UNIDAD 12: Estadística Bidimensional

EJERCICIOS Y ACTIVIDADES - PÁG. 238

1. Dadas dos variables estadísticas, hemos obtenido los datos que se resumen en la siguiente tabla:

X \ Y	2	4	6	8	10	12	Totales
10	7	10	12	17	13	6	•••
20	9	12	15	18	12	7	•••
30	16	12	11	8	4	3	•••
40	21	16	10	9	5	2	•••
Totales	•••	•••	•••	•••	•••	•••	•••

a) **Completa en tu cuaderno la columna y la fila de totales.**

X \ Y	2	4	6	8	10	12	Totales
10	7	10	12	17	13	6	65
20	9	12	15	18	12	7	73
30	16	12	11	8	4	3	54
40	21	16	10	9	5	2	63
Totales	53	50	48	52	34	18	255

b) **Calcula el tamaño de la población.** $N = 255$

c) **¿Qué valor tiene la frecuencia absoluta** $f_{2,3}$ **?** $f_{2,3} = 15$

EJERCICIOS Y ACTIVIDADES - PÁG. 239

2. Calcula la media aritmética y la varianza para las distribuciones marginales asociadas a la siguiente tabla:

X \ Y	10	20	30	40
100	12	8	6	4
200	16	14	11	9
300	18	20	22	10
400	21	23	16	10

La distribución marginal para X es:

x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$	$x_i^2 \cdot f_{i\cdot}$
100	30	3000	300000
200	50	10000	2000000
300	70	21000	6300000
400	70	28000	11200000
Total	220	62000	19800000

La media es $\overline{x} = \dfrac{62000}{220} = 281,81$ y la varianza: $\sigma_X{}^2 = \dfrac{\sum_{i=1}^{4} x_i{}^2 \cdot f_{i\cdot}}{N} - \overline{x}^2 = 10578,51$

La distribución marginal para Y es:

y_j	$f_{\cdot j}$	$y_j \cdot f_{\cdot j}$	$y_j{}^2 \cdot f_{\cdot j}$
10	67	670	67000
20	65	1300	26000
30	55	1650	49500
40	33	1320	52800
Total	220	4940	13500

La media es $\overline{y} = \dfrac{4940}{220} = 22,45$ y la varianza: $\sigma_Y{}^2 = \dfrac{\sum_{i=1}^{4} y_j{}^2 \cdot f_{\cdot j}}{N} - \overline{y}^2 = 109,43$

EJERCICIOS Y ACTIVIDADES - PÁG. 240

3. Una variable bidimensional se distribuye mediante la siguiente tabla:

X \ Y	**10**	**11**	**12**	**13**
10	12	18	25	14
20	16	21	24	16
30	12	15	18	20

Calcula:

a) **La media aritmética y la varianza para la distribución condicionada** $X / y = 12$

Calculamos la tabla de frecuencias para la distribución:

x_i	f_{i3}	$x_i \cdot f_{i3}$	$x_i{}^2 \cdot f_{i3}$
10	25	250	2500
20	24	480	9600
30	18	540	16200
Total	67	1270	28300

La media es $\overline{x} = \dfrac{1270}{67} = 18,9552$ y la varianza: $\sigma^2_{X/y=12} = \dfrac{\sum_{i=1}^{3} x_i{}^2 \cdot f_{i3}}{N} - \overline{x}^2 = 63,0876$

b) **La media aritmética y la varianza para la distribución condicionada** $Y / x = 20$

Calculamos la tabla de frecuencias para la distribución:

y_j	f_{2j}	$y_j \cdot f_{2j}$	$y_j^2 \cdot f_{2j}$
10	16	160	1600
11	21	231	2541
12	24	288	3456
13	16	208	2704
Total	77	887	10301

La media es $\overline{y} = \frac{887}{77} = 11,5195$ y la varianza: $\sigma^2_{Y/x=20} = \frac{\sum_{i=1}^{4} y_j^2 \cdot f_{2j}}{N} - \overline{y}^2 = 1,0808$

EJERCICIOS Y ACTIVIDADES - PÁG. 241

4. Calcula la covarianza de la siguiente distribución bidimensional:

X \ Y	**5**	**10**	**15**	**20**
10	8	12	16	4
20	21	9	8	2
30	18	7	0	0

Hallamos la media de las dos variables:

x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$
10	40	400
20	40	800
30	25	750
Total	105	1950

y_j	$f_{\cdot j}$	$y_j \cdot f_{\cdot j}$
5	47	235
10	28	280
15	24	360
20	6	120
Total	105	995

Las medias son: $\overline{x} = \frac{1950}{105} \approx 18,5714$ y $\overline{y} = \frac{995}{105} = 9,4762$.

Escribimos una doble entrada con el producto $f_{ij} \cdot x_i \cdot y_j$

X \ Y	**5**	**10**	**15**	**20**	**Totales**
10	400	1200	2400	800	4800
20	2100	1800	2400	800	7100
30	2700	2100	0	0	4800
Totales	5200	5100	4800	1600	16700

La covarianza es: $\sigma_{XY} = \dfrac{\sum_j \left(\sum_i f_{ij} \cdot x_i \right) \cdot y_j}{N} - \bar{x} \cdot \bar{y} = -16,9388$

EJERCICIOS Y ACTIVIDADES - PÁG. 242

5. Calcula el coeficiente de correlación lineal para la distribución bidimensional siguiente:

X \ Y	1	2	3	4
10	0	0	6	9
11	0	4	12	8
12	9	16	4	0
13	21	12	0	0

Calculamos las medias y las desviaciones típicas de las dos variables:

x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$	$x_i^2 \cdot f_{i\cdot}$
10	15	150	1500
11	24	264	2904
12	29	348	4176
13	33	429	5577
Total	101	1191	14157

y_j	$f_{\cdot j}$	$y_j \cdot f_{\cdot j}$	$y_j^2 \cdot f_{\cdot j}$
1	30	30	30
2	32	64	128
3	22	66	198
4	17	68	272
Total	101	228	628

Las medias son : $\bar{x} = \dfrac{1191}{101} \approx 11,7921$ y $\bar{y} = \dfrac{228}{101} = 2,2574$, mientras que las varianzas son: $\sigma_X{}^2 = 1,1152$ y $\sigma_Y{}^2 = 1,1219$ y las desviaciones típicas: $\sigma_X = 1,056$ y $\sigma_Y = 1,0592$.

Para la covarianza construimos una tabla con los productos: $f_{ij} \cdot x_i \cdot y_j$

X \ Y	1	2	3	4	Totales
10	0	0	180	360	540
11	0	88	396	352	836
12	108	384	144	0	636
13	273	312	0	0	585
Totales	381	784	720	712	2597

El valor de la covarianza es $\sigma_{XY} = \dfrac{\sum_j \left(\sum_i f_{ij} \cdot x_i \right) \cdot y_j}{N} - \bar{x} \cdot \bar{y} = -0,9069$ y, finalmente, el coeficiente de correlación lineal es: $r = \dfrac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = -0,8108$

EJERCICIOS Y ACTIVIDADES - PÁG. 243

6. Calcula la recta de regresión de Y sobre X y la de X sobre Y de la siguiente distribución:

a) **Estima el valor de Y para X igual a 80**

b) **Estima el valor de X para Y igual a 10**

X \ Y	1	2	3	4	5	6
10	0	0	0	0	3	5
20	0	0	0	0	4	4
30	0	0	3	3	4	7
40	0	3	1	5	3	0
50	3	4	5	4	3	0
60	4	2	3	2	0	0
70	3	1	1	0	0	0

<u>RECTA DE REGRESIÓN DE Y SOBRE X</u>: $y-\bar{y}=\frac{\sigma_{XY}}{\sigma_X^2}\left(x-\bar{x}\right)$

Calculamos las medias, las desviaciones típicas y la covarianza:

x_i	$f_{i\cdot}$	$x_i\cdot f_{i\cdot}$	$x_i^2\cdot f_{i\cdot}$
10	8	80	800
20	8	160	3200
30	17	510	15300
40	12	480	19200
50	19	950	47500
60	11	660	39600
70	5	350	24500
Total	80	3190	150100

y_j	$f_{\cdot j}$	$y_j\cdot f_{\cdot j}$	$y_j^2\cdot f_{\cdot j}$
1	10	10	10
2	10	20	40
3	13	39	117
4	14	56	224
5	17	85	425
6	16	96	576
Total	80	306	1392

Las medias son: $\bar{x}=\frac{3190}{80}=39{,}875$ y $\bar{y}=\frac{306}{80}=3{,}825$, mientras que las varianzas son: $\sigma_X{}^2=286{,}234$ y $\sigma_Y{}^2=2{,}7694$ y las desviaciones típicas: $\sigma_X=16{,}9185$ y $\sigma_Y=1{,}6641$.

Para la covarianza construimos una tabla con los productos: $f_{ij}\cdot x_i\cdot y_j$

X \ Y	1	2	3	4	5	6	Totales
10	0	0	0	0	150	300	450
20	0	0	0	0	400	480	880
30	0	0	270	360	600	1260	2490
40	0	240	120	800	600	0	1760
50	150	400	750	800	750	0	2850
60	240	240	540	480	0	0	1500
70	210	140	210	0	0	0	560

Totales	600	1020	1890	2440	2500	2040	10490

El valor de la covarianza es $\sigma_{XY}=\dfrac{\sum_j\left(\sum_i f_{ij}\cdot x_i\right)\cdot y_j}{N}-\bar{x}\cdot\bar{y}=-21,3969$ y $\dfrac{\sigma_{XY}}{\sigma_X^2}=-0,075$

Por tanto, la recta de regresión de Y sobre X es: $y-3,825=-0,075\left(x-39,875\right)$

RECTA DE REGRESIÓN DE X SOBRE Y: $x-\bar{x}=\dfrac{\sigma_{XY}}{\sigma_Y^2}\left(y-\bar{y}\right)$.

En este caso solo falta calcular $\dfrac{\sigma_{XY}}{\sigma_Y^2}=-7,7262$, por tanto la ecuación de la recta de regresión de X sobre Y es: $x-39,875=-7,7262\left(y-3,825\right)$

Estimación del valor de Y para X igual a 80: $y-3,825=-0,075\left(80-39,875\right)\Rightarrow y=-26,6689$

Estimación del valor de X para Y igual a 10: $x-39,875=-7,7262\left(10-3,825\right)\Rightarrow x=-7,8346$

EJERCICIOS Y ACTIVIDADES DE RECAPITULACIÓN - PÁGS. 246-248

VARIABLES ESTADÍSTICAS BIDIMENSIONALES. DISTRIBUCIONES MARGINALES

1. La siguiente tabla resume los datos obtenidos de una variable estadística bidimensional:

X \ Y	10	20	30	40
10	0	3	4	4
11	4	5	3	6
12	3	4	5	0
13	7	6	1	0

a) **Determina las frecuencias absolutas para las distribuciones marginales**

x_i	$f_{i\cdot}$	$x_i\cdot f_{i\cdot}$	$x_i^2\cdot f_{i\cdot}$
10	11	110	1100
11	18	198	2178
12	12	144	1728
13	14	182	2366
Total	55	634	7372

y_j	$f_{\cdot j}$	$y_j\cdot f_{\cdot j}$	$y_j^2\cdot f_{\cdot j}$
10	14	140	1400
20	18	360	7200
30	13	390	11700
40	10	400	16000
Total	55	1290	36300

b) **Calcula la media aritmética de las distribuciones marginales**

Las medias son: $\bar{x}=\dfrac{634}{55}=11,527$ y $\bar{y}=\dfrac{1290}{55}=23,45$.

c) **Calcula la varianza de las distribuciones marginales.**

Las varianzas son: $\sigma_X{}^2=1,15835$ y $\sigma_Y{}^2=109,8843$.

d) **Calcula la desviación típica de las distribuciones marginales.**

Las desviaciones típicas son: $\sigma_X = 1,0763$ y $\sigma_Y = 10,4826$.

e) **Calcula la mediana de la distribución marginal correspondiente a la variable X.**
Construimos la tabla marginal con las frecuencias acumuladas:

x_i	$f_{i\cdot}$	$F_{i\cdot}$
10	11	11
11	18	**29**
12	12	41
13	14	55
Total	55	

La mediana es $Me = 11$

2. Dada la siguiente distribución bidimensional, calcula la media y la varianza de las distribuciones marginales:

X \ Y	[0,8)	[8,16)	[16,24)
10	4	9	3
20	5	3	8
30	8	7	3
40	10	3	4

Construimos las tablas de las distribuciones marginales:

x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$	$x_i^2 \cdot f_{i\cdot}$
10	16	160	1600
20	16	320	6400
30	18	540	16200
40	17	680	27200
Total	67	1700	51400

I_j	y_j	$f_{\cdot j}$	$y_j \cdot f_{\cdot j}$	$y_j^2 \cdot f_{\cdot j}$
[0,8)	4	27	108	432
[8,16)	12	22	264	3168
[16,24)	20	18	360	7200
	Total:	67	732	10800

Las medias son: $\bar{x} = \frac{1700}{67} = 25,3731$ y $\bar{y} = \frac{732}{67} = 10,9254$, mientras que las varianzas son: $\sigma_X^2 = 123,3682$ y $\sigma_Y^2 = 41,8303$

3. La siguiente tabla resume los datos obtenidos de una variable estadística bidimensional:

X \ Y	[0, 6)	[6, 12)	[12, 18)	[18, 24)
10	0	3	6	0
20	0	3	5	6
30	8	6	10	8
40	3	4	9	12

a) **Determina las frecuencias absolutas para las distribuciones marginales.**
Obtenemos la tabla de frecuencias con los datos necesarios para este y el resto de apartados:

x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$	$x_i^2 \cdot f_{i\cdot}$
10	9	90	900
20	14	280	5600
30	32	960	28800
40	28	1120	44800
Total	83	2450	80100

I_j	y_j	$f_{\cdot j}$	$y_j \cdot f_{\cdot j}$	$y_j^2 \cdot f_{\cdot j}$
[0, 6)	3	11	33	99
[6, 12)	9	16	144	1296
[12, 18)	15	30	450	6750
[18, 24)	20	26	520	10400
	Total:	83	1147	18545

b) **Calcula la media aritmética de las distribuciones marginales.**

Las medias son: $\overline{x} = \dfrac{2450}{83} = 29,5181$ y $\overline{y} = \dfrac{1147}{83} = 13,8193$.

c) **Calcula la varianza de las distribuciones marginales.**

Las varianzas son: ${\sigma_X}^2 = 93,7436$ y ${\sigma_Y}^2 = 32,4613$.

d) **Calcula las desviaciones típicas de las distribuciones marginales.**

Las desviaciones típicas son: $\sigma_X = 9,6821$ y $\sigma_Y = 5,6975$.

e) **Calcula la mediana de la distribución marginal correspondiente a la variable Y.**

Construimos la tabla marginal con las frecuencias acumuladas:

I_j	y_j	$f_{\cdot j}$	$F_{\cdot j}$
[0, 6)	3	11	11
[6, 12)	9	16	27
[12, 18)	15	30	**57**
[18, 24)	20	26	83
	Total:	83	

La mediana se encuentra en el intervalo [12, 18) y se calcula:

$$\frac{Me-12}{41,5-27} = \frac{18-12}{57-27} \Rightarrow Me = 12 + \frac{14,5 \cdot 6}{30} = 14,9$$

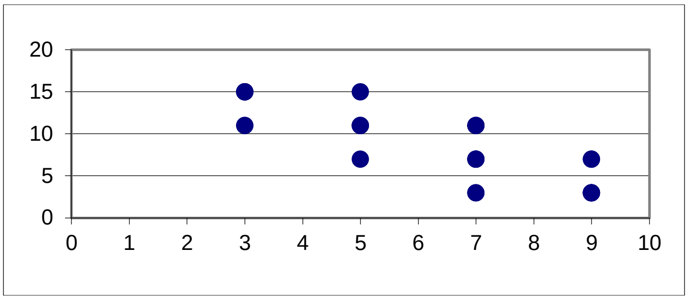
4. La siguiente tabla resume los datos de una variable bidimensional:

X \ Y	[1, 5)	[5, 9)	[9, 13)	[13, 17)
3	0	0	2	4
5	0	1	3	1
7	2	3	3	0
9	5	2	0	0

a) **Haz una gráfica con una nube de puntos**

Para realizar la gráfica de nube de puntos necesitamos los 26 puntos:

x	3	3	3	3	3	3	5	5	5	5	5	7	7	7	7	7	7	7	7	9	9	9	9	9	9	9
y	11	11	15	15	15	15	7	11	11	11	15	3	3	7	7	7	11	11	11	3	3	3	3	3	7	7



b) **Determina las frecuencias absolutas para las distribuciones marginales.**

x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$	$x_i^2 \cdot f_{i\cdot}$
10	6	60	600
20	5	100	2000
30	8	240	7200
40	7	280	11200
Total	26	680	21000

I_j	y_j	$f_{\cdot j}$	$y_j \cdot f_{\cdot j}$	$y_j^2 \cdot f_{\cdot j}$
[1, 5)	3	7	21	63
[5, 9)	7	6	42	294
[9, 13)	11	8	88	968
[13, 17)	15	5	75	1125
Total:		26	226	2450

c) **Calcula la media aritmética de las distribuciones marginales.**

Las medias son: $\bar{x} = \dfrac{680}{26} = 26{,}1538$ y $\bar{y} = \dfrac{1147}{83} = 8{,}6923$.

d) **Calcula la varianza de las distribuciones marginales.**

Las varianzas son: $\sigma_X^{\ 2} = 123{,}6686$ y $\sigma_Y^{\ 2} = 18{,}6746$.

e) **Calcula la desviación típica de las distribuciones marginales.**

Las desviaciones típicas son: $\sigma_X = 11{,}1206$ y $\sigma_Y = 4{,}3214$.

f) **Calcula la mediana de la distribución marginal correspondiente a la variable Y.**
Construimos la tabla marginal con las frecuencias acumuladas:

I_j	y_j	$f_{\cdot j}$	$F_{\cdot j}$
[1, 5)	3	7	7
[5, 9)	7	6	**13**
[9, 13)	11	8	21
[13, 17)	15	5	26
Total:		26	

La mediana es el extremo superior del intervalo [9, 13):
$Me = 9$

DISTRIBUCIONES CONDICIONADAS

5. La siguiente tabla corresponde al estudio de una variable bidimensional:

X \ Y	**5**	**10**	**15**	**20**
2	5	3	5	1
3	2	4	4	3
4	6	6	5	4

a) **Calcula la media aritmética de la distribución $X / y = 10$**

Construimos la tabla de frecuencias de la distribución $X\ /\ y = 10$

x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$
10	3	30
20	4	80
30	6	180
Total	13	290

La media es $\bar{x}_{y=10} = \dfrac{290}{13} = 22,3077$.

b) **Calcula la media aritmética de la distribución $Y / x = 3$.**

Construimos la tabla de frecuencias de la distribución $Y / x = 3$:

y_j	$f_{\cdot j}$	$y_j \cdot f_{\cdot j}$
5	2	10
10	4	40
15	4	60
20	3	60
Total:	13	170

La media es $\bar{x}_{y=10} = \dfrac{290}{13} = 22,3077$

6. Dada la siguiente distribución bidimensional, calcula la media y la varianza de la distribución condicionada $X\ /\ y = 11$:

Y / **X**	**11**	**12**	**13**
5	5	6	1
6	7	8	2
7	8	9	3
8	2	3	5

Calculamos la tabla de frecuencias de la distribución condicionada:

x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$	$x_i^2 \cdot f_{i\cdot}$
5	5	25	125
6	7	42	252
7	8	56	392
8	2	16	128
Total	22	139	897

La media es $\bar{x} = \dfrac{139}{22} = 6,318$ y la varianza: $\sigma^2_{X/y=11} = 0,8533$

7. Dada la tabla:

Y / X	[1, 5)	[5, 9)	[9, 13)	[13, 17)
[0, 2)	1	2	8	2
[2, 4)	3	5	3	4
[4, 6)	4	7	4	6
[6, 8)	7	8	7	1
[8, 10)	8	4	6	6
[10,12)	2	2	5	8
[12, 14)	1	1	3	2

a) **Calcula la media aritmética y la varianza de la distribución** $Y / x = [2,4)$

Construimos la tabla de frecuencias de la distribución condicionada:

I_j	y_j	$f_{\cdot j}$	$y_j \cdot f_{\cdot j}$	$y_j^2 \cdot f_{\cdot j}$
[1, 5)	3	3	9	27
[5, 9)	7	5	35	245
[9, 13)	11	3	33	363
[13, 17)	15	4	60	900
Total:		15	137	1535

La media es $\overline{y}_{x=[2,4)} = \dfrac{137}{15} = 9,1\widehat{3}$ y la varianza $\sigma^2_{Y/x=[2,4)} = 18,9156$

b) **Calcula la media aritmética y la varianza de la distribución** $X / y = [5,9)$

Construimos la tabla de frecuencias de la distribución condicionada:

I_i	x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$	$x_i^2 \cdot f_{i\cdot}$
[0, 2)	1	2	2	2
[2, 4)	3	5	15	45
[4, 6)	5	7	35	175
[6, 8)	7	8	56	392
[8, 10)	9	4	36	324
[10,12)	11	2	22	242
[12, 14)	13	1	13	169
Total		29	179	1349

La media es $\overline{x}_{y=[5,9)} = \dfrac{179}{29} = 6,1724$ y la varianza $\sigma^2_{X/y=[5,9)} = 8,4185$

COVARIANZA

8. Calcula la covarianza de la siguiente distribución:

X \ Y	10	15	20
3	2	5	1
5	4	3	2
7	3	4	5
9	1	2	7

Calculamos las medias de las distribuciones marginales:

x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$
3	8	24
5	9	45
7	12	84
9	10	90
Total	39	243

y_j	$f_{\cdot j}$	$y_j \cdot f_{\cdot j}$
10	10	100
15	14	210
20	15	300
Total:	39	610

Las medias son: $\bar{x} = \dfrac{243}{39} = 6,2308$ y $\bar{y} = \dfrac{610}{39} = 15,641$.

Para la covarianza construimos una tabla con los productos $f_{ij} \cdot x_i \cdot y_j$:

X \ Y	10	15	20	Totales
3	60	225	60	345
5	200	225	200	625
7	210	420	700	1330
9	90	270	1260	1620
Totales	560	1140	2220	3920

Por tanto, la covarianza es $\sigma_{XY} = \dfrac{\sum_j \left(\sum_i f_{ij} \cdot x_i \right) \cdot y_j}{N} - \bar{x} \cdot \bar{y} = 3,0572$

9. Calcula la covarianza de la siguiente distribución:

X \ Y	5	6	7
10	1	3	1
11	3	4	0
12	7	5	6
13	4	2	6

Construimos la tabla de frecuencias de las distribuciones marginales para calcular sus medias:

x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$
10	5	50
11	7	77
12	18	216
13	12	156
Total	42	499

y_j	$f_{\cdot j}$	$y_j \cdot f_{\cdot j}$
5	15	75
6	14	84
7	13	91
Total:	42	250

Las medias son: $\overline{x} = \dfrac{499}{42} = 11{,}881$ y $\overline{y} = \dfrac{250}{42} = 5{,}9524$.

Para la covarianza construimos una tabla con los productos $f_{ij} \cdot x_i \cdot y_j$:

X \ Y	**5**	**6**	**7**	**Totales**
10	50	180	70	300
11	165	264	0	429
12	420	360	504	1284
13	260	156	546	962
Totales	895	960	1120	2975

Por tanto, la covarianza es $\sigma_{XY} = \dfrac{\sum_j \left(\sum_i f_{ij} \cdot x_i \right) \cdot y_j}{N} - \overline{x} \cdot \overline{y} = 0{,}1134$

10. Calcula la covarianza de la siguiente distribución:

X \ Y	**[5, 7)**	**[7, 9)**	**[9, 11)**
10	2	5	8
20	4	7	4
30	5	4	2
40	9	3	1

Construimos las tablas de frecuencias de las distribuciones marginales para calcular las medias:

x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$
10	15	150
20	15	300
30	11	330
40	13	520

Total	54	1300

I_j	y_j	$f_{\cdot j}$	$y_j \cdot f_{\cdot j}$
[5, 7)	6	20	120
[7, 9)	8	19	152
[9, 11)	10	15	150
Total:		54	422

Las medias son: $\bar{x} = \dfrac{1300}{54} = 24,0741$ y $\bar{y} = \dfrac{422}{54} = 7,8148$.

Para la covarianza construimos una tabla con los productos $f_{ij} \cdot x_i \cdot y_j$:

X \ Y	6	8	10	Totales
10	120	400	800	1320
20	480	1120	800	2400
30	900	960	600	2460
40	2160	960	400	3520
Totales	3660	3440	2600	9700

Por tanto, la covarianza es $\sigma_{XY} = \dfrac{\sum_j \left(\sum_i f_{ij} \cdot x_i \right) \cdot y_j}{N} - \bar{x} \cdot \bar{y} = -8,5048$

COEFICIENTE DE CORRELACIÓN. RECTAS DE REGRESIÓN

11. Calcula la covarianza y el coeficiente de correlación lineal de la siguiente distribución:

X \ Y	[10, 15)	[15, 20)	[20, 25)
5	5	7	9
6	8	4	2
7	10	7	1
8	15	10	0

Construimos las tablas de frecuencias de las distribuciones marginales para calcular sus medias y sus desviaciones típicas:

x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$	$x_i^2 \cdot f_{i\cdot}$
5	21	105	525
6	14	84	504
7	18	126	882
8	25	200	1600
Total	78	515	3511

I_j	y_j	$f_{\cdot j}$	$y_j \cdot f_{\cdot j}$	$y_j^2 \cdot f_{\cdot j}$
[10, 15)	12,5	38	475	5937,5
[15, 20)	17,5	28	490	8575
[20, 25)	22,5	12	270	6075
Total:		78	1235	20587,5

Las medias son $\bar{x} = \frac{515}{78} = 6,6026$ y $\bar{y} = \frac{1235}{78} = 15,8\hat{3}$, las varianzas: $\sigma_X{}^2 = 1,419$ y $\sigma_Y{}^2 = 13,2479$ y las desviaciones típicas: $\sigma_X = 1,1912$ y $\sigma_Y = 3,6398$.

Para la covarianza construimos una tabla con los productos: $f_{ij} \cdot x_i \cdot y_j$

X \ Y	12,5	17,5	22,5	Totales
5	312,5	612,5	1012,5	1937,5
6	600	420	270	1290
7	875	857,5	157,5	1890
8	1500	1400	0	2900
Totales	3287,5	3290	1440	8017,5

El valor de la covarianza es $\sigma_{XY} = \frac{\sum_j \left(\sum_i f_{ij} \cdot x_i \right) \cdot y_j}{N} - \bar{x} \cdot \bar{y} = -1,7521$ y, finalmente, el coeficiente de correlación lineal es: $r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = -0,4041$

12. En el estudio de cierta variable bidimensional se han obtenido los siguientes datos: $\bar{x} = 35,5$, $\sigma_X = 1,2$, $\bar{y} = 135$, $\sigma_Y = 5,3$, $\sigma_{XY} = 5,36$.

a) **Calcula el coeficiente de correlación lineal.**

$$r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = 0,8428$$

b) **Determina la ecuación de la recta de regresión de Y sobre X**

La ecuación de la recta es: $y - \bar{y} = \frac{\sigma_{XY}}{\sigma_X^2}\left(x - \bar{x}\right)$, por tanto: $y - 135 = 3,7\hat{2}\left(x - 35,5\right)$

c) **Determina la ecuación de la recta de regresión de X sobre Y**

La ecuación de la recta es: $x - \bar{x} = \frac{\sigma_{XY}}{\sigma_Y^2}\left(y - \bar{y}\right)$, por tanto: $x - 35,5 = 0,1908\left(y - 135\right)$

d) **Estima el valor de la variable X para Y igual a 20.**

Utilizando la recta de regresión de X sobre Y, tenemos:

$x = 35,5 + 0,1908\left(20 - 135\right) = 13,0562$

13. En un instituto se realiza un estudio para determinar la relación entre el número de horas dedicadas a ver la televisión (variable *X*) y el número de asignaturas suspensas (variable *Y*), y se obtienen los datos que se resumen a continuación.

X \ Y	0	1	2	3	4	5	6	7
1	12	7	0	0	0	0	0	0

2	15	11	6	0	0	0	0	0
3	7	4	2	6	7	8	0	0
4	0	0	3	7	9	9	10	15
5	0	0	1	6	7	7	17	21

a) **Calcula la media aritmética de las distribuciones marginales.**
Construimos la tabla de frecuencias de las distribuciones marginales:

x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$	$x_i^2 \cdot f_{i\cdot}$
1	19	19	19
2	32	64	128
3	34	102	306
4	53	212	848
5	59	295	1475
Total	197	692	2776

y_j	$f_{\cdot j}$	$y_j \cdot f_{\cdot j}$	$y_j^2 \cdot f_{\cdot j}$
0	34	0	0
1	22	22	22
2	12	24	48
3	19	57	171
4	23	92	368
5	24	120	600
6	27	162	972
7	36	252	1764
Total:	197	729	3945

Las medias son: $\bar{x} = \frac{692}{197} = 3,5127$ y $\bar{y} = \frac{729}{197} = 3,7005$

b) **Calcula la varianza y la desviación típica de las distribuciones marginales.**
Las varianzas son: $\sigma_X{}^2 = 1,7524$ y $\sigma_Y{}^2 = 6,3316$ y las desviaciones típicas: $\sigma_X = 1,3238$ y $\sigma_Y = 2,5163$.

c) **Calcula la covarianza de la variable bidimensional.**
Para la covarianza construimos una tabla con los productos $f_{ij} \cdot x_i \cdot y_j$:

X \ Y	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Totales**
1	0	7	0	0	0	0	0	0	7
2	0	22	24	0	0	0	0	0	46
3	0	12	12	54	84	120	0	0	282
4	0	0	24	84	144	180	240	420	1092
5	0	0	10	90	140	175	510	735	1660
Totales	0	41	70	228	368	475	750	1155	3087

El valor de la covarianza es $\sigma_{XY} = \frac{\sum_j \left(\sum_i f_{ij} \cdot x_i \right) \cdot y_j}{N} - \bar{x} \cdot \bar{y} = 2,6713$

d) **Calcula las rectas de regresión** X/Y **e** Y/X.
Recta de regresión X/Y

La ecuación de la recta es: $x - \bar{x} = \frac{\sigma_{XY}}{\sigma_Y^2}\left(y - \bar{y}\right)$, por tanto: $x - 3,5127 = 1,5244\left(y - 3,7005\right)$

Recta de regresión Y/X

La ecuación de la recta es: $y-\bar{y}=\frac{\sigma_{XY}}{\sigma_X^2}\left(x-\bar{x}\right)$, por tanto: $y-3,7005=0,4219\left(x-3,5127\right)$

14. La Dirección General de Tráfico realiza un estudio para determinar la relación existente entre el consumo de alcohol y los errores que se cometen al volante que pueden ocasionar accidentes. Escogido un grupo de conductores al azar, lo someten a 6 pruebas de conducción, primero sin tomar nada, luego una copa, después dos y finalmente 3, y se obtienen los siguientes resultados. La variable *X* es el número de copas consumidas y la variable *Y*, el número de pruebas superadas.

X \ Y	0	1	2	3	4	5	6
0	0	0	0	0	0	18	23
1	0	0	0	3	7	15	15
2	10	23	7	0	0	0	0
3	25	16	0	0	0	0	0

a) **Calcula la media aritmética de las distribuciones marginales.**
Construimos la tabla de frecuencia de las distribuciones marginales:

x_i	$f_{i\cdot}$	$x_i\cdot f_{i\cdot}$	$x_i^2\cdot f_{i\cdot}$
0	41	0	0
1	40	40	40
2	40	80	160
3	41	123	369
Total	162	243	569

y_j	$f_{\cdot j}$	$y_j\cdot f_{\cdot j}$	$y_j^2\cdot f_{\cdot j}$
0	35	0	0
1	39	39	39
2	7	14	28
3	3	9	27
4	7	28	112
5	33	165	825
6	38	228	1368
Total:	162	483	2399

Las medias son $\bar{x}=\frac{243}{162}=1,5$ y $\bar{y}=\frac{483}{162}=2,9815$

b) **Calcula la varianza y la desviación típica de las distribuciones marginales.**
Las varianzas son: $\sigma_X{}^2=1,2623$ y $\sigma_Y{}^2=5,9194$ y las desviaciones típicas: $\sigma_X=1,1235$ y $\sigma_Y=2,433$.

c) **Calcula la covarianza de la variable bidimensional.**
Para la covarianza construimos una tabla con los productos $f_{ij}\cdot x_i\cdot y_j$:

X \ Y	0	1	2	3	4	5	6	Totales
0	0	0	0	0	0	0	0	0
1	0	0	0	9	28	75	90	202
2	0	46	28	0	0	0	0	74
3	0	48	0	0	0	0	0	48
Totales	0	94	28	9	28	75	90	324

El valor de la covarianza es $\sigma_{XY} = \dfrac{\sum_j \left(\sum_i f_{ij} \cdot x_i \right) \cdot y_j}{N} - \overline{x} \cdot \overline{y} = -2,4722$

d) **Calcula e interpreta el coeficiente de correlación lineal.**

El coeficiente de correlación lineal es: $r = \dfrac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = -0,9044$.

Es un valor negativo (esto es, la pendiente de la recta es negativa: cuanto mayor sea la variable X menor será la variable Y) y cercano al -1, por lo que la aproximación por una recta es buena. El estudio concluye que hay correlación lineal negativa entre el número de copas y el número de pruebas superadas.

e) **Calcula las rectas de regresión X/Y e Y/X.**

Recta de regresión X/Y

La ecuación de la recta es: $x - \overline{x} = \dfrac{\sigma_{XY}}{\sigma_Y^2}\left(y - \overline{y}\right)$, por tanto:

$$x - 1,5 = -0,4177\left(y - 2,9815\right)$$

Recta de regresión Y/X

La ecuación de la recta es: $y - \overline{y} = \dfrac{\sigma_{XY}}{\sigma_X^2}\left(x - \overline{x}\right)$, por tanto:

$$y - 2,9815 = -1,9584\left(x - 1,5\right)$$

f) **Predice los errores que cometería una persona que se tome 5 copas.**

Utilizamos la recta de regresión Y/X de para predecir el valor de Y si $x = 5$:

$$y = 2,9815 - 1,9584\left(5 - 1,5\right) = -3,8729$$

Este valor significa que no conseguiría superar ninguna prueba, esto es, cometería errores en las 6 pruebas.

15. En un periódico local se ha publicado un estudio estadístico de una variable estadística bidimensional. De los datos publicados se obtiene que $\sigma_X = 2,3$, $\sigma_Y = 5,4$ y $\sigma_{XY} = 15$. ¿Es posible el valor obtenido de la covarianza? Razona tu respuesta.

No es posible, ya que, si calculamos el valor del coeficiente de correlación lineal se obtiene: $r = \dfrac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = 1,21 > 1$

El coeficiente debe ser menor que 1 y, por tanto, hay un error en los datos publicados.

16. La ecuación de la recta de regresión X sobre Y de un estudio estadístico bidimensional es $x - 3,5 = 2\left(y - 5\right)$.

a) **Determina $\overline{x}$ e $\overline{y}$.**

Las medias son $\overline{x} = 3,5$ y $\overline{y} = 5$

b) **¿El coeficiente de correlación lineal es positivo? Razona tu respuesta.**

Sí, ya que si $\dfrac{\sigma_{XY}}{\sigma_Y^2} = 2$, entonces la covarianza es positiva y por tanto el coeficiente de correlación lineal también.

c) **¿Se puede estimar el valor de X para Y igual a 2? En caso afirmativo, estímalo.**
Si el valor 2 está entre los posibles valores de la variable Y en el contexto del estudio, se puede realizar la estimación, que será: $x = 3,5 + 2(2-5) = -2,5$

(En el estudio realizado puede que no tenga sentido un valor negativo de X, en cuyo caso habría que interpretar este resultado).

d) **Si $\sigma_Y = 0,35$, determina el valor de la covarianza.**

Puesto que $\dfrac{\sigma_{XY}}{\sigma_Y^2} = 2$, tenemos que:

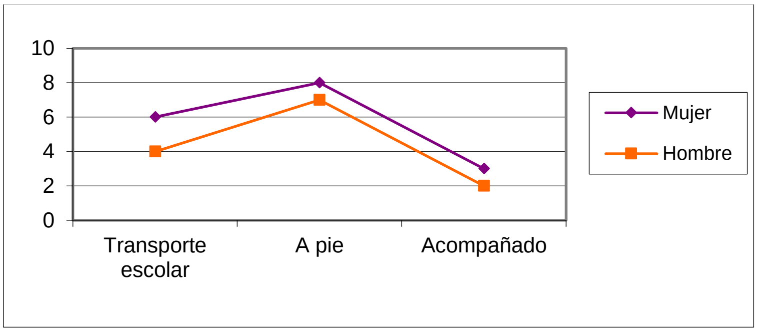
$$\sigma_{XY} = 2\sigma_Y^2 = 2 \cdot 0,35^2 = 0,245$$

PROBLEMAS

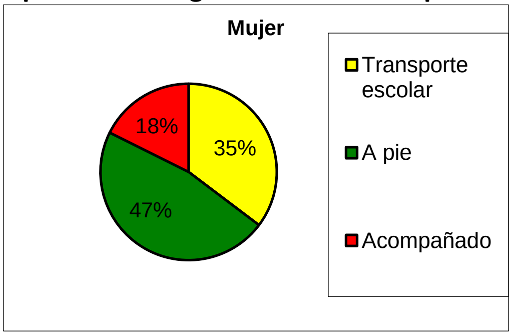
17. En mi clase hay compañeros que usan el transporte escolar, otros a los que los traen los padres con su vehículo y otros que vienen andando. La siguiente tabla muestra los datos según el sexo:

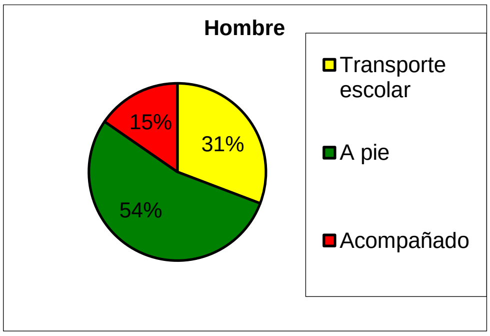
X \ Y	Mujer	Hombre
Transporte escolar	6	4
A pie	8	7
Acompañado	3	2

a) **Dibuja el polígono de frecuencias para la variable X para cada sexo. Utiliza el mismo eje para tal gráfica.**



b) **Representa una gráfica de sectores para cada sexo.**





18. En mi instituto se realiza un estudio para determinar la relación entre el número de asignaturas suspensas y las horas de dedicación al estudio:

X \ Y	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	15	17
1	0	0	0	8	12	3	0	0
2	0	12	9	2	0	0	0	0
3	18	16	0	0	0	0	0	0
4	25	5	0	0	0	0	0	0

a) **Calcula la media aritmética de las distribuciones marginales.**

Construimos las tablas de frecuencia de las distribuciones marginales *(en la siguiente página)*

Las medias son: $\overline{x} = \dfrac{291}{142} = 2,0493$ e $\overline{y} = \dfrac{353}{142} = 2,4859$

x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$	$x_i^2 \cdot f_{i\cdot}$
0	32	0	0
1	23	23	23
2	23	46	92
3	34	102	306
4	30	120	480
Total	142	291	901

y_j	$f_{\cdot j}$	$y_j \cdot f_{\cdot j}$	$y_j^2 \cdot f_{\cdot j}$
0	43	0	0
1	33	33	33
2	9	18	36
3	10	30	90
4	12	48	192
5	3	15	75
6	15	90	540
7	17	119	833
Total:	142	353	1799

b) **Calcula la varianza y la distribución típica de las distribuciones marginales.**

Las varianzas son: $\sigma_X^2 = 2,1455$ y $\sigma_Y^2 = 6,4892$

Las desviaciones típicas: $\sigma_X = 1,4647$ y $\sigma_Y = 2,5474$.

c) **Calcula la covarianza de la variable bidimensional.**

Para la covarianza construimos una tabla con los productos $f_{ij} \cdot x_i \cdot y_j$:

X \ Y	0	1	2	3	4	5	6	7	Totales
0	0	0	0	0	0	0	0	0	0
1	0	0	0	24	48	15	0	0	87
2	0	24	36	12	0	0	0	0	72
3	0	48	0	0	0	0	0	0	48
4	0	20	0	0	0	0	0	0	20
Totales	0	92	36	36	48	15	0	0	227

El valor de la covarianza es $\sigma_{XY} = \dfrac{\sum\limits_j \left(\sum\limits_i f_{ij} \cdot x_i \right) \cdot y_j}{N} - \overline{x} \cdot \overline{y} = -3,4958$

d) **Calcula e interpreta el coeficiente de correlación lineal.**

El coeficiente de correlación lineal es: $r = \dfrac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = -0,9369$.

Es un valor negativo (esto es, la pendiente de la recta es negativa: cuanto mayor sea la variable X menor será la variable Y y viceversa) y cercano al -1, por lo que la aproximación por una recta es buena. El estudio concluye que hay correlación lineal negativa entre el número de asignaturas suspensas y las horas dedicadas al estudio.

19. Para probar la eficacia de un medicamento, se hacen unas pruebas de evolución de una enfermedad en pacientes que han tomado distintas dosis y se observan los días que tarda el paciente en recuperarse de la enfermedad.
La variable *X* es el número de dosis del medicamento que consume el paciente e *Y* es el número de días que tarda el paciente en curarse.

X \ Y	1	2	3	4	5	6	7
0	0	0	0	0	0	8	12
1	0	0	0	0	6	12	0
2	0	0	0	6	7	0	0
3	12	21	18	0	0	0	0
4	17	15	0	0	0	0	0

a) **Calcula la media aritmética de las distribuciones marginales.**
Construimos una tabla de frecuencias para las distribuciones marginales:

x_i	$f_{i\cdot}$	$x_i \cdot f_{i\cdot}$	$x_i^2 \cdot f_{i\cdot}$
0	20	0	0
1	18	18	18
2	13	26	52
3	51	153	459
4	32	128	512
Total	134	325	1041

y_j	$f_{\cdot j}$	$y_j \cdot f_{\cdot j}$	$y_j^2 \cdot f_{\cdot j}$
1	29	29	29
2	36	72	144
3	18	54	162
4	6	24	96
5	13	65	325
6	20	120	720
7	12	84	588
Total:	134	448	2064

Las medias son: $\bar{x} = \dfrac{325}{134} = 2,04254$ e $\bar{y} = \dfrac{448}{134} = 3,3433$

b) **Calcula la varianza y la desviación típica de las distribuciones marginales.**

Las varianzas son: $\sigma_X^{\ 2} = 1,8862$ y $\sigma_Y^{\ 2} = 4,2254$ y las desviaciones típicas: $\sigma_X = 1,3734$ y $\sigma_Y = 2,0556$.

c) **Calcula la covarianza de la variable bidimensional.**

Para la covarianza construimos una tabla con los productos $f_{ij} \cdot x_i \cdot y_j$:

X \ Y	1	2	3	4	5	6	7	Totales
0	0	0	0	0	0	0	0	0

1	0	0	0	0	30	72	0	102
2	0	0	0	48	70	0	0	118
3	36	126	162	0	0	0	0	324
4	68	120	0	0	0	0	0	188
Totales	104	246	162	48	100	72	0	732

El valor de la covarianza es $\sigma_{XY} = \dfrac{\sum_j \left(\sum_i f_{ij} \cdot x_i \right) \cdot y_j}{N} - \bar{x} \cdot \bar{y} = -2,646$

d) **Calcula e interpreta el coeficiente de correlación lineal.**

El coeficiente de correlación lineal es: $r = \dfrac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = -0,9373$.

Es un valor negativo (esto es, la pendiente de la recta es negativa: cuanto mayor sea la variable X menor será la variable Y y viceversa) y cercano al -1, por lo que la aproximación por una recta es buena.

El estudio concluye que hay correlación lineal negativa entre el número de dosis y los días que tarda el paciente en curarse.

e) **Calcula las rectas de regresión X/Y e Y/X.**

<u>Recta de regresión X/Y</u>

La ecuación de la recta es: $x - \bar{x} = \dfrac{\sigma_{XY}}{\sigma_Y^2}\left(y - \bar{y}\right)$, por tanto:

$$x - 2,4254 = -0,6262\left(y - 3,3433\right)$$

<u>Recta de regresión Y/X</u>

La ecuación de la recta es: $y - \bar{y} = \dfrac{\sigma_{XY}}{\sigma_X^2}\left(x - \bar{x}\right)$, por tanto:

$$y - 3,3433 = -1,4028\left(x - 2,4254\right)$$

f) **A la vista de estos resultados, ¿consideras que es eficaz el medicamento? Razona tu respuesta.**

Es posible afirmar que el medicamento es bueno ya que hay una alta correlación lineal negativa entre el número de dosis y el tiempo que tarda el paciente en curarse.

(No obstante, en los estudios reales habría que realizar más pruebas para descartar que la causa de la mejora de los pacientes sea debida a otros factores ajenos al medicamento)

DESAFÍO PISA- PÁG. 238

¿CUÁNTAS PERSONAS VIVEN EN UN PISO?

En la localidad de Moradoria se ha realizado un estudio estadístico para establecer la relación existente entre el número de personas que habita una vivienda y los metros cuadrados de esta.

La siguiente tabla muestra los resultados de dicho estudio, en el que la variable *X* representa los metros cuadrados de la vivienda y la variable *Y* el número de personas que vive en dicha vivienda:

X \ Y	0	1	2	3	4	5	6	7	8
[0, 45)	3	14	8	5	0	0	0	0	0
[45, 50)	4	15	9	2	0	0	0	0	0
[50, 60)	6	10	12	7	0	0	0	0	0
[60, 70)	7	0	8	10	3	0	0	0	0
[70, 80)	4	0	0	7	8	9	0	0	0
[80, 90)	6	0	0	12	15	18	0	0	0
[90, 100)	7	0	0	0	8	15	18	6	0
[100, 120)	4	1	0	0	0	8	8	11	12

ACTIVIDAD 1. ¿Cuántas viviendas vacías hay en el municipio?

B: 41, ya que la suma de viviendas vacías es $f_{\cdot 1} = 41$

ACTIVIDAD 2. El número de habitantes de Moradoria es de:

C: 1000, ya que si multiplicamos los habitantes de cada casa por el número de ellas y luego sumamos se obtiene $\sum_{i,j} f_{ij} \cdot y_j = 1000$

ACTIVIDAD 3. El número de viviendas del municipio es:

B: 300, ya que $\sum_{i,j} f_{ij} = 300$

ACTIVIDAD 4. Los metros cuadrados de la vivienda media son, aproximadamente:

A: 75, ya que $\dfrac{\sum_i f_{i\cdot} \cdot x_i}{300} = 74,1\widehat{6}$

ACTIVIDAD 5. El tamaño de la vivienda vacía en metros cuadrados es, aproximadamente:

A: 73, ya que $\dfrac{\sum_i f_{i1} \cdot x_i}{41} \approx 72,13$