

UNIDAD DIDÁCTICA 2

ESTADÍSTICA DESCRIPTIVA

(2ª parte)

Contenidos

2. Estadística descriptiva bidimensional

2.1 **Tablas de frecuencias cruzadas**

2.2 Diagrama de dispersión

2.3 Covarianza y coeficiente de correlación

NOTA: los apartados 2.2 y 2.3 se estudian antes del tema de regresión (2º parcial)

Tablas de frecuencias cruzadas

- Primer análisis de la relación existente entre las dos componentes de una v.a.bidimensional:



Tabla de frecuencias cruzada

- La tabla contiene la frecuencia con que se ha observado cada combinación de los valores posibles de ambas componentes.

Tablas de frecuencias cruzadas

- En el caso de que una o ambas componentes sean de tipo continuo es preciso primero proceder a un agrupamiento de sus valores en intervalos.
- Estas tablas son especialmente útiles para analizar v.a. bidimensionales con las dos componentes de tipo cualitativo:



Tablas de contingencia

Tablas de frecuencias cruzadas

- La tabla siguiente está obtenida a partir de las respuestas dadas por los alumnos a la encuesta, y cruza las variables **SEXO** y **REPITE**.

REPITE SEXO	SI	NO	Row Total
CHICOS	5 10.9	41 89.1	46 64.8
CHICAS	1 4.0	24 96.0	25 35.2
COLUMN TOTAL	6 8.5	65 91.5	

Tablas de frecuencias cruzadas

- Cada casilla recoge el número de individuos que tienen los valores correspondientes para las dos variables (SEXO y REPITE). Son las **frecuencias absolutas conjuntas**.

REPITE SEXO	SI	NO	Row Total
CHICOS	5 10.9	41 89.1	46 64.8
CHICAS	1 4.0	24 96.0	25 35.2
COLUMN TOTAL	6 8.5	65 91.5	

Tablas de frecuencias cruzadas

- A la derecha de la tabla se recogen las frecuencias marginales totales, tanto absolutas como relativas (estas últimas expresadas en %), para los dos valores de SEXO.
- A estas frecuencias se les denomina respectivamente, **frecuencias marginales absolutas** (en rojo) y **frecuencias marginales relativas** (en azul).

Tablas de frecuencias cruzadas

REPITE SEXO	SI	NO	Frecuencias marginales relativas de SEXO
CHICOS	5 10.9	41 89.1	46 64.8
CHICAS	1 4.0	24 96.0	25 35.2
COLUMN TOTAL	6 8.5	65 91.5	Frecuencias marginales absolutas de SEXO

Tablas de frecuencias cruzadas

- En la parte inferior de la tabla se recogen las **frecuencias marginales** para la variable REPITE.

REPITE SEXO	SI	NO	Row Total
CHICOS	5 10.9	41 89.1	46 64.8
CHICAS	1 4.0	24 96.0	25 35.2

COLUMN
TOTAL

Frecuencias
marginales
relativas de
REPITE

6

8.5

65

91.5

Frecuencias
marginales
absolutas de
REPITE

Tablas de frecuencias cruzadas

- Con el fin de estudiar si la proporción de repetidores es similar en los dos sexos conviene calcular la frecuencia relativa de cada casilla respecto al total de la fila correspondiente.
- Estas frecuencias relativas, que se recogen en la tabla en % en la parte inferior de cada casilla, se denominan **frecuencias relativas condicionales** de REPITE en función de los valores de SEXO.
- Por ejemplo, la proporción de chicos que repite es igual a **10,9%**. Sin embargo la proporción de chicas que repite es **4%**.

Tablas de frecuencias cruzadas

REPITE SEXO	SI	NO	Row Total
CHICOS	5 10.9	41 89.1	46 64.8
CHICAS	1 4.0	24 96.0	25 35.2
COLUMN TOTAL	6 8.5	65 91.5	

Frecuencias relativas de
REPITE condicionada a
SEXO

Tablas de frecuencias cruzadas

- Es importante que las frecuencias relativas se calculen adecuadamente respecto al total de la fila o de la columna correspondiente, según sea relevante para los objetivos perseguidos en un determinado estudio.

EJERCICIO 19: *Calcular a partir de los datos de la tabla anterior las frecuencias relativas condicionales de SEXO frente a REPITE. ¿Cuál de los dos conjuntos de frecuencias condicionales consideras que puede prestarse a una interpretación más interesante?*

Tablas de frecuencias cruzadas

SOLUCIÓN: Se calculan las frecuencias relativas condicionales de SEXO frente a REPITE, dividiendo las frecuencias marginales conjuntas por el total de cada columna

REPITE SEXO	SI	NO	Row Total
CHICOS	5 $(5/6) \times 100 = 83,3\%$	41 $(41/65) \times 100 = 63,08\%$	46 64.8
CHICAS	1 $(1/6) \times 100 = 16,7\%$	24 $(24/65) \times 100 = 36,92\%$	25 35.2
COLUMN TOTAL	6 8.5	65 91.5	

Tablas de frecuencias cruzadas

SOLUCIÓN: *las frecuencias relativas condicionales de SEXO frente a REPITE reflejan que hay mayor proporción de chicos en la muestra.*

Esto ya se sabía mirando las frecuencias marginales de SEXO.

Tablas de frecuencias cruzadas

- Cuando sobre cada individuo de la población se observan **dos** características de tipo cuantitativo se tiene una **variable aleatoria bidimensional cuantitativa**
- **Ejemplo 1**: En la población constituida por los estudiantes universitarios españoles se observa la ESTATURA (cms) y el PESO (kgs) de cada estudiante.

Una muestra de esta variable bidimensional puede estar formada por los 130 pares de valores constatados en los 130 alumnos que contestaron la encuesta.

Tablas de frecuencias cruzadas

- **Ejemplo 2**: En la población constituida por los días se funcionamiento de la calefacción de una factoría se observa diariamente la TEMPERATURA ($^{\circ}\text{C}$ a las 12) y el CONSUMO (termias) de energía de la calefacción.
Una muestra de esta variable bidimensional puede estar formada por los 57 pares de valores constatados en 57 días laborables de invierno de un determinado año.

Tablas de frecuencias cruzadas

- En el ejemplo anterior se expuso cómo podía describirse, mediante una Tabla de Contingencia, la relación entre las dos componentes de una variable bidimensional en el caso de que ambas fueran de tipo **cualitativo o atributo**.
- Cuando las dos variables sean de tipo **cuantitativo**, y especialmente cuando se trate de variables **continuas** (como sucede en los dos ejemplos anteriores) es posible utilizar técnicas más adecuadas para describir y analizar la relación existente entre ambas.

Tablas de frecuencias cruzadas

- Por supuesto es posible, en primer lugar, para estudiar la relación que hay entre las dos componentes de la v.a. bidimensional cuantitativa, construir una **tabla de frecuencias cruzada**.
- Previamente será necesario agrupar cada componente en intervalos o tramos.
- A continuación se obtienen las frecuencias observadas en cada combinación de tramos.

Tablas de frecuencias cruzadas

Ejemplo

ESTATURA PESO	145 155	155 165	165 175	175 185	185 195	Row Total
40 55	9 75,0%	17 44,7%	0 0,0%	0 0,0%	0 0,0%	26 20,0%
55 70	3 25,0%	18 47,4	31 53,4%	5 29,4%	0 0,0%	57 43,8%
70 85	0 0,0%	3 7,9%	24 41,4%	12 70,6%	3 60,0%	42 32,3%
85 100	0 0,0%	0 0,0%	3 5,2%	0 0,0%	2 40,0%	5 3,8%
Column Total	12 9,2%	38 29,2%	58 44,6	17 13,1%	5 3,8%	130 100

Frecuencias (absolutas y relativas) marginales de PESO

Tablas de frecuencias cruzadas

- En el margen derecho se recogen las frecuencias (absolutas y relativas, estas últimas expresadas como %) de los 4 tramos considerados para PESO.
- Estas frecuencias, que están obtenidas sumando para todos los valores posibles de ESTATURA se denominan **marginales**.

Tablas de frecuencias cruzadas

Ejemplo

ESTATURA PESO	145 155	155 165	165 175	175 185	185 195	Row Total
40 55	9 75,0%	17 44,7%	0 0,0%	0 0,0%	0 0,0%	26 20,0%
55 70	3 25,0%	18 47,4	31 53,4%	5 29,4%	0 0,0%	57 43,8%
70 85	0 0,0%	3 7,9%	24 41,4%	12 70,6%	3 60,0%	42 32,3%
85 100	0 0,0%	0 0,0%	3 5,2%	0 0,0%	2 40,0%	5 3,8%
Column Total	12 9,2%	38 29,2%	58 44,6	17 13,1%	5 3,8%	130 100

Frecuencias (absolutas y relativas) marginales de ESTATURA

Tablas de frecuencias cruzadas

- Dentro de cada columna se recogen las frecuencias observadas para los diferentes tramos de PESO en los individuos cuya ESTATURA se halla en el tramo considerado.
- Las frecuencias relativas están calculadas respecto a la frecuencia total de la columna considerada y se denominan **frecuencias relativas condicionales**.
- Así de los individuos cuya ESTATURA está en el tramo 145-155 el 75% pesan entre 40 y 55 Kg y el 25% entre 55 y 70 Kg, mientras que de los que miden entre 175 y 185 cm, el 29.4% pesan entre 55 y 70 Kg y el 70.6% pesan entre 70 y 85 Kg.

Distribuciones marginales y condicionales

Ejemplo

ESTATURA PESO	145 155	155 165	165 175	175 185	185 195	Row Total
40 55	9 75,0%	17 44,7%	0 0,0%	0 0,0%	0 0,0%	26 20,0%
55 70	3 25,0%	18 47,4	31 53,4%	5 29,4%	0 0,0%	57 43,8%
70 85	0 0,0%	3 7,9%	24 41,4%	12 70,6%	3 60,0%	42 32,3%
85 100	0 0,0%	0 0,0%	3 5,2%	0 0,0%	2 40,0%	5 3,8%
Column Total	12 9,2%	38 29,2%	58 44,6	17 13,1%	5 3,8%	130 100

Frecuencias observadas para el PESO condicionado a ESTATURA [145,155)

Ejercicios evaluación

- 1.- Con el fin de determinar la calidad de los programas, se seleccionan al azar 10 y se registra el número de errores que hay en cada uno. Los datos se exponen a continuación:

13, 12, 8, 18, 24, 12, 16, 30, 15, 10

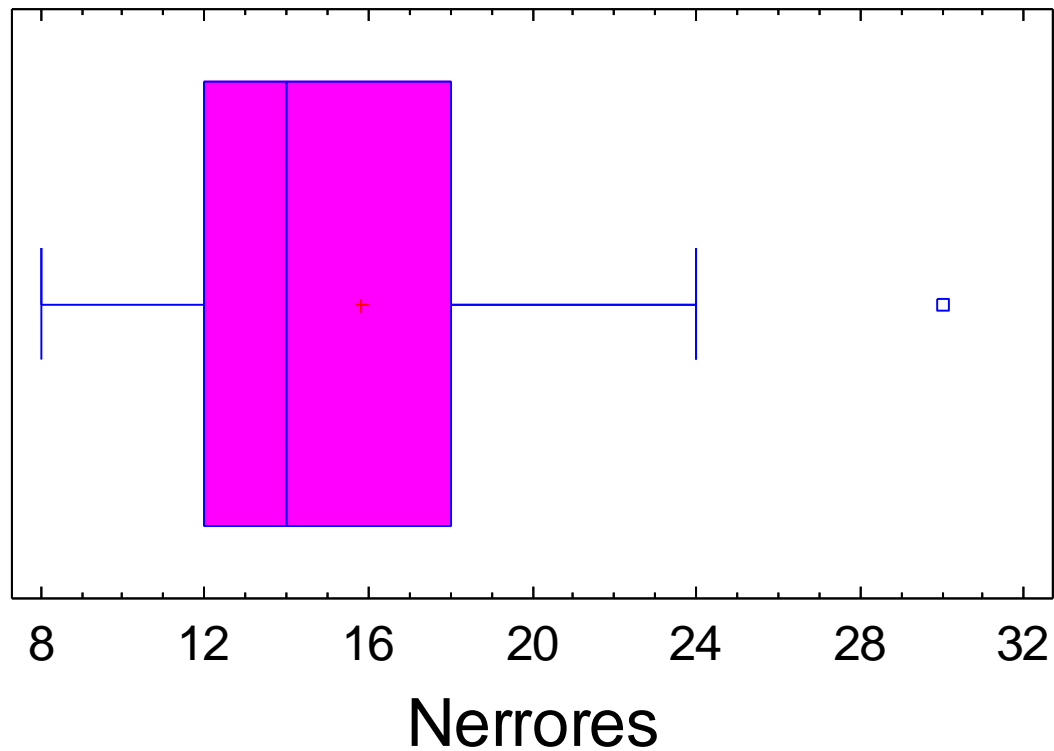
¿Cuál es la población en estudio? ¿Cuál es la variable aleatoria implicada? ¿Tipo de variable? ¿Mediana? ¿Cuartiles 1º y 3º? ¿Recorrido? ¿Recorrido intercuartílico?

A la vista de los datos y del diagrama *Box&Whisker* de este conjunto de datos y justificando las respuestas, ¿cuáles de las siguientes afirmaciones son falsas?:

- a) El recorrido intercuartílico vale 5
- b) Existe un único punto aislado o anómalo
- c) La mediana vale 14
- d) El parámetro de posición más adecuado para esta distribución sería la mediana
- e) El parámetro de dispersión más adecuado para esta distribución sería la desviación típica

Ejercicios evaluación

Box-and-Whisker Plot



Ejercicios evaluación

2- Tras realizar una encuesta a 129 alumnos de la UPV sobre el medio de transporte y el tipo de residencia que utilizaban durante el curso se ha obtenido la siguiente tabla:

	COCHE	BICI/MOTO	A PIE	AMIGO	BUS	Total Row
FAMILIA	28 21.7	7	8 6.2	5 3.9	19 14.7	67 51.9
RESIDEN	1 0.8	0 0.0	9 7.0	0 0.0	2 1.6	12 9.3
PISO	7	0.8	27	0 0.0	1 0.8	36 27.9
PENSIÓN	0 0.0	0 0.0		0 0.0	1 0.8	1 0.8
OTRO	4 3.1	2 1.6	3 2.3	0 0.0	4 3.1	13 10.1
Total Column	40 31.0	10 7.8	47	5 3.9	27 20.93	129 100.0

Ejercicios evaluación

Con la información derivada de esta tabla, rellenar los valores que faltan, marcar cuáles de las siguientes afirmaciones son ciertas e indicar a qué tipo de frecuencia corresponden:

- a)** El 5,4% de los alumnos que vienen en BICI/MOTO viven con su FAMILIA.
- b)** El 20,93% de los alumnos vienen a PIE y viven en un PISO
- c)** El 20,93% de los alumnos vienen en BUS
- d)** El 5,4% de los alumnos que viven en PISO vienen en COCHE

Frecuencias absolutas conjuntas

	COCHE	BICI/MOTO	A PIE
FAMILIA	28	7	8
RESIDEN	1	0	9
PISO	7	3	27

- a) Frecuencia relativa marginal de venir en bici/moto
- b) Frecuencia relativa de vivir en piso condicionada a venir en bici/moto
- c) Frecuencia relativa conjunta de vivir en familia y venir a pie
- d) Frecuencia relativa de venir en coche condicionada a vivir en piso