

1. Solución: La tipificación es un procedimiento que se utiliza para **comparar relativamente** dos valores de distintas muestras, es decir, para comparar esos valores considerando cada uno en relación con su muestra correspondiente.

La tipificación trata de cuantificar si un valor x_i es grande o pequeño en relación con el resto de la muestra. Para ello se calcula la diferencia a la media $x_i - \bar{x}$, lo que nos informa de la distancia que hay entre el x_i y la media y si está por encima (positivo) o por debajo (negativo). Como esa diferencia o distancia será grande o pequeña dependiendo de cómo sean las distancias a la media en conjunto y de esto nos informa la desviación típica, parece razonable dividir por la desviación para conseguir la medida adimensional que necesitamos para establecer comparaciones. Así conseguimos el valor tipificado: $\frac{x_i - \bar{x}}{S_x}$.

Por ejemplo, sabemos que una persona sacó un 6.3 en Estadística Administrativa I (EAI) y un 5.8 en Estadística Administrativa II (EAI). Además la nota media de EAI fue de 5.7 con una desviación de 1.5 puntos y la de EAI fue de 6.1 con una desviación de 1.8. Si queremos comparar las notas de esa persona en términos relativos haríamos:

$$\text{EAI} : \frac{6,3 - 5,7}{1,5} = 0,4 \quad \text{EAI} : \frac{5,8 - 6,1}{1,8} = -0,1667$$

Diríamos que sacó peor nota relativa en EAI porque el valor tipificado es menor. De hecho, vemos que en EAI está por debajo de la media y en EAI está por encima (a pesar de hacer sacado mayor nota en valor absoluto en EAI que en EAI).

2. Solución: En primer lugar debemos plantear el problema. El experimento consiste en observar el número de pases que conceden (variable) en cada trimestre (individuos), luego la población serán todos los trimestres y tenemos una muestra de 4 trimestres. La variable es cardinal (porque los valores son números), de razón (porque 0 significa ausencia) y discreta (porque pueden dar 456 o 457 pases pero no 456.5).

En el apartado a) nos piden que calculemos un “valor central”. Como no hay valores extremos, sabemos que el mejor valor central en este caso es la media aritmética. Además, la dispersión entorno a la media se mide con la desviación típica, así que construiremos la tabla de frecuencias (ordenada) para calcular ambos valores:

x_i	n_i	$x_i n_i$	$x_i^2 n_i$
815	1	$815 \times 1 = 815$	$815^2 \times 1 = 664225$
904	1	$904 \times 1 = 904$	$904^2 \times 1 = 817216$
926	1	$926 \times 1 = 926$	$926^2 \times 1 = 857476$
1023	1	$1023 \times 1 = 1023$	$1023^2 \times 1 = 1046529$
Total	4	3668	3385446

$$\bar{x} = \frac{\text{nº de pases concedidos en total}}{\text{nº de trimestres}} = \frac{\sum_{i=1}^n x_i n_i}{N} = \frac{3668}{4} = 917 \text{ pases}$$

$$\overline{x^2} = \frac{\sum_{i=1}^n x_i^2 n_i}{N} = \frac{3385446}{4} = 846361,5 \text{ pases}^2 \quad S_x^2 = \overline{x^2} - \bar{x}^2 = 846361,5 - 917^2 = 5472,5 \text{ pases}^2$$

$$S_x = \sqrt{S_x^2} = \sqrt{5472,5} = 73,9763 \text{ pases}$$

Por lo que el número de pases concedidos por trimestre varía alrededor de 917 con una dispersión de 73.9763 pases en conjunto alrededor de ese valor.

En el apartado b) nos piden la tasa media de crecimiento trimestral, por lo que tendremos que calcular las tasas o razones de crecimiento trimestrales y posteriormente construir la tabla de la media geométrica.

$$1^a \text{ Tasa de crecimiento} = \frac{\text{número de pases del 2º trimestre}}{\text{número de pases del 1º trimestre}} = \frac{926}{815} = 1,1362,$$

como la tasa es mayor que 1, significa que hubo un aumento. En concreto, como $1.1362 - 1 = 0.1362$, podemos decir que en el segundo trimestre hubo un incremento del número de pases concedidos del 13.64 % en relación al primero.

$$2^{\text{a}} \text{ Tasa de crecimiento} = \frac{\text{número de pases del 3}^{\text{o}} \text{ trimestre}}{\text{número de pases del 2}^{\text{o}} \text{ trimestre}} = \frac{904}{826} = 0,9762,$$

como la tasa es menor que 1, significa que hubo un descenso. En concreto, como $1-0.9762=0.028$, podemos decir que en el tercer trimestre hubo un descenso del número de pases concedidos del 2.8 % en relación al segundo.

$$3^{\text{a}} \text{ Tasa de crecimiento} = \frac{\text{número de pases del 4}^{\text{o}} \text{ trimestre}}{\text{número de pases del 3}^{\text{o}} \text{ trimestre}} = \frac{1023}{904} = 1,1316,$$

como la tasa es mayor que 1, significa que hubo un aumento. En concreto, como $1-1.1316=0.1316$, podemos decir que en el cuarto trimestre hubo un incremento del número de pases concedidos del 13.16 % en relación al tercero. Tabularmente, quedaría:

x_i	n_i	$x_i^{n_i}$
1.1362	1	$1,1362^1 = 1,1362$
0.9762	1	$0,9762^1 = 0,9762$
1.1316	1	$1,1316^1 = 1,1316$
Total	3	1.2551

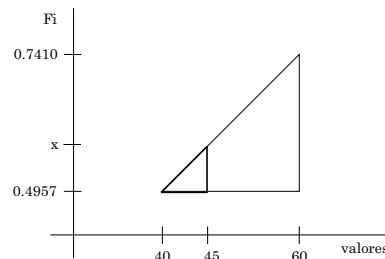
$$x_G = \sqrt[3]{x_1^{n_1} \times \dots \times x_k^{n_k}} = \sqrt[3]{1,1362 \times 0,9762 \times 1,1316} = 1,0787,$$

La razón media de crecimiento es de 1.0787, lo que significa que el crecimiento trimestral varió alrededor del 7.87 %.

3. Solución: En primer lugar debemos plantear el problema. El experimento consiste en observar la edad (variable) de los asturianos (individuos), luego la población serán todos los asturianos y tenemos una muestra (que coincide con la población) de $213237 + \dots + 47517 = 1088173$ individuos. La variable es cardinal (porque los valores son números), de razón (porque 0 significa ausencia) y continua (porque se pueden tener 22.5 años y 22.578, etc.).

En el apartado a) nos piden que hallemos el percentil en el que está una persona de 45 años, es decir, el porcentaje de población que se acumula entre 0 y 45 años. Para ello debemos hallar las frecuencias relativas acumuladas. Como tenemos datos agrupados, para hallar el percentil con el que se corresponde justamente con la edad de 45 años, podemos suponer que la edad de los asturianos entre 40 y 60 se distribuye uniformemente y aplicar entonces triángulos semejantes. Es decir:

Clases	n_i	F_i
[0,20]	213237	0.1960
(20,40]	326174	0.4957
(40,60]	266955	0.7410
(60,80]	234290	0.9563
80,100]	47517	1
Total	1088173	—



Siguiendo la regla de triángulos semejantes (base grande es a altura grande lo mismo que base pequeña es a altura pequeña), tenemos la siguiente regla de tres:

$$\begin{array}{lcl} \text{base grande}=60-40 & \text{—} & \text{altura grande}=0.7410-0.4957 \\ \text{base pequeña}=45-40 & \text{—} & \text{altura pequeña}=x-0.4957 \end{array} \Rightarrow \begin{array}{lcl} 20 & \text{—} & 0.2453 \\ 5 & \text{—} & x-0.4957 \end{array}$$

Despejando, tenemos que $x = 0,5570$, luego una persona de 45 años se encuentra en el cuantil 0,557, entre los percentiles 55 y 56, esto quiere decir que el 55.7 % de los asturianos tenían la misma edad (45 años) o eran más jóvenes que esa persona (y el resto serían de la misma edad o mayores).

En el apartado b) nos piden interpretar los valores de la tabla.

Media recortada al 5 % : eliminando las personas más jóvenes (concretamente el 2.5 %) y las personas más viejas (otro 2.5 %), la edad media del resto es de 41.5958 años (es decir, la edad del 95 % central oscila entorno a los 41.4958 años).

Mediana: La mitad de los asturianos tienen 50 años o menos, y el resto 50 años o más.

Amplitud intercuartil: mide la diferencia máxima de edad de la gente de “mediana edad” $C_3 - C_1$, luego cuantifica lo próximos que están entre sí esos asturianos en cuanto a edad. En concreto, quitando los más jóvenes (el 25 %) y los más viejos (otro 25 %) la diferencia máxima de edad es de 40 años.

Asimetría: al ser un valor positivo, aunque no en excesivamente grande (0.197) indica que la distribución de edad en Asturias ese año era ligeramente asimétrica a la derecha. Es decir, había más mayores que jóvenes (aunque no era una diferencia excesiva).

Curtois: es un valor negativo, lo que indica que la distribución es más plana que la normal, por lo que la gente de edad cercana a la media está menos concentrada que en una distribución normal.

4. Solución: En primer lugar debemos plantear el problema. El experimento consiste en observar el precio de compra (variable 1) y el precio de venta (variable 2) de cada botella de vino (individuo), luego la población serán todas las botellas de vino y tenemos una muestra de $97 + \dots + 13 = 330$ botellas. La variable es cardinal (porque los valores son números), de razón (porque 0 significa ausencia) y continua (porque el precio puede ser, en principio, cualquiera).

En el apartado a) nos piden que hallemos la ganancia que han obtenido con esas 330 botellas. Si consideramos la ganancia como los ingresos menos los gastos, deberíamos calcular el total de la diferencia entre el precio de venta y el precio de compra, para lo que tenemos que realizar la tabla de frecuencias de la variable ganancia=precio de venta-precio de compra:

Precios de venta	Precio de compra	Ganancia (x_i)	n_i	$x_i n_i$
3.10	1.05	2.05	97	$2.05 \times 9 = 198.85$
4.90	1.95	2.95	131	$2.95 \times 131 = 386.45$
6.60	2.85	3.75	63	$3.75 \times 63 = 236.25$
9.50	4.45	5.05	26	$5.05 \times 23 = 131.3$
16	6.95	9.05	13	$9.05 \times 13 = 117.65$
Total	—	—	330	1070.5

Por lo que, la ganancia que han obtenido con la venta de esas 330 botellas, será $x_T = \sum_{i=1}^k x_i n_i = 1070,5$ euros.

En el apartado b) nos piden hacer una predicción del precio de venta de una botella sabiendo que el precio de compra es de 2.15 euros, para lo que tendremos que hacer una recta de regresión con $Y =$ precio de venta (que es lo que pretendemos predecir) y $X =$ precio de compra (que es lo que conocemos). Debemos fijarnos en el precio de compra que nos dan está entre el precio mínimo (que es 3.10) y el máximo (que es 16), porque si no fuera así, no podríamos hacer la predicción. Sabemos que la recta de regresión es $y^* = a^*x + b^*$ con $a^* = S_{xy}/S_x^2$ y $b^* = \bar{y} - a^*\bar{x}$, luego tenemos que hacer los cálculos:

$$\overline{xy} = \frac{1,05 \times 3,1 \times 97 + \dots + 6,95 \times 16 \times 13}{330} = 16,0522$$

$$\bar{x} = \frac{1,05 \times 97 + \dots + 6,95 \times 13}{330} = 2,2512 \quad \bar{y} = \frac{3,10 \times 97 + \dots + 16 \times 13}{330} = 5,4952$$

$$\overline{x^2} = \frac{1,05^2 \times 97 + \dots + 6,95^2 \times 13}{330} = 6,8472 \quad S_x^2 = \overline{x^2} - \bar{x}^2 = 6,8472 - 2,2512^2 = 1,7793$$

$$S_{xy} = \overline{xy} - \bar{x}\bar{y} = 16,0522 - 2,2512 \times 5,4952 = 3,6814$$

$$a^* = \frac{3,6814}{1,7793} = 2,0690 \quad b^* = 5,4952 - 2,0690 \times 2,2512 = 0,8375$$

por lo que $y^* = 2,0690x + 0,8375$. Para realizar la predicción únicamente tenemos que sustituir x por el valor que nos dan, y nos quedaría $y^*(2,15) = 2,0690 \times 2,15 + 0,8375 = 5,2859$ euros. Diríamos entonces que creemos que una botella que cuesta 2.15 euros, la venderán a 5.2859 euros aproximadamente.

Para saber si nuestra predicción es fiable, debemos calcular el coeficiente de determinación $R^2 = \frac{S_{xy}^2}{S_x^2 S_y^2}$, por lo que tenemos que calcular en primer lugar la varianza de Y :

$$\overline{y^2} = \frac{3,10^2 \times 97 + \dots + 6,95^2 \times 13}{330} = 37,8675; \quad S_y^2 = \overline{y^2} - \bar{y}^2 = 37,8675 - 5,4952^2 = 7,6703$$

$$R^2 = \frac{3,6814^2}{1,7793 \times 7,6703} = 0,9930$$

lo que significa que el 99.30 % de la variación del precio de venta queda explicada por la regresión con el precio de compra. Así que el precio de venta queda determinado por el precio de compra casi en su totalidad, por lo que la regresión es muy buena y, por lo tanto, la predicción es muy fiable.