Estadística descriptiva

Resúmenes y representaciones de muestras

1. En 1798 el científico inglés Henry Cavendish midió la densidad de la Tierra haciendo un cuidadoso experimento con una balanza de torsión. He aquí sus 29 medidas de la densidad de la Tierra respecto de la del agua (realizadas con el mismo instrumento):

5.50	5.61	4.88	5.07	5.26	5.55	5.36	5.29	5.58	5.65	5.57	5.53	5.62	5.29	5.44
5.34	5.79	5.10	5.27	5.39	5.42	5.47	5.63	5.34	5.46	5.30	5.75	5.68	5.85	

a) Halla la mediana, los cuartiles, el rango total y el rango intercuartílico.

b) Dibuja un histograma con los datos. ¿Es una distribución simétrica? ¿Hay valores atípicos?

c) La media de las 29 medidas es la mejor estimación de Cavendish de la densidad de la Tierra. Encuéntrese esta media y la desviación estándar.

2. Un investigador de una piscifactoría recopiló los siguientes datos sobre la longitud de las carpas hembras de 6 años (en milímetros):

217	230	220	221	225	223	219	217	225	228	234	222	231
222	220	222	222	223	225	214	221	233	227	234	223	225
253	220	213	224	235	283	210	218	235	231			

Agrupa los datos en 5 clases de 15 mm de anchura, comenzando en 210. Dibuja el histograma correspondiente. Calcula la media y desviación estándar ¿Hay algún valor atípico? Si los eliminamos, ¿cuál es la nueva media?

3. En 1879 Michelson obtuvo los siguientes valores para la velocidad de la luz en el aire (se dan los resultados restando 299.000 a los datos originales, en km./sg.):

850, 740, 900, 1070, 930, 850, 950, 980, 980, 880, 1000, 980, 930, 650, 760.

En 1882 Newcomb, utilizando otro procedimiento, obtuvo (restando de nuevo 299.000):

883, 816, 778, 796, 682, 711, 611, 599, 1051, 781, 578, 796, 774, 820, 771.

Dibuja histogramas para ambas distribuciones. Calcula las medias y desviaciones típicas. ¿Qué conclusiones pueden extraerse?

4. Determina razonadamente si las siguientes afirmaciones son verdaderas o falsas:

a) Si añadimos 7 a todos los datos de una muestra, el primer cuartil aumenta en 7 unidades y el rango intercuartílico no cambia.

b) Al restar 1 a cada dato de una muestra, la desviación típica siempre disminuye.

c) Si cambiamos el signo de todos los datos de una muestra, el coeficiente de asimetría también cambia de signo.

d) Al multiplicar por 3 todos los datos de una muestra, el coeficiente de asimetría no varía.

e) Si a una muestra de datos con media $\bar{x}$ se le añade un nuevo dato que coincide con $\bar{x}$, la media no cambia y la desviación típica disminuye.

5. La media de las variaciones mensuales del PIB de la Comunidad de Murcia de los nueve primeros meses del año ha sido del 0.1%. ¿Cuál debe ser la media de los últimos tres meses para que la media anual cumpla el objetivo del 0.2%?

6. a) Disponemos de una serie de datos $(x_1, \ldots, x_{100})$, ya ordenados de menor a mayor, cuya media muestral es $\bar{x}$. Ahora formamos una nueva serie añadiendo a la anterior los valores x_1 y x_{100}. ¿Qué condición se debe cumplir para que la media muestral de la nueva muestra coincida con $\bar{x}$, la media muestral de la muestra original?

b) Tenemos una serie de datos $(x_1, x_2, \ldots, x_n)$, cuya media es a y cuya varianza muestral es b. Duplicamos ahora el tamaño de la serie añadiendo los opuestos (en signo) de los datos originales: $(x_1, x_2, \ldots, x_n, -x_1, -x_2, \ldots, -x_n)$. Llamemos b' a la varianza muestral de la nueva serie de datos. ¿Quién es mayor, b ó b'?

Correlaciones, covarianzas y recta de regresión

7. En un experimento sobre la ley de Hooke se colocaron pesos de varios tamaños en el extremo de una cuerda de piano. La tabla de la derecha da los pesos colocados y la longitud que adquirió la cuerda. Calcula el coeficiente de correlación y la recta de regresión. ¿Cuál es el error cuadrático medio de los datos?

Peso (kg)	Longitud (cm)
0	439.00
2	439.12
4	439.21
6	439.31
8	439.40
10	439.50

8. Una empresa quiere conocer la relación entre el tamaño de su equipo de ventas y sus ingresos anuales (en millones de pesetas). Se examinan las cifras de los 10 últimos años, obteniéndose los datos siguientes:

Año	75	76	77	78	79	80	81	82	83	84
Plantilla	15	18	24	22	25	29	30	32	35	38
Ventas	13.5	16.3	23.3	24.1	26.3	29.3	34.1	32.6	36.3	41.5

a) Dibuja la nube de puntos ($X =$ plantilla, $Y =$ ventas). ¿Hay relación lineal entre las dos series?

b) Calcula el coeficiente de correlación y halla la recta de regresión de Y sobre X.

9. Los datos de mortalidad infantil (muertes por mil partos) en un país durante los años 2008-2012 fueron (tomando 2010 como año 0):

X: año	-2	-1	0	1	2
Y: mortandad	14.5	13.8	12.7	11.9	11.4

a) Ajusta a estos datos una ecuación de la forma $Y = ae^{bX}$, transformando primero a una regresión lineal.

b) Calcula el coeficiente de correlación de la regresión lineal y comenta la bondad del ajuste.

c) ¿Qué mortalidad infantil se espera para 2020 (año 10) si se da por bueno el ajuste anterior?

10. En la tabla se recogen medidas (bajo ciertas condiciones) del volumen de un determinado gas al someterlo a distintas presiones:

P presión	1	1.5	2	2.5	3
V volumen	1	0.76	0.62	0.52	0.46

a) Ajusta a los datos una ecuación de la forma $V = aP^b$, transformando primero a una regresión lineal.

b) Calcula el coeficiente de correlación lineal en el problema transformado y comenta la bondad del ajuste.

c) ¿Qué volumen corresponde a $P = 3.5$ si se da por bueno el ajuste anterior?

Ejercicios adicionales

11. Dada una muestra $(x_1, x_2, \ldots, x_n)$, comprueba que la media $\overline{x}$ es aquel valor respecto del cual la dispersión cuadrática media es mínima, es decir, considera la función d dada por

$$a \in \mathbb{R} \longmapsto d(a) = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2 ,$$

y comprueba que el mínimo de d se alcanza en $a = \overline{x}$ (por tanto, el valor mínimo de d es justamente la varianza de los x_i).

12. Dada una muestra $(x_1, y_1), \ldots, (x_n, y_n)$ con $n \geq 2$, se pide obtener la recta $y = \hat{b}x$ que pasa por el origen $(0, 0)$ y que da el menor error cuadrático medio de entre todas las rectas de ecuación $y = bx$. Da la expresión del error cuadrático mínimo en términos de la muestra.