

## 4 Descripción conjunta de varias variables. Ejemplos y ejercicios.

### 4.1 Ejemplos.

**Ejemplo 4.1** La siguiente tabla de frecuencias absolutas corresponde a 200 observaciones de una variable bidimensional. La última fila contiene la distribución marginal de la variable  $Y$ , y la última columna contiene la distribución marginal de la variable  $X$ .

$X \setminus Y$	10	15	20	25	30	35
8	8	10	10	6	0	10
10	12	20	0	14	10	20
12	24	10	10	6	20	10

Calcular:

- Las distribuciones marginales.
- La distribución de  $X$  condicionada a que  $Y = 25$ .
- La distribución de  $Y$  condicionada a que  $X = 12$ .

**Ejemplo 4.2** Con los datos del ejemplo 4.1 calcular las medias, y varianzas marginales. ¿Cuál de las dos variables presenta mayor variación?

**Ejemplo 4.3** Preguntamos a dos hermanos sus preferencias sobre diferentes deportes. La siguiente tabla contiene los órdenes de preferencias dados por cada hermano:

deporte	$x_i$	$y_i$
tennis	1	7
fútbol	5	3
baloncesto	6	2
natación	2	6
waterpolo	3	5
voleibol	7	1
golf	4	4

Averigüad si los gustos deportivos de estos dos hermanos son iguales, opuestos o independientes.

## 4.2 Ejercicios.

**Ejercicio 4.1** La siguiente tabla muestra la distribución conjunta de frecuencias relativas de la variable  $X$ , que representa el número de tarjetas de crédito que posee una persona, y la variable  $Y$ , que refleja el número de compras semanales pagadas con tarjeta de crédito.

$X = \text{“Num. tarjetas”}$	$Y = \text{“Num. de compras por semana”}$				
	0	1	2	3	4
1	0.08	0.13	0.09	0.06	0.03
2	0.03	0.08	0.08	0.09	0.07
3	0.01	0.03	0.06	0.08	0.08

- Si se sabe que en el estudio han participado 300 personas, hallar la distribución conjunta de frecuencias absolutas.
- Hallar la distribución marginal de  $Y$ . Cuál es el número medio y la desviación típica del número de compras semanales pagadas con tarjeta de crédito?
- Obtener la distribución del número de tarjetas de crédito que poseen las personas de dicho estudio. Cuál es el número más frecuente de tarjetas de crédito que posee una de estas personas?
- Calcular la distribución del número de compras semanales pagadas con tarjetas de crédito que realizan las personas que poseen tres tarjetas. Cuál es la media de esta distribución?

**Ejercicio 4.2** Se han clasificado 100 familias según el número de hijos e hijas, en la siguiente tabla:

$H \setminus M$	0	1	2	3	4
0	4	6	9	4	1
1	5	10	7	4	2
2	7	8	5	3	1
3	5	5	3	2	1
4	2	3	2	1	0

- Hallar las medias, varianzas y desviaciones típicas marginales.
- Qué número medio de hijas hay en aquellas familias que tienen 2 hijos?
- Qué número medio de hijos hay en aquellas familias que no tienen hijas?
- Qué número medio de hijos tienen aquellas familias que a lo sumo tienen 2 hijas?

e) Hallar la covarianza.

**Ejercicio 4.3** Responde a las siguientes cuestiones:

- a) Supón que se miden dos variables y que la segunda siempre toma un valor más pequeño que la primera. Es verdad que el coeficiente de correlación lineal de Pearson será negativo? Pon un ejemplo.
- b) Supón que en las parejas de hermanos gemelos, uno siempre fuese un 2% más bajo que el otro. Cuál sería la correlación lineal entre las alturas de los dos hermanos? Justifica la respuesta.
- c) En general, qué se puede deducir si el coeficiente de correlación lineal entre dos variables es exactamente igual a cero?

**Ejercicio 4.4** El product manager de una empresa publicitaria sospecha que la campaña llevada a cabo durante el último mes para promocionar una nueva bebida refrescante no ha sido efectiva. Con el fin de verificar esta impresión, encarga que se encueste a 1000 personas y se les pida cuántas veces han visto el anuncio por televisión y cuántos litros del producto han consumido. La tabla siguiente muestra los resultados:

	Y = "veces que ha visto el anuncio"			
X = "litros"	0	1-3	4-6	7-10
0	314	230	104	77
(0,1]	75	35	28	12
(1,2]	45	21	7	2
(2,4]	16	13	6	0
(4,10]	0	1	5	9

- a) Qué porcentaje de personas no han probado el producto a pesar de haber visto el anuncio? Qué porcentaje lo han probado sin haber visto el anuncio?
- b) Cuál es el número medio de veces que han visto el anuncio las 1000 personas encuestadas?
- c) De qué manera ha influido la publicidad sobre el consumo de este producto? (Indicación: Calculad el coeficiente de correlación lineal entre estas dos variables e interpretad el resultado).
- d) Si la población total es de 3 millones de personas, y el litro de refresco vale 1 euro, estimad el gasto total en producto.

**Ejercicio 4.5** Para probar la efectividad de una vacuna se realiza el siguiente experimento a un grupo de personas: se vacuna a una parte de ellas y se observa si contraen o no la enfermedad. Los datos se recogen en la siguiente tabla:

	vacunados	no vacunados
enfermos	13	236
no enfermos	215	450

Estudiar la efectividad de la vacuna.

**Ejercicio 4.6** La tabla siguiente contiene 901 individuos clasificados según las variables categóricas  $X = \text{"sueldo (en dólares USA)"}$  e  $Y = \text{"satisfacción con el trabajo"}$ . Las categorías de la variable  $Y$  son:  $VI = \text{"muy insatisfecho"}$ ,  $MI = \text{"moderadamente insatisfecho"}$ ,  $MS = \text{"moderadamente satisfecho"}$ ,  $VS = \text{"muy satisfecho"}$ .

sueldo	VI	MI	MS	VS
< 6000	20	24	80	82
6000 – 15000	22	38	104	125
15000 – 25000	13	28	81	113
> 25000	7	18	54	92

Utilizando los datos de esta tabla calcular las medidas de asociación que se crean oportunas.

**Ejercicio 4.7** Se ha realizado una encuesta a 80 hombres casados en la que se les ha preguntado por el número de hermanos ( $X$ ) y el número de hijos ( $Y$ ) que han tenido. Los resultados son los que aparecen a continuación:

$X$	$Y$	$n_{ij}$	$X$	$Y$	$n_{ij}$	$X$	$Y$	$n_{ij}$
0	1	4	2	3	3	4	0	2
1	1	3	3	1	6	4	1	7
1	3	4	3	2	12	4	2	15
2	0	2	3	3	5	4	4	1
2	2	9	3	4	2	5	2	5

- Construye una tabla de doble entrada y representa la distribución mediante un diagrama de dispersión.
- Según el diagrama anterior, qué valor crees que tendrá la covarianza entre  $X$  e  $Y$ ? Y el coeficiente de correlación lineal? Justifica tu respuesta.
- Calcula el coeficiente de correlación lineal.

**Ejercicio 4.8** Se ha realizado un estudio de la mano dominante en cierta población. La tabla siguiente recoge las frecuencias observadas en una muestra aleatoria de esta población:

	<i>sexo</i>	
<i>mano dominante</i>	<i>hombres</i>	<i>mujeres</i>
<i>derecha</i>	2780	3281
<i>izquierda</i>	311	300

Puede afirmarse que existe relación entre el sexo y el ser diestro o siniestro? En caso que exista asociación entre estas variables, cuál es el grado de asociación? Justifica todas las medidas que utilices.

**Ejercicio 4.9** La tabla siguiente recoge los datos obtenidos en un estudio sobre la relación entre la educación que reciben los hijos y la orientación política de los padres.

	<i>orientación política</i>		
<i>educación</i>	<i>conservadores</i>	<i>moderados</i>	<i>liberales</i>
<i>no autoritaria</i>	7	9	14
<i>moderada</i>	10	10	8
<i>autoritaria</i>	15	11	6

Puede decirse que existe asociación entre estas dos variables? En caso que exista, cuál es el grado de asociación? Justifica las medidas que utilices.