

6. DISTRIBUCIONES BIDIMENSIONALES

Estadística Descriptiva
Dr. Francisco Rabadán Pérez

Índice

1. Distribución Bidimensional de frecuencias
 1. *Independencia y relación funcional*
 2. *Tablas de doble entrada: correlación y contingencia*
 3. *Distribuciones Marginales*
 4. *Distribuciones condicionadas*
 5. *Independencia Estadística*
2. Representaciones gráficas
 - Apéndice: Momentos de distribuciones bidimensionales

1. Distribución bidimensional de frecuencias

- Variable bidimensional (x_i, y_j)
- Podemos estudiar ambas variables por separado.
- Lo mas interesante es ver qué relación existe entre ambas.
- Si observamos una relación de dependencia, **podemos formular la hipótesis de una relación causal**.

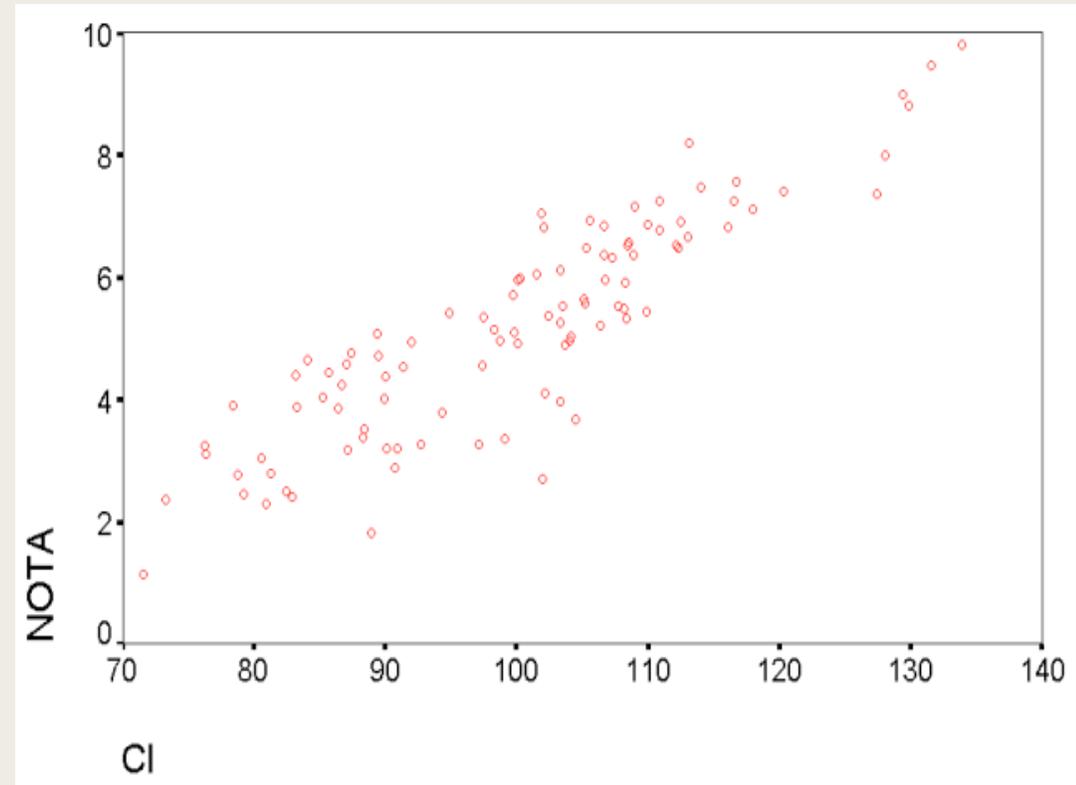


Diagrama de dispersión:

<http://matematicas1bc.blogspot.com.es/2012/05/estadistica.html>

1.1. Independencia y relación funcional de (X,Y)

Queremos conocer el **grado y modo** en que se relacionan dos caracteres (variables o atributos).

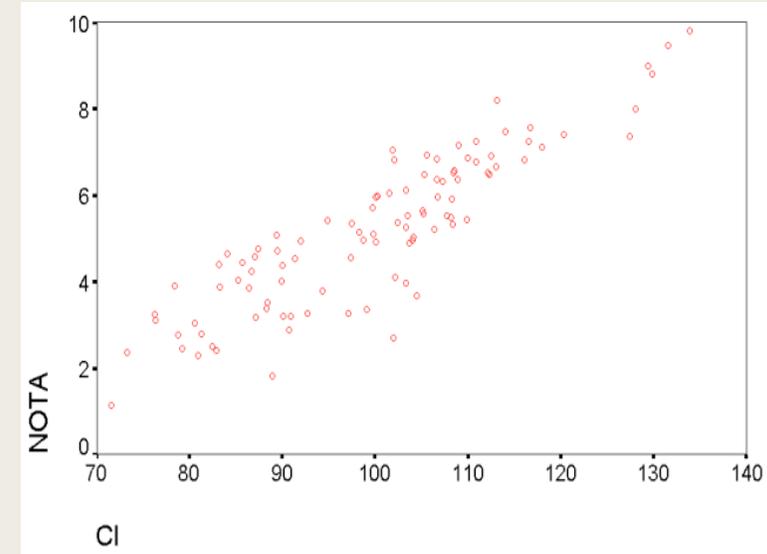
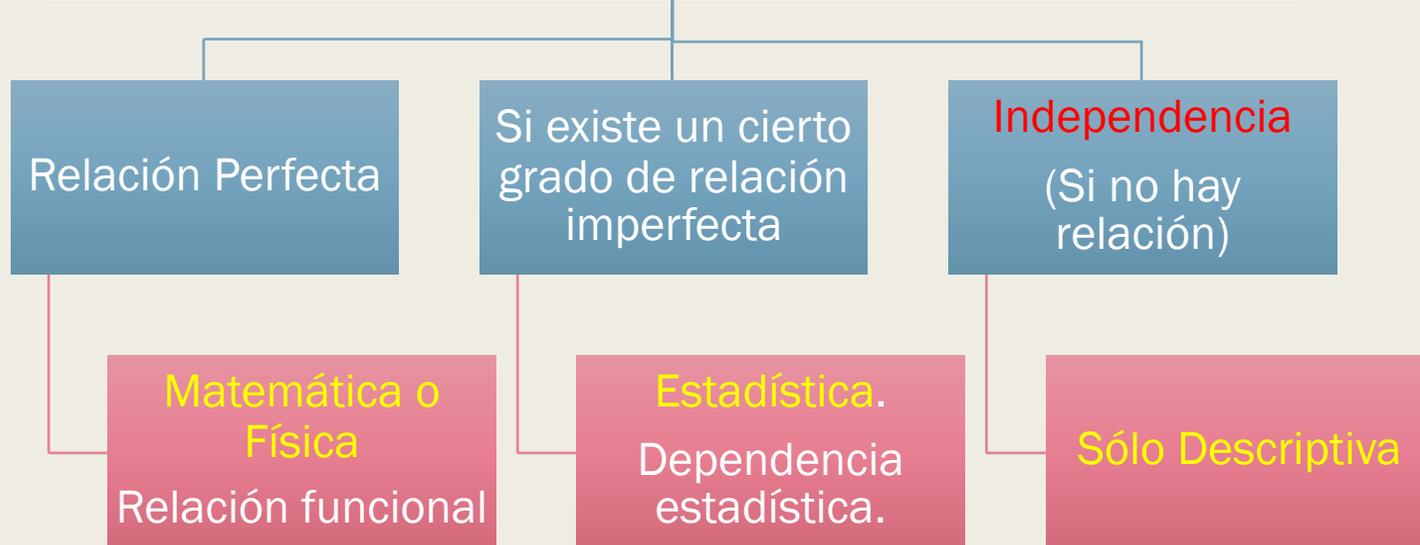
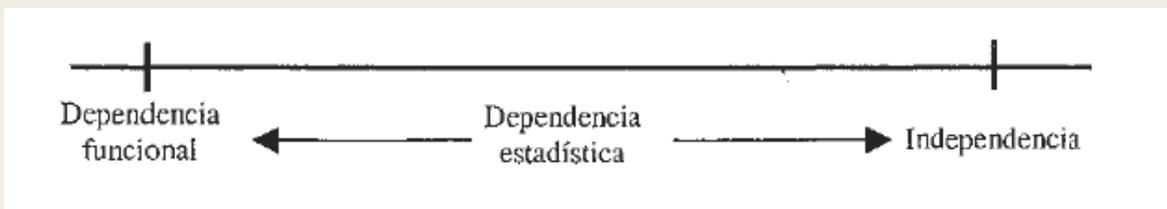
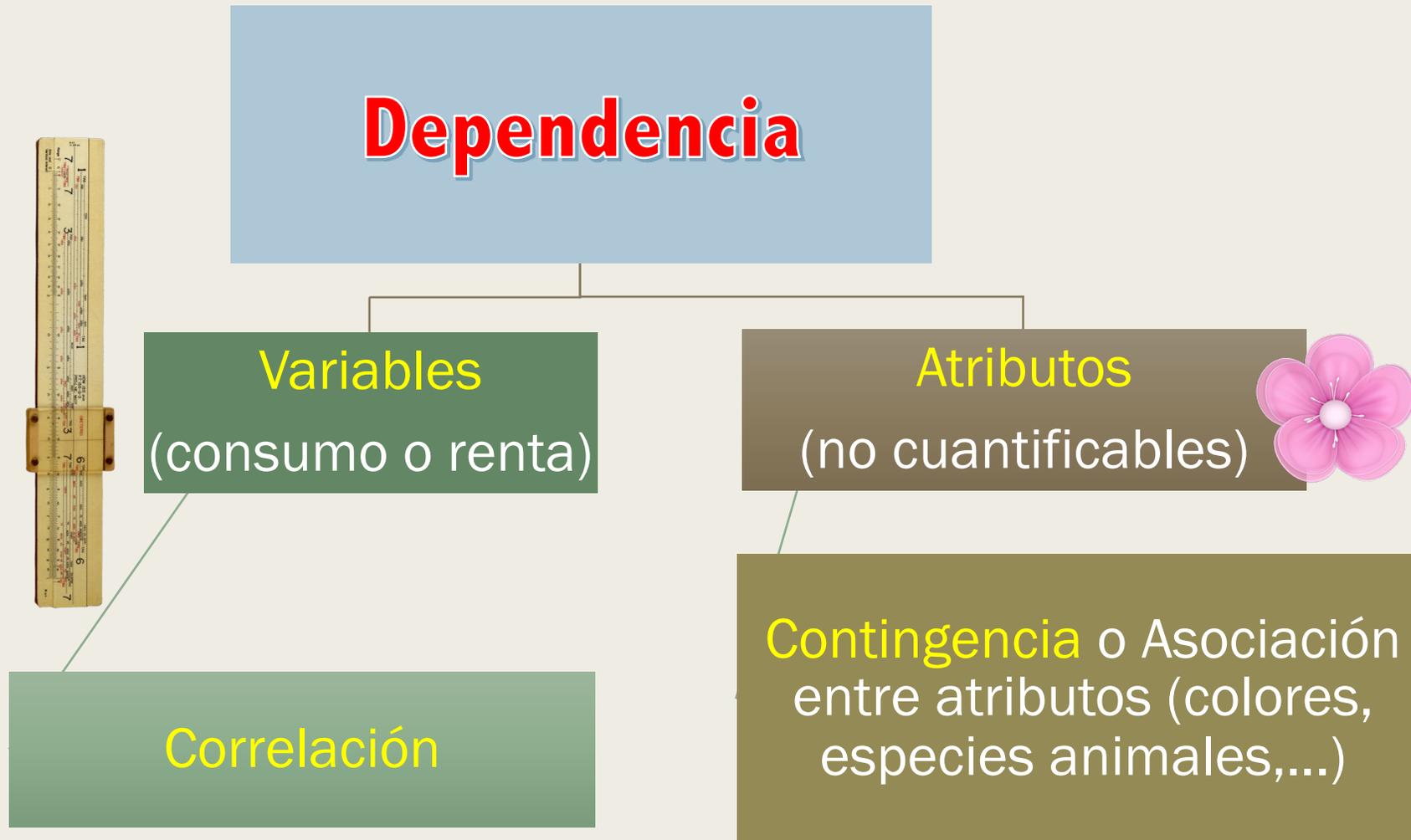


Diagrama de dispersión:
<http://matematicas1bc.blogspot.com.es/2012/05/estadistica.html>



(Martín-Pliego, 2011, pág. 199)

1.1. Independencia y relación funcional de (X,Y)



1.2. Tablas de doble entrada: correlación y contingencia

Tablas de doble entrada

De Correlación

- Variables Cuantitativas



De Contingencia

- Variables Cualitativas
- Atributos



	y_1	y_2	...	y_j	...	y_k	$n_{i.}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2k}	$n_{2.}$
x_3	n_{31}	n_{32}	...	n_{3j}	...	n_{3k}	$n_{3.}$
...	n_{4k}	$n_{4.}$
...	n_{ij}	$n_{i.}$
x_n	n_{n1}	n_{n2}	...	n_{nj}	...	n_{nk}	$n_{n.}$
$n_{.j}$	$n_{.1}$	$n_{.2}$	$n_{.j}$	$n_{.j}$...	$n_{.k}$	N

1.2. Tablas de doble entrada: correlación y contingencia

- La tabla de doble entrada se puede transformar, por ejemplo, para facilitar el cálculo en una **tabla de 3 columnas**.
- Las dos primeras columnas indican el par (x_i, y_j)
- La tercera columna contiene la frecuencia conjunta asociada (n_{ij} u otras)

x_i	y_j	n_{ij}
x_1	y_1	n_{11}
x_2	y_2	n_{12}
x_3	y_3	n_{12}
...
x_i	x_j	
...
x_n	y_k	n_{h2}

1.2. Tablas de doble entrada: correlación y contingencia

Tipos de frecuencias

- Frecuencia absoluta conjunta (n_{ij})
- Frecuencia relativa conjunta (f_{ij})

$$f_{ij} = \frac{n_{ij}}{N}$$

- Frecuencia acumuladas (F_{i*} , o F_{*j})::
 - Acumular respecto a x
 - Acumular respecto a y
 - Acumular simultáneamente.

x_i	y_j	n_{ij}
x_1	y_1	n_{11}
x_2	y_2	n_{12}
x_3	y_3	n_{12}
...
x_i	x_j	
...
x_n	y_k	n_{h2}

1.2. Tablas de doble entrada: correlación y contingencia

- La primera fila (y_j) y la primera columna (x_i) muestran los valores ordenados de las variables.
- Dentro de la tablas vemos la **frecuencia conjunta** (n_{ij}): podrían ser también
 - relativas (f_{ij}),
 - acumuladas (N_{ij}), o
 - relativas acumuladas (F_{ij}).

	y_1	y_2	...	y_j	...	y_k	$n_{i.}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2k}	$n_{2.}$
x_3	n_{31}	n_{32}	...	n_{3j}	...	n_{3k}	$n_{3.}$
...	n_{4k}	$n_{4.}$
...	n_{ij}	$n_{i.}$
x_n	n_{n1}	n_{n2}	...	n_{nj}	...	n_{nk}	$n_{n.}$
$n_{.j}$	$n_{.1}$	$n_{.2}$	$n_{.j}$	$n_{.j}$		$n_{.k}$	N

1.3. Distribuciones marginales

- **Frecuencia marginal:**
 - No nos importa el valor que toma la otra variable.
 - Coincide con la frecuencia de la variable unidimensional.
 - Ponemos un *, donde la variable no importa que tome cualquier valor.

$$n_{i.} = \sum_{j=1}^k n_{ij}$$

f. abs. Marginal de X

$$n_{.j} = \sum_{i=1}^h n_{ij}$$

f. abs. Marginal de Y

Es claro que:

$$\sum_{i=1}^h n_{i.} = \sum_{j=1}^k n_{.j} = \sum_{i=1}^h \sum_{j=1}^k n_{ij} = N$$

	y_1	y_2	...	y_j	...	y_k	$n_{i.}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2k}	$n_{2.}$
x_3	n_{31}	n_{32}	...	n_{3j}	...	n_{3k}	$n_{3.}$
...	n_{4k}	$n_{4.}$
...	n_{ij}	$n_{i.}$
x_n	n_{n1}	n_{n2}	...	n_{nj}	...	n_{nk}	$n_{n.}$
$n_{.j}$	$n_{.1}$	$n_{.2}$	$n_{.j}$	$n_{.j}$		$n_{.k}$	N

1.4. Distribuciones condicionadas

(Martín-Pliego, 2011, pág. 203)

- Estudiamos las frecuencias de una variable para un conjunto concreto de valores de la otra.
- Ej:
 - $\mathbf{x_i / (y_1, y_2)}$: (x_i, n_{ij}) siendo $y_j = (y_1, y_2)$
 - $\mathbf{y_j / x_4}$: (y_j, n_{ij}) siendo $x_i = x_4$
- Notese que **“/”** no significa división, sino **“condicionado a”**
- En el caso de que esté condicionado a un único valor, la anotaremos como

$$n_{i/j}; n_{j/i}$$

x_i / y_2	$n_{i/2}$
x_1	n_{12}
x_2	n_{22}
\vdots	\vdots
x_i	n_{i2}
\vdots	\vdots
x_h	n_{h2}
	$n_{.2}$

x_i / y_j	$n_{i/j}$
x_1	n_{1j}
x_2	n_{2j}
\vdots	\vdots
x_i	n_{ij}
\vdots	\vdots
x_h	n_{hj}
	$n_{.j}$

Podemos calcular las frecuencias condicionadas relativas **dividiendo por el total de datos que cumplen la característica condicionante**

$$f_{i/j} = \frac{n_{ij}}{n_{.j}}$$

f. Relativa de x_i condicionada a y_j

$$f_{j/i} = \frac{n_{ij}}{n_{i.}}$$

f. Relativa de y_j condicionada a x_i

1.4. Distribuciones condicionadas

EJEMPLO

Sea la siguiente tabla de doble entrada

X \ Y	1	2	3	4	$n_{i\cdot}$
5	1	2	1	3	7
10	2	1	3	2	8
15	3	2	1	2	8
$n_{\cdot j}$	6	5	5	7	23

Se pide calcular:

- (a) La distribución marginal de la Y.
- (b) La distribución condicionada de $X/Y = 2$.

y_j	$n_{\cdot j}$	$n_{\cdot j}/N$
1	6	6/23
2	5	5/23
3	5	5/23
4	7	7/23
	23	1

$x_{i/y=2}$	$n_{i/2}$	$f_{i/2}$
5	2	2/5
10	1	1/5
15	2	2/5
	5	1

1.5. Independencia estadística

- “Dos variables X e Y se dice que son estadísticamente independientes cuando la frecuencia relativa conjunta es igual al producto de las frecuencias relativas marginales” (Martín-Pliego, 2011; pág 205)

$$\frac{n_{ij}}{N} = \frac{n_{i.}}{N} \frac{n_{.j}}{N}, \quad \forall (i, j) \rightarrow \begin{cases} f_{i/j} = \frac{n_{ij}}{n_{.j}} = \frac{n_{i.} \frac{n_{.j}}{N}}{n_{.j}} = \frac{n_{i.}}{N} \\ f_{j/i} = \frac{n_{ij}}{n_{i.}} = \frac{n_{.j} \frac{n_{i.}}{N}}{n_{i.}} = \frac{n_{.j}}{N} \end{cases}$$

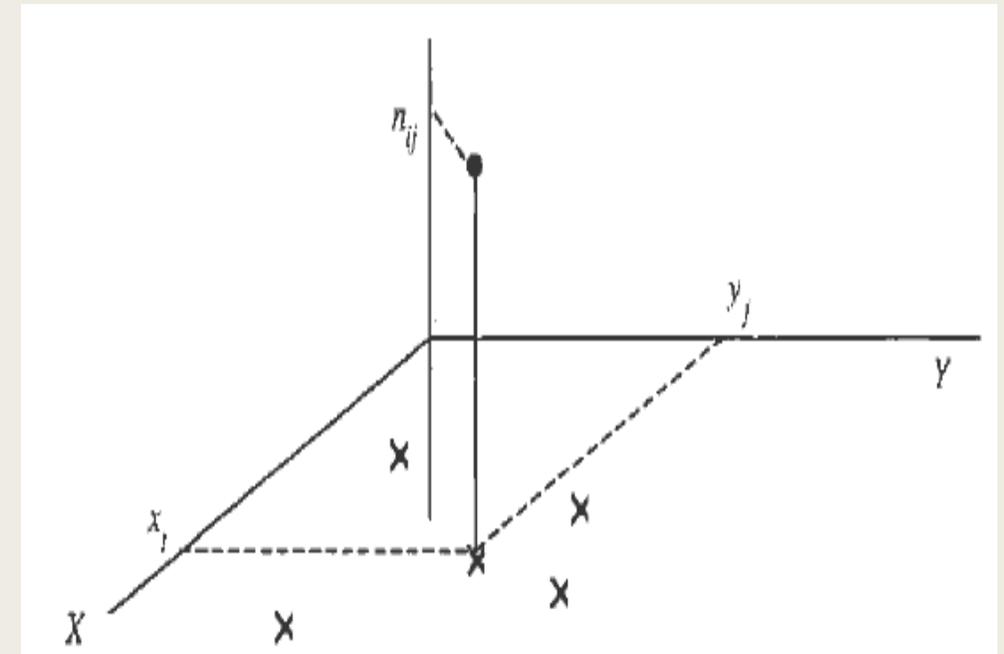
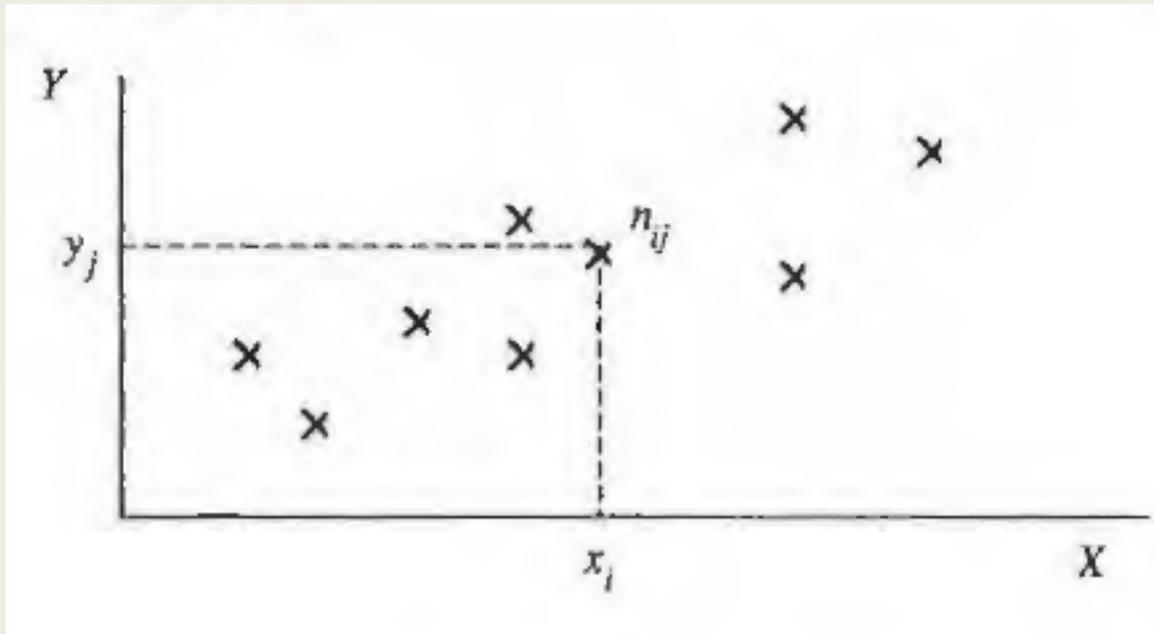
IMPORTANTE

(Martín-Pliego, 2011, pág. 205)

“..., las frecuencias relativas condicionadas son iguales a sus correspondientes frecuencias relativas marginales, lo que nos indica que el condicionamiento, en cuanto tal, no existe: las variables son independientes, puesto que en las distribuciones marginales se estudia el comportamiento de una variable con independencia de los valores que pueda tomar la otra” (Martín-Pliego, 2011; pág 205)

2. Representaciones gráficas

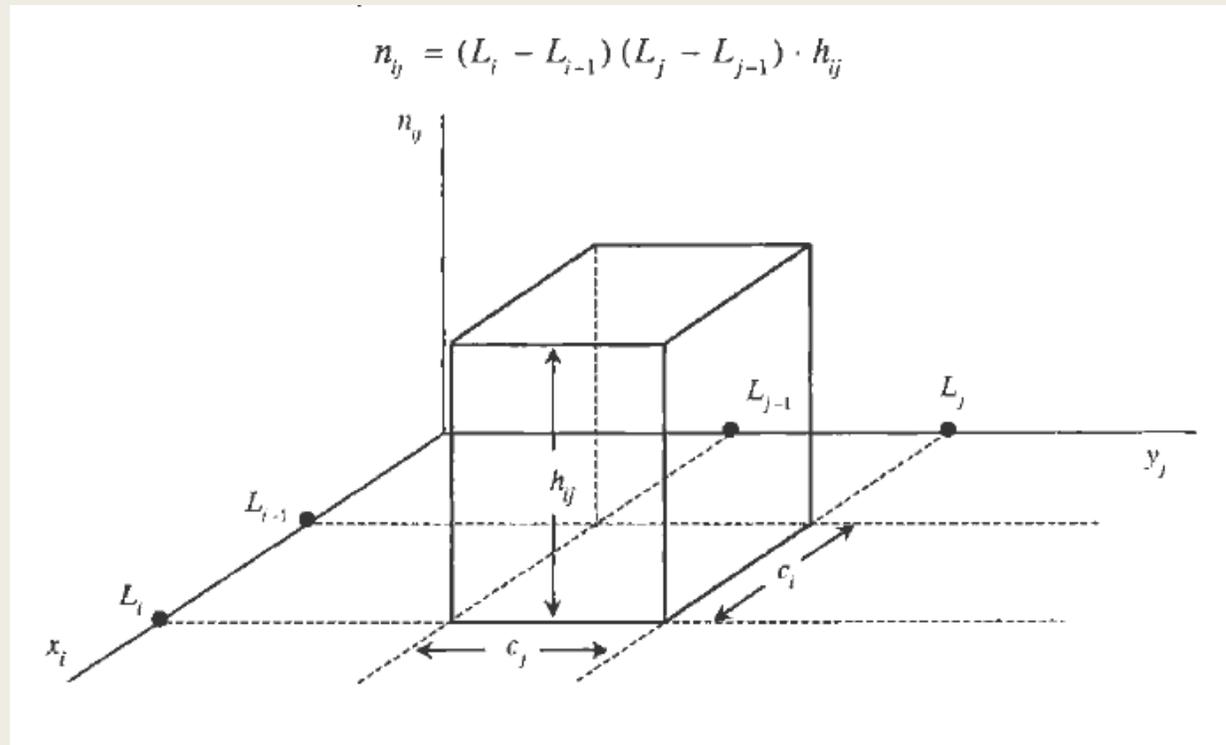
Nube de puntos o diagrama de dispersión



Se puede realizar también con datos agrupados utilizando la marca de clase.

2. Representaciones gráficas

Escalograma (datos agrupados)



Apéndice: momentos en distribuciones bidimensionales

Respecto al origen

$$r = 1, s = 0 \rightarrow a_{10} = \bar{x}$$

$$r = 0, s = 1 \rightarrow a_{01} = \bar{y}$$

$$a_{rs} = \sum_{i=1}^h \sum_{j=1}^k \frac{x_i^r y_j^s n_i}{N}$$

Respecto a la media

$$m_{11} = \text{cov}(x, y)$$

$$m_{rs} = \sum_{i=1}^h \sum_{j=1}^k \frac{(x_i - \bar{x})^r (y_j - \bar{y})^s n_i}{N}$$

IMPORTANTE

Apéndice: momentos en distribuciones bidimensionales

Recordemos

- Los momentos respecto a la media y respecto al origen se diferencian en un cambio de escala.
 - *Los a_r tienen como origen el cero, y los m_r en \bar{x}*
- Los momentos respecto a la media se pueden calcular partiendo de los momentos respecto al origen aplicando el binomio de Newton.

$$m_{20} = a_{20} - a_{20}^2 V(x)$$

$$m_{02} = a_{02} - a_{02}^2 V(y)$$

$$m_{11} = a_{11} - a_{10}a_{01} COV(x,y)$$

IMPORTANTE

Textos recomendados

- Martín-Pliego, *Introducción a la Estadística Económica y Empresarial*, Editorial AC, 2011, 3ª Edición

Prácticas recomendadas

- Ejercicios: los veremos en el capítulo 9
- Ejercicios resueltos en clase
- Prácticas y recursos web (aula virtual)

¿Quieres saber más?

- Ver 7.3.4 y 7.3.5 en Martín-Pliego
- El Capítulo 8 (Interpolación y ajuste) no entra en el temario del curso 2016/2017, pero es muy interesante.

