1	IT2	O1	IT3	IT4	IT5
3	1	2	2	2	2	1	1
4	2	3	2	3	3	3	3
5	3	3	3	3	3	2	2
6	4	3	3	3	2	2	2
β_0	5	3	2	3	3	3	3
β_1	6	1	1	1	1	2	2
β_3	7	2	3	3	3	3	2
10	8	2	2	2	1	1	1
11	9	1	2	2	1	1	1
12	10	1	2	2	2	1	1
13	11	2	2	2	2	2	2
14	12	2	3	3	3	2	2
15	13	3	2	3	3	2	2
16	14	2	3	3	2	2	3
17	15	2	2	2	2	2	1
18	16	2	2	2	3	2	3
19	17	2	2	2	2	2	2
20	18	1	2	2	1	1	2
21	19	3	2	3	3	3	3
22	20	3	3	3	3	2	2
23	21	1	1	1	1	2	2
24	22	3	3	3	3	3	2
25	23	3	2	3	3	3	3
26	24	3	2	3	3	3	3
27	25	2	3	3	3	3	2
28	26	2	2	2	1	1	1

29	27	1	2	2	1	2	2
30	28	3	2	3	3	2	2
31	29	3	2	3	3	3	3
32	30	1	2	2	2	1	1
33	31	3	3	3	3	3	2
34	32	3	3	3	2	3	3
$\sum_{i=1}^n x_i$	33	1	1	1	1	1	1
36	34	1	1	1	1	1	1
37	35	3	2	3	2	2	2
38	$\sum_{i=1}^n x_i$						

Figures and tables with captions will be placed in `figure` and `table` environments, respectively **Figura 5.1**

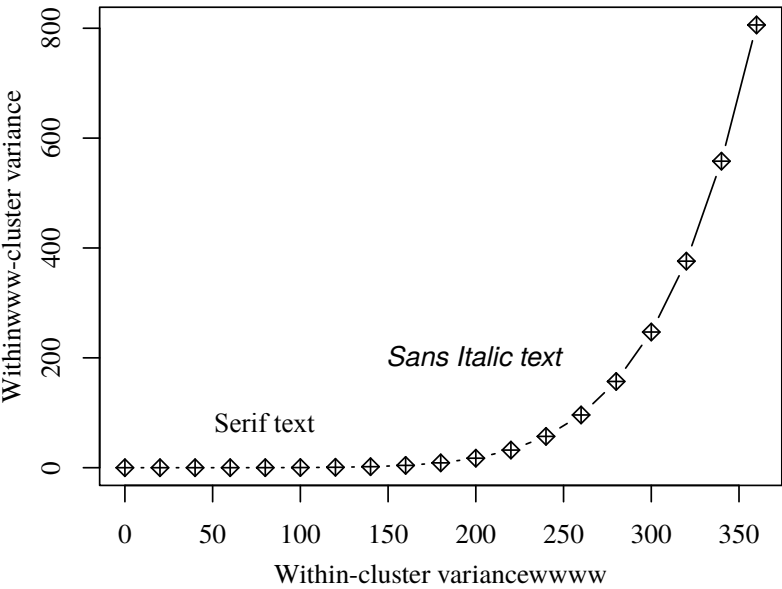


Figura 5.1 Here is a nice figure 100!

Centering

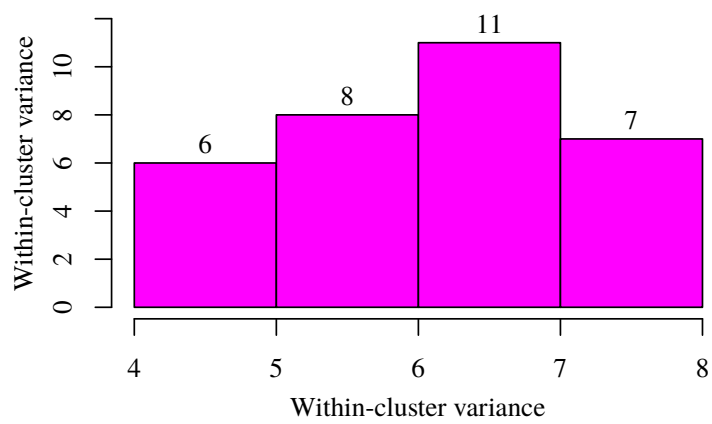


Figura 5.2 Here is a nice figure 100!



6

Matriz y rango



Distribución de frecuencias

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 7.1

Tabla 7.1: 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 además si los datos son muy variados, que para ser 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 Sturges 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 intervállica se obtiene con $l = \frac{L}{r}$

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^*	$\frac{f_2}{n}$	$\frac{F_2}{n}$	$\frac{F_2^*}{n}$...

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 $n = f_1 + f_2 + \dots + f_r = \sum_{i=1}^r f_i$ 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

$$F_m = f_1 + f_2 + \dots + f_m = \sum_{i=1}^m f_i$$

- Las **frecuencias absolutas acumuladas mayor que** F_i^* obedecen a la fórmula

$$F_m^* = f_m + f_{m+1} + \dots + f_r = \sum_{i=m}^r f_i = n - \sum_{i=1}^{m-1} f_i = n - (f_1 + f_2 + \dots + f_{m-1})$$

- 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 7.1. Sean los datos

Solución. Entonces

8

Gráficos estadísticos



9

Medidas de tendencia central

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

9.1. La media (\bar{x})

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

9.1.1. 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

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i \quad (9.1)$$

$$\frac{d[P; F_1]}{d[P; \mathcal{L}_1]} = e = \frac{d[P; F_2]}{d[P; \mathcal{L}_2]} \quad (9.2)$$

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

9.1.2. Media de datos agrupados

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

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

Clase	Clase	f_i	F_i	F_i^*	h_i	H_i	H_i^*	...
$< y_1 - y_2 >$	y_1	f_1	F_1	F_1^*	$\frac{f_1}{n}$	$\frac{F_1}{n}$	$\frac{F_1^*}{n}$...
$< y_2 - y_3 >$	y_2	f_2	F_2	F_2^*	$\frac{f_2}{n}$	$\frac{F_2}{n}$	$\frac{F_2^*}{n}$...
$< y_3 - y_4 >$	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^*	$\frac{f_r}{n}$	$\frac{F_r}{n}$	$\frac{F_r^*}{n}$...

Ejercicio 9.1. Si el promedio de n datos es \bar{x} entonces el promedio del conjunto inicial más un dato adicional x_{n+1} es

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

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

$$\bar{x}' = \frac{n\bar{x} + y_1 + y_2 + \dots + y_r}{n + r}$$

Solución. En efecto sea el promedio

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

9.2. La moda (Mo)

9.2.1. 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 $Mo = x_2$

9.2.2. 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^*	...
$< y_{r-1} - y_r]$	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
-

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

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 $B_{r \times s}$

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

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

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

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

10

Medidas de dispersión



11

Medidas de asimetría



Parte II

Probabilidades



1

Frecuencia relativa clásica



2

Calculando



3

Variable aleatoria



4

Variable acumulativa



5

ww



A

Sumatorias

Una suma de números representados por x_1, x_2, \dots, 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

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

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. $\sum_{i=1}^n x_i$

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

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

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}_{n \times n}$$

de **orden** $n \times m$ tiene **entradas** a_{ij} donde el primer subíndice 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. Álgebra 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 $A \pm B = [a_{ij} \pm b_{ij}]_{n \times m}$, $i = 1, 2, \dots, n$ 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

$$A \cdot B = \left[\sum_{k=1}^p a_{ik} b_{kj} \right]_{n \times q}$$

$$= \begin{pmatrix} \sum_{k=1}^m a_{1k} b_{k1} & \sum_{k=1}^m a_{1k} b_{k2} & \cdots & \sum_{k=1}^m a_{1k} b_{kq} \\ \sum_{k=1}^m a_{2k} b_{k1} & \sum_{k=1}^m a_{2k} b_{k2} & \cdots & \sum_{k=1}^m a_{2k} b_{kq} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^m a_{nk} b_{k1} & \sum_{k=1}^m a_{nk} b_{k2} & \cdots & \sum_{k=1}^m a_{nk} b_{kq} \end{pmatrix}_{n \times q}$$

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 verifican 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

frecuencias absolutas, 20
frecuencias absolutas acumuladas menor
 que, 20
frecuencias absolutas relativas, 20
frecuencias absolutas relativas menor
 que, 20
traza, 45