



CUADERNOS DE TRABAJO

FACULTAD DE ESTUDIOS ESTADÍSTICOS

Estadística Bidimensional Teoría

María Gamboa

Departamento de Estadística y Ciencia de los Datos

Cuaderno de Trabajo número CT-02/2022



UCM

**UNIVERSIDAD
COMPLUTENSE
MADRID**

Los Cuadernos de Trabajo de la Facultad de Estudios Estadísticos constituyen una apuesta por la publicación de los trabajos en curso y de los informes técnicos desarrollados desde la Facultad para servir de apoyo tanto a la docencia como a la investigación.

Los Cuadernos de Trabajo se pueden descargar de la página de la Biblioteca de la Facultad www.ucm.es/BUCM/est/, en la página del Repositorio Institucional UCM E-Prints Complutense y en la sección de investigación de la página del centro www.ucm.es/centros/webs/eest/

CONTACTO:

Biblioteca de la Facultad de Estudios Estadísticos

Universidad Complutense de Madrid

Av. Puerta de Hierro, S/N

28040 Madrid

Tlf. 913944035

buc_est@buc.ucm.es

Los trabajos publicados en la serie Cuadernos de Trabajo de la Facultad de Estudios Estadísticos no están sujetos a ninguna evaluación previa. Las opiniones y análisis que aparecen publicados en los Cuadernos de Trabajo son responsabilidad exclusiva de sus autores.

ISSN: 2341-2550

Resumen

Este trabajo presenta una iniciación al estudio de las distribuciones bidimensionales. En particular va enfocado a alumnos de primer curso de los Grados de Estadística Aplicada y Ciencia de los Datos Aplicada de la Facultad de Estudios Estadísticos. Esta primera toma de contacto permitirá al alumno abordar el estudio conjunto de dos variables estadísticas. En concreto, se presentan las diferentes distribuciones obtenidas a partir de dos características medidas en un conjunto de datos, formas de organizar la información en tablas de frecuencia, estudio de independencia y correlación, características numéricas de las distribuciones analizadas y representaciones gráficas de las mismas. Cada una de las secciones vienen acompañadas de ejemplos y ejercicios para que el alumno asimile todos los conocimientos presentados.

Contents

Resumen	1
List of figures	5
List of tables	6
1 ESTADISTICA BIDIMENSIONAL	1
1.1 Distribución conjunta	1
1.2 Distribuciones marginales	4
1.3 Tablas de doble entrada	4
1.4 Distribuciones condicionadas	11
1.5 Dependencia/ Independencia entre 2 variables	15
1.6 Características numéricas de una distribución bidimensional .	16
1.6.1 Vector de medias	16
1.6.2 Momentos respecto al origen de orden r y s (no centrados)	18
1.6.3 Momentos respecto a la media de orden r y s (centrados)	18
1.6.3.1 Covarianza	19
1.7 Diagramas de dispersión	22

List of Figures

1.1	Gráfico de dispersión de Peso sobre Edad	23
1.2	Gráficos de dispersion de Y sobre X en 4 poblaciones diferentes	24

List of Tables

1.1	Tabla de doble entrada de frecuencias absolutas	5
1.2	Tabla de doble entrada de frecuencias relativas	5
1.3	Tabla de doble entrada de frecuencias absolutas	7
1.4	Tabla de doble entrada de frecuencias relativas	8
1.5	Tabla de doble entrada de frecuencias porcentuales	9
1.6	Distribución de frecuencias absolutas de $X/Y = 10$	11
1.7	Distribución de frecuencias absolutas de $X/Y = 10$	12
1.8	Distribución de frecuencias absolutas de $Y/X = 2$	13
1.9	Distribución de frecuencias absolutas de $Y/X = 2$	13
1.10	Distribución de frecuencias relativas de $Y/X = 2$	14
1.11	Tabla de doble entrada de frecuencias relativas	16
1.12	Temperaturas y número de aparatos de refrigeración vendidos durante una semana	17
1.13	Tabla de doble entrada de frecuencias relativas conjuntas . . .	20
1.14	Tabla de doble entrada de frecuencias relativas conjuntas . . .	21
1.15	Tabla de doble entrada de frecuencias relativas conjuntas . . .	21
1.16	Tabla de doble entrada de frecuencias relativas conjuntas . . .	25
1.17	Tabla de doble entrada de frecuencias absolutas conjuntas . .	25

Chapter 1

ESTADISTICA BIDIMENSIONAL

En este capítulo estudiaremos el comportamiento estadístico de dos características sobre un mismo conjunto de datos. Aprenderemos a extraer la información necesaria para poder obtener resultados y conclusiones sobre el comportamiento de las mismas, así como las relaciones existentes entre ellas.

1.1 Distribución conjunta

Vamos a considerar una población de tamaño n y obtendremos información conjunta sobre ella acerca de dos características X e Y .

Definición 1. *Se denomina variable bidimensional al par (X, Y) formado por dos variables estadísticas estudiadas sobre una población o muestra. Ambas variables estadísticas, X e Y , pueden ser de tipo cualitativo o cuantitativo.*

Ejemplo 1. *La siguiente tabla muestra diversos ejemplos de variables bidimensionales:*

(X, Y)	X	Y
(sexo, color de pelo)	<i>cualitativa</i>	<i>cualitativa</i>
(altura, peso)	<i>cuantitativa</i>	<i>cuantitativa</i>
(profesión, ingresos)	<i>cualitativa</i>	<i>cuantitativa</i>
(nivel de colesterol, riesgo de infarto)	<i>cuantitativa</i>	<i>cualitativa</i>

Cada uno de los valores de la variable (X, Y) viene representado por el par ordenado (x_i, y_j) con $i \in \{1, \dots, k\}$, $j \in \{1, \dots, p\}$ y por lo tanto el número de modalidades de la variable X puede tomar los valores $\{x_1, \dots, x_k\}$ mientras que para la variable Y son $\{y_1, \dots, y_p\}$.

Ejemplo 2. *Vamos a hacer un estudio en la Facultad de Estudios Estadísticos y para ello vamos a analizar las variables: notas obtenidas y tipo de grupo de los alumnos de primer curso del año 2021/2022.*

$X \equiv$ *nota media obtenida en el curso 2021/2022, con $X = \{0, 1, 2, \dots, 10\}$. Es decir en este caso la variable tiene once modalidades y por lo tanto $k = 11$.*
 $Y \equiv$ *grupo de clase al que pertenece, con $Y = \{\text{mañana}, \text{tarde}\}$ y por lo tanto $p = 2$.*

De aquí en adelante nos centraremos en el estudio de variables bidimensionales cuantitativas pudiendo ser estas de tipo discreto o continuo.

Definición 2. *Se denomina frecuencia absoluta conjunta $(n_{i,j})$ de la modalidad (x_i, y_j) al número de individuos de la población que presentan conjuntamente el valor x_i de la variable X y el valor y_j de la variable Y .*

Es importante comprobar que la suma de todas las frecuencias absolutas sea igual al tamaño de la población. Por lo tanto, para una población de tamaño n se cumple

$$\sum_{i=1}^k \sum_{j=1}^p n_{i,j} = n. \quad (1.1)$$

Definición 3. Se denomina frecuencia relativa conjunta ($f_{i,j}$) de la modalidad (x_i, y_j) a la proporción de individuos de la población que presentan conjuntamente el valor x_i de la variable X y el valor y_j de la variable Y .

La frecuencia relativa conjunta se calcula como

$$f_{i,j} = \frac{n_{i,j}}{n}. \quad (1.2)$$

Es importante comprobar que la suma de todas las frecuencias relativas sea igual a uno. Por lo tanto para una población de tamaño n se cumple

$$\sum_{i=1}^k \sum_{j=1}^p f_{i,j} = 1. \quad (1.3)$$

Definición 4. Se denomina frecuencia porcentual conjunta ($p_{i,j}$) de la modalidad (x_i, y_j) al porcentaje de individuos de la población que presentan conjuntamente el valor x_i de la variable X y el valor y_j de la variable Y .

La frecuencia porcentual conjunta se calcula como

$$p_{i,j} = f_{i,j} \times 100. \quad (1.4)$$

Es importante comprobar que la suma de todas las frecuencias porcentuales sea igual a cien. Por lo tanto, para una población de tamaño n dada se cumple

$$\sum_{i=1}^k \sum_{j=1}^p p_{i,j} = 100. \quad (1.5)$$

1.2 Distribuciones marginales

Definición 5. *Dada una variable bidimensional (X, Y) , se denomina distribución marginal de X a la distribución en X de todas las observaciones independientemente del valor que toman en Y .*

Las frecuencias absolutas y relativas marginales, $n_{i.}$ y $f_{i.}$, de X pueden calcularse a partir de las siguientes expresiones:

$$n_{i.} = \sum_{j=1}^p n_{ij}, \text{ con } \sum_{i=1}^k n_{i.} = n, \quad (1.6)$$

$$f_{i.} = \sum_{j=1}^p f_{ij} = \frac{n_{i.}}{n}, \text{ con } \sum_{i=1}^k f_{i.} = 1. \quad (1.7)$$

Definición 6. *Dada una variable bidimensional (X, Y) , se denomina distribución marginal de Y a la distribución en Y de todas las observaciones independientemente del valor que toman en X .*

Las frecuencias absolutas y relativas marginales, $n_{.j}$ y $f_{.j}$, de Y pueden calcularse a partir de las siguientes expresiones:

$$n_{.j} = \sum_{i=1}^k n_{ij}, \text{ con } \sum_{j=1}^p n_{.j} = n, \quad (1.8)$$

$$f_{.j} = \sum_{i=1}^k f_{ij} = \frac{n_{.j}}{n}, \text{ con } \sum_{j=1}^p f_{.j} = 1. \quad (1.9)$$

1.3 Tablas de doble entrada

Para organizar los datos se utilizan las tablas de doble entrada. En los laterales izquierdo y superior aparecen los distintos valores que pueden tomar

las variables X e Y y en el interior cualquiera de las frecuencias descritas. En particular las tabla de doble entrada de frecuencias absolutas tienen la siguiente estructura:

X/Y	$y_1 \dots y_j \dots y_p$	$\sum_{j=1}^p n_{i,j} = n_{i.}$
x_1	$n_{11} \dots n_{1j} \dots n_{1p}$	$n_{1.}$
\dots	\dots	\dots
x_i	$n_{i1} \dots n_{ij} \dots n_{ip}$	$n_{i.}$
\dots	\dots	\dots
x_k	$n_{k1} \dots n_{kj} \dots n_{kp}$	$n_{k.}$
$\sum_{i=1}^k n_{i,j} = n_{.j}$	$n_{.1} \dots n_{.j} \dots n_{.p}$	n

Table 1.1: Tabla de doble entrada de frecuencias absolutas

y las de frecuencias relativas la siguiente:

X/Y	$y_1 \dots y_j \dots y_p$	$\sum_{j=1}^p f_{i,j} = f_{i.}$
x_1	$f_{11} \dots f_{1j} \dots f_{1p}$	$f_{1.}$
\dots	\dots	\dots
x_i	$f_{i1} \dots f_{ij} \dots f_{ip}$	$f_{i.}$
\dots	\dots	\dots
x_k	$f_{k1} \dots f_{kj} \dots f_{kp}$	$f_{k.}$
$\sum_{i=1}^k f_{i,j} = n_{.j}$	$f_{.1} \dots f_{.j} \dots f_{.p}$	1

Table 1.2: Tabla de doble entrada de frecuencias relativas

Ejemplo 3. *En una empresa hay una vacante para un puesto de trabajo. Se presentan 12 personas al mencionado puesto y se decide hacer un test de aptitud profesional para seleccionar al individuo que ocupará el cargo. Además se pregunta a cada uno de los aspirantes por el número de horas dedicadas a preparar dicho test.*

Se obtienen las siguientes observaciones de la variable bidimensional (X, Y) donde X e Y denotan la puntuación obtenida en la prueba y el número de horas dedicadas a la preparación del test, respectivamente. Contruir la tabla de doble entrada de frecuencias absolutas, relativas y porcentuales. Además mostrar las correspondientes distribuciones marginales.

X	Y
0	20
0	25
1	20
3	10
1	10
2	20
2	15
2	25
2	25
3	10
1	10
3	25

A la vista de las observaciones podemos decir que las variables X e Y son

variables cuantitativas discretas tomando los siguientes valores:

$$X = \{x_1, x_2, x_3, x_4\} = \{0, 1, 2, 3\},$$

$$Y = \{y_1, y_2, y_3, y_4\} = \{10, 15, 20, 25\}.$$

Para contruir la tabla de frecuencias absolutas conjuntas $(n_{i,j})$, debemos de contar el número de individuos de la población que presentan conjuntamente la característica i de X y j de Y , para todo $1 \leq i, j \leq 4$. Las distribuciones marginales se encuentran en la parte inferior y derecha de la tabla. En más detalle, la distribución de frecuencias absoutas marginal de X se obtiene sumando por filas las frecuencias absolutas conjuntas para cada valor de X , sin considerar los valores que toma en Y y la de Y sumando por columnas las frecuencias absolutas conjuntas para cada valor de Y , sin considerar los valores que toma en X .

X/Y	10	15	20	25	$\sum_{j=1}^p n_{i,j} = n_{i.}$
0	0	0	1	1	2
1	2	0	1	0	3
2	0	1	1	2	4
3	2	0	0	1	3
$\sum_{i=1}^k n_{i,j} = n_{.j}$	4	1	3	4	12

Table 1.3: Tabla de doble entrada de frecuencias absolutas

Para contruir la tabla de frecuencias relativas conjuntas $(f_{i,j})$, debemos de calcular la proporción de individuos de la población que presentan conjuntamente la característica i de X y j de Y , para todo $1 \leq i, j \leq 4$, es decir

$$f_{ij} = \frac{n_{ij}}{n}.$$

X/Y	10	15	20	25	$\sum_{j=1}^p f_{ij} = f_{i.}$
0	0/12	0/12	1/12	1/12	2/12
1	2/12	0/12	1/12	0/12	3/12
2	0/12	1/12	1/12	2/12	4/12
3	2/12	0/12	0/12	1/12	3/12
$\sum_{i=1}^k f_{ij} = f_{.j}$	4/12	1/12	3/12	4/12	1

Table 1.4: Tabla de doble entrada de frecuencias relativas

Para contruir la tabla de frecuencias porcentuales conjuntas (p_{ij}) , debemos de calcular el porcentaje de individuos de la población que presentan conjuntamente la característica i de X y j de Y , para todo $1 \leq i, j \leq 4$, es decir $p_{i,j} = f_{i,j} \times 100$.

X/Y	10	15	20	25	$\sum_{j=1}^p p_{i,j} = p_{i.}$
0	0%	0%	$8'\widehat{3}\%$	$8'\widehat{3}\%$	$16'\widehat{6}\%$
1	$16'\widehat{6}\%$	0%	$8'\widehat{3}\%$	0%	25%
2	0%	$8'\widehat{3}\%$	$8'\widehat{3}\%$	$16'\widehat{6}\%$	$33'\widehat{3}\%$
3	$16'\widehat{6}\%$	0%	0%	$8'\widehat{3}\%$	25%
$\sum_{i=1}^k p_{i,j} = p_{.j}$	$33'\widehat{3}\%$	$8'\widehat{3}\%$	25%	$33'\widehat{3}\%$	100%

Table 1.5: Tabla de doble entrada de frecuencias porcentuales

Ejercicio 1. *En una oficina se quiere estudiar la condición familiar de los 12 trabajadores que la componen. Para ello se pregunta acerca del número de hijos y casas en propiedad que tiene cada uno de los trabajadores. Analizamos la variable bidimensional (X, Y) donde X e Y denotan el número de hijos y casas en propiedad que se tienen, respectivamente y cuyas observaciones se muestran en la siguiente tabla:*

X	Y
0	2
0	2
1	2
2	1
1	1
2	2
2	1
2	2
2	2
2	1
1	1
2	2

1. *Calcular las tablas de doble entrada (absolutas, relativas y porcentuales).*
2. *Calcular el número de individuos que tienen al menos un hijo.*
3. *Porcentaje de individuos que tienen dos propiedades.*
4. *Proporción de individuos que tienen un hijo y una propiedad.*
5. *Número de individuos que tienen al menos un hijo y dos propiedades.*
6. *Calcular la media de la variable X .*
7. *Calcular la moda de la variable Y .*

1.4 Distribuciones condicionadas

Definición 7. *Dada una variable bidimensional (X, Y) , se denomina distribución condicionada de X por un valor de y_j y se denota $(X/Y = y_j)$, a la distribución en X de todas aquellas observaciones que presentan el valor y_j en Y .*

En el Ejemplo 3, la distribución de $(X/Y = 10)$ se obtiene quedándonos con la columna de $Y = 10$ y sus respectivos valores en X , es decir:

$X/Y = 10$	$n_{.1}$
0	0
1	2
2	0
3	2

Table 1.6: Distribución de frecuencias absolutas de $X/Y = 10$

Como podemos observar, cuando $Y = 10$ la variable X no presenta ninguna observación con los valores $X = 0$ o $X = 2$. Por ello, podrían eliminarse esas filas de la tabla anterior y la distribución condicionada quedaría de la siguiente forma:

$X/Y = 10$	$n_{.1}$
1	2
3	2

Table 1.7: Distribución de frecuencias absolutas de $X/Y = 10$

Las frecuencias relativas condicionadas, f_i^j , denotan a la proporción de individuos que presentan el valor x_i de X entre todos aquellos que presentan el valor y_j de Y y pueden calcularse a partir de la siguiente expresión:

$$f_i^j = \frac{n_{ij}}{n_{.j}}, \quad (1.10)$$

y además debe cumplirse que

$$\sum_{i=1}^k f_i^j = 1. \quad (1.11)$$

Definición 8. *Dada una variable bidimensional (X, Y) , se denomina distribución condicionada de Y por un valor de x_i y se denota $(Y/X = x_i)$, a la distribución en Y de todas aquellas observaciones que presentan el valor x_i en X .*

En el Ejemplo 3, la distribución de $(Y/X = 2)$ se obtiene quedándonos con la fila de $X = 2$ y sus respectivos valores en Y , es decir:

$Y/X = 2$	$n_{3.}$
10	0
15	1
20	1
25	2

Table 1.8: Distribución de frecuencias absolutas de $Y/X = 2$

Como podemos observar cuando $X = 2$ la variable Y no presenta ninguna observación con el valor $Y = 10$. Por ello, podría eliminarse esa fila de la tabla anterior y la distribución condicionada quedaría de la siguiente forma:

$Y/X = 2$	$n_{3.}$
15	1
20	1
25	2

Table 1.9: Distribución de frecuencias absolutas de $Y/X = 2$

Las frecuencias relativas condicionadas f_j^i denotan a la proporción de individuos que presentan el valor y_j de Y entre todos aquellos que presentan el valor x_i de X y pueden calcularse a partir de la siguiente expresión:

$$f_j^i = \frac{n_{ij}}{n_{i.}}, \quad (1.12)$$

y además debe cumplirse que

$$\sum_{j=1}^p f_j^i = 1. \quad (1.13)$$

En el Ejemplo 3 estas frecuencias relativas condicionadas, f_j^i , quedarían de la forma:

$Y/X = 2$	f_j^i
15	1/4
20	1/4
25	2/4
Σ	1

Table 1.10: Distribución de frecuencias relativas de $Y/X = 2$

Ejercicio 2. *A partir de los datos del Ejercicio 1:*

- *Mostrar todas las distribuciones condicionadas de frecuencias absolutas y sus respectivas frecuencias relativas.*
- *Calcular la media de la variable $X/Y = 1$.*
- *Calcular la moda de la variable $Y/X = 0$.*

Relación entre los tres tipos de distribuciones:

$$\begin{aligned} f_{ij} &= \frac{n_{ij}}{n} \\ &= \frac{n_{i.}}{n} \frac{n_{.j}}{n_{i.}} = f_{i.} f_{.j}^i \\ &= \frac{n_{.j}}{n} \frac{n_{ij}}{n_{.j}} = f_{.j} f_i^j \end{aligned}$$

1.5 Dependencia/ Independencia entre 2 variables

Definicion 9. *Las variables X e Y son independientes funcionalmente cuando no existe ninguna relación funcional entre ellas es decir, el comportamiento de una de ellas no se ve afectado por la otra. En este caso, $X/Y = y_j$ es siempre la misma sea cual sea el valor de y_j en Y .*

Para ver si dos variables estadísticas son independientes se comprueba que para todos los pares de valores (x_i, y_j) se cumple que la frecuencia relativa conjunta es igual al producto de las frecuencias relativas marginales, es decir:

$$f_{ij} = f_{i.}f_{.j}, \quad (1.14)$$

o equivalentemente

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

Definicion 10. *Las variables X e Y tienen dependencia funcional recíproca cuando $k=p$ y además:*

Y depende funcionalmente de X si: $j = f^{-1}(i)$,

X depende funcionalmente de Y si: $i = f^{-1}(j)$.

En el Ejemplo 3, las variables X e Y no son independientes pues hemos encontrado al menos un producto que no satisface la Expresión 1.14:

$$f_{ij} = f_{i.}f_{.j} \longrightarrow \frac{0}{12} \neq \frac{4}{12} \times \frac{2}{12},$$

X/Y	10	15	20	25	$\sum_{j=1}^p f_{i,j} = f_{i.}$
0	0/12	0/12	1/12	1/12	2/12
1	2/12	0/12	1/12	0/12	3/12
2	0/12	1/12	1/12	2/12	4/12
3	2/12	0/12	0/12	1/12	3/12
$\sum_{i=1}^k f_{i,j} = f_{.j}$	4/12	1/12	3/12	4/12	1

Table 1.11: Tabla de doble entrada de frecuencias relativas

Cuando hay un cierto grado de dependencia se dice que hay dependencia estadística.

Ejercicio 3. *Con los datos del Ejercicio 1, decidir si hay dependencia estadística.*

1.6 Características numéricas de una distribución bidimensional

1.6.1 Vector de medias

Consideramos la distribución estadística bidimensional, (X, Y) , el centro de gravedad de la distribución viene dado por el par (\bar{x}, \bar{y}) y se le denomina vector de medias donde cada una de las componentes respresentan las medias marginales de las variables X e Y , respectivamente. Si los datos los tenemos agrupados por intervalos procedemos de la misma forma que lo hacíamos en las tablas de frecuencias de variables unidimensionales es decir, obtendremos las marcas de clase de cada uno de los intervalos.

Ejemplo 4. *En una tienda de electrodomésticos se anotan durante una semana las diferentes temperaturas que marca el termómetro situado en el exterior del centro comercial donde se halla situado a las 12:00 de la mañana y el número de aparatos de refrigeración vendidos ese mismo día. Calcular el vector de medias si las medidas obtenidas son las siguientes:*

X	Y
35.0	9
33.0	2
37.0	11
34.0	4
29.0	3
30.0	2
30.5	3

Table 1.12: Temperaturas y número de aparatos de refrigeración vendidos durante una semana

Para poder calcular el vector de medias utilizo las distribuciones marginales de ambas variables y me calculo sus correspondientes medias aritméticas.

$$\bar{x} = \frac{\sum_{i=1}^7 x_i}{7} = \frac{35 + 33 + 37 + 34 + 29 + 30 + 30.5}{7} = 32.64. \quad (1.15)$$

$$\bar{y} = \frac{\sum_{j=1}^7 y_j}{7} = \frac{9 + 2 + 11 + 4 + 3 + 2 + 3}{7} = 4.85. \quad (1.16)$$

Por lo tanto el vector de medias es $(\bar{x}, \bar{y}) = (32.64, 4.85)$.

Si hubieramos tenido los datos organizados en una tabla de doble entrada deberíamos de calcular previamente las distribuciones marginales y a partir de ellas obtendríamos las medias aritméticas para cada una de las variables.

1.6.2 Momentos respecto al origen de orden r y s (no centrados)

El momento respecto al origen de orden (r, s) de la variable bidimensional (X, Y) denotado por m_{rs} , se define como

$$\begin{aligned} m_{rs} &= \frac{\sum_{i=1}^k \sum_{j=1}^p n_{ij} x_i^r y_j^s}{n} \\ &= \sum_{i=1}^k \sum_{j=1}^p f_{ij} x_i^r y_j^s. \end{aligned} \quad (1.17)$$

Casos particulares:

m_{10} = es la media marginal de X .

m_{01} = es la media marginal de Y .

m_{r0} = es el momento respecto al origen de orden r marginal de X .

m_{0s} = es el momento respecto al origen de orden s marginal de Y .

1.6.3 Momentos respecto a la media de orden r y s (centrados)

El momento respecto a la media de orden (r, s) de la variable bidimensional (X, Y) denotado por M_{rs} , se define como

$$\begin{aligned} M_{rs} &= \frac{\sum_{i=1}^k \sum_{j=1}^p n_{ij} (x_i - \bar{x})^r (y_j - \bar{y})^s}{n} \\ &= \sum_{i=1}^k \sum_{j=1}^p f_{ij} (x_i - \bar{x})^r (y_j - \bar{y})^s. \end{aligned} \quad (1.18)$$

Casos particulares:

M_{20} = es la varianza de X .

M_{02} = es la varianza de Y .

M_{r0} = es el momento respecto a la media de orden r marginal de X .

M_{0s} = es el momento respecto a la media de orden s marginal de Y .

$M_{01} = M_{10} = 0$.

Relación entre momentos conjuntos y marginales en caso de independencia

Cuando las variables X e Y son independientes entre sí los momentos respecto del origen y de la media satisfacen las siguientes expresiones

$$m_{rs} = m_{r0}m_{0s},$$

$$M_{rs} = M_{r0}M_{0s}.$$

1.6.3.1 Covarianza

La covarianza denotada por $cov(X, Y)$, es una característica numérica de las distribuciones bidimensionales que mide la variación conjunta de X e Y . Se corresponde con el momento centrado $M_{1,1}$ y se calcula a partir de cualquiera de las siguientes expresiones:

$$\begin{aligned} Cov(X, Y) = M_{11} &= \frac{\sum_{i=1}^k \sum_{j=1}^p n_{ij}(x_i - \bar{x})(y_j - \bar{y})}{n} \\ &= \frac{\sum_{i=1}^k \sum_{j=1}^p n_{ij}x_i y_j}{n} - \bar{x}\bar{y}. \end{aligned}$$

Si el valor de covarianza obtenido es positivo indica que si una variable aumenta la otra variable también lo hace es decir, existe una relación positiva o directa entre ellas. En cambio valores de covarianzas negativos indicarán que a medida de que una de las variables aumenta la otra disminuye y por

lo tanto, tienen una relación negativa o inversa. Valores de covarianza nulos o muy próximas a cero indican que hay independencia entre las variables X e Y o de existir relación, no sería lineal.

Propiedad 1. X, Y Independientes $\Rightarrow Cov(X, Y) = 0$.

Debemos de tener el cuenta que el contrarrecíproco de la Propiedad 1 siempre se cumple es decir, un valor de la covarianza distinto a 0 indica que ambas variables no son independientes entre sí.

Propiedad 2. *La covarianza solamente se ve afectada por los cambios de escala. En particular, dadas las transformaciones $X' = aX + b$, $Y' = cY + d$,*

$$Cov(X', Y') = acCov(X, Y). \quad (1.19)$$

Ejemplo 5. *Calcular la covarianza entre las variables X e Y que denotan dos tipos de intereses de un préstamo bancario a partir de la tabla de frecuencias relativa conjunta siguiente*

X/Y	5-10	10-15
0 – 5	0.06	0.02
5 – 10	0.14	0.38
10 – 15	0.30	0.10

Table 1.13: Tabla de doble entrada de frecuencias relativas conjuntas

Debo de calcular previamente las marcas de clase y las medias marginales para después obtener las diferencias entre las marcas de clase de cada una de las variables con sus respectivas medias.

X/Y	5-10	10-15	$f_{i.}$
0 – 5	0.06	0.02	0.08
5 – 10	0.14	0.38	0.52
10 – 15	0.30	0.10	0.40
$f_{.j}$	0.50	0.50	1

Table 1.14: Tabla de doble entrada de frecuencias relativas conjuntas

Las marcas de clase de la variable X son $c_i = \{2.5, 7.5, 12.5\}$ con $i = \{1, 2, 3\}$ y de la variable Y son: $c_j = \{7.5, 12.5\}$ con $j = \{1, 2\}$

c_i/c_j	7.5	12.5	$f_{i.}$
2.5	0.06	0.02	0.08
7.5	0.14	0.38	0.52
12.5	0.30	0.10	0.40
$f_{.j}$	0.50	0.50	1

Table 1.15: Tabla de doble entrada de frecuencias relativas conjuntas

Por lo tanto las medias de ambas variables se obtienen a partir de las siguientes expresiones

$$\bar{x} = 2.5 \times 0.08 + 7.5 \times 0.52 + 12.5 \times 0.40 = 9.1, \quad (1.20)$$

$$\bar{y} = 7.5 \times 0.5 + 12.5 \times 0.5 = 10. \quad (1.21)$$

$$\begin{aligned} Cov(X, Y) &= \sum_{i=1}^k \sum_{j=1}^p f_{ij} x_i y_j - \bar{x} \bar{y} \\ &= 0.06 \times 2.5 \times 7.5 + 0.02 \times 2.5 \times 12.5 + 0.14 \times 7.5 \times 7.5 \\ &\quad + 0.38 \times 7.5 \times 12.5 + 0.30 \times 12.5 \times 7.5 + 0.10 \times 12.5 \times 12.5 \\ &\quad - 9.1 \times 10 = 89 - 91 = -2. \end{aligned}$$

El valor de la covarianza es distinto de cero y negativo por lo que podemos afirmar que ambas variables no son independientes entre sí y que entre ellas existe una relación inversa. Esto significa que a medida que un tipo de interés crece el otro decrece.

1.7 Diagramas de dispersión

En los estudios estadísticos es importante obtener representaciones gráficas de las distribuciones analizadas para poder detectar las posibles relaciones existentes entre las variables involucradas.

En esta sección mostramos cómo representar gráficamente una variable bidimensional (X, Y) mediante diagramas de dispersión.

El diagrama de dispersión es un gráfico muy utilizado en estadística donde se representan los pares observados de datos en los ejes cartesianos permitiendo describir la relación existente entre las dos variables analizadas de una forma visual.

Para poder identificar si existe alguna relación entre las variables debemos de observar si en la nube de puntos, obtenida tras la representación gráfica, aparece alguna estructura matemática conocida tales como una línea recta

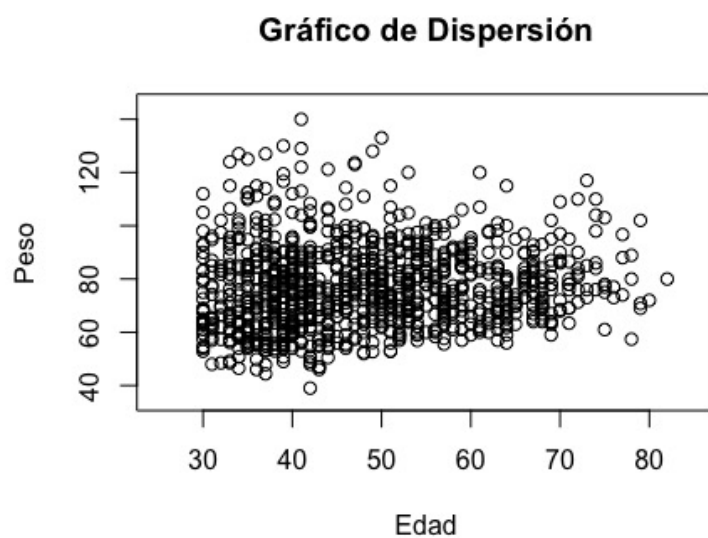
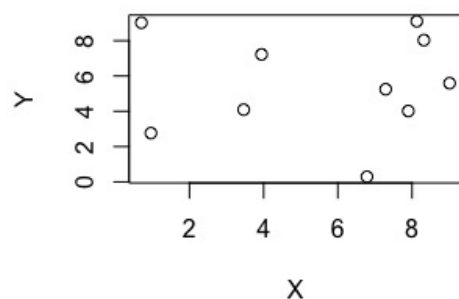
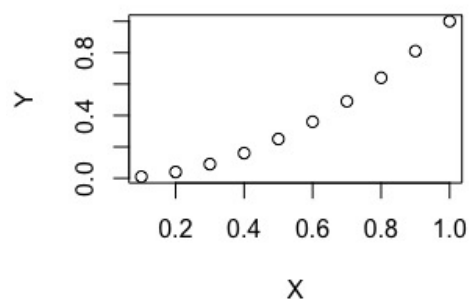


Figure 1.1: Gráfico de dispersión de Peso sobre Edad

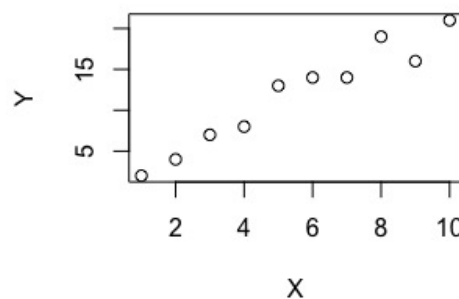
o parábola o si de lo contrario la disposición de los puntos no se asemeja a ninguna de ellas.



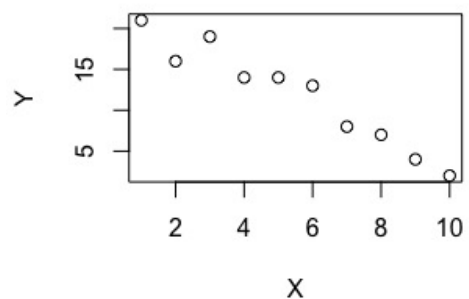
(a) Ausencia de relación



(b) Fuerte asociación pero no lineal



(c) Asociación lineal positiva



(d) Asociación lineal negativa

Figure 1.2: Gráficos de dispersion de Y sobre X en 4 poblaciones diferentes

Ejercicio 4. *La siguiente tabla recoge, en frecuencias relativas, la distribución del número de exámenes que deben de realizar los aspirantes a un determinado puesto trabajo (variable X) y las puntuaciones obtenidas por cada individuo en un examen preliminar (variable Y)*

X/Y	5	10	15
0	0.16	0	0
1	0.14	0.18	0.06
2	0.20	0.20	0.06

Table 1.16: Tabla de doble entrada de frecuencias relativas conjuntas

- *Representar gráficamente la distribución conjunta.*
- *Calcular la covarianza entre las dos variables.*
- *¿Son independientes las variables?*
- *Distribución del número de exámenes a realizar de entre los que han obtenido una puntuación preliminar mayor o igual a 10.*
- *Calcular el centro de gravedad de la distribución (X, Y)*

Ejercicio 5. *La siguiente tabla recoge, en frecuencias absolutas, la distribución de la nota obtenida en una primera prueba teórica para el acceso a un máster de música electrónica (variable X) y las puntuaciones obtenidas por los mismos individuos en un segundo examen práctico (variable Y)*

X/Y	2	4	6	8
10	16	20	20	10
20	14	18	6	1
30	20	20	16	15

Table 1.17: Tabla de doble entrada de frecuencias absolutas conjuntas

- *Representar gráficamente la distribución conjunta.*
- *Calcular la covarianza entre las dos variables.*
- *¿Son independientes las variables?*
- *Si no hay independencia entre las variables, ¿qué tipo de relación podrían tener a la vista de los datos?*
- *Calcular la varianza de X .*
- *Calcular la varianza de Y .*
- *Calcular M_{01}*
- *Calcular la distribución de puntos totales obtenidos en ambos exámenes.*

Bibliography

- BARÓ LLANS. "Estadística descriptiva". Ed. Parramón.
- BYRKIT. "Statistics today. a comprehensive introduction". The Benjamin/ Cummings Publ. Co., Inc. California 1987.
- I. CASTILLO- M. GUIJARRO: Estadística descriptiva y cálculo de probabilidades. Pearson, 2005.
- DURÁ PEIRO LÓPEZ CUÑAT: Fundamentos de estadística. estadística descriptiva y modelos probabilísticos para la influencia. Ed. Ariel.
- MOORE, D.S. McCabe, G. P.: Introduction to the practice of statistics. N York W.H. Freeman, 2003.
- PEÑA ROMO: Introducción a la estadística para ciencias sociales. Mc Graw Hill 1995.
- QUESADA ISIDORO LÓPEZ: Curso y ejercicios de estadística. Ed. Alhambra.
- V. TOMEU- I. UÑA: Lecciones de estadística descriptiva. Ed. Thompson.
- ROSS, S.: Introducción a la Estadística. Ed. Reverté, 2008