Ejercicios - Estadística descriptiva

Curso: Universidad 1 · 20

Se debe mostrar el desarrollo de cada ejercicio.

Ejercicio 1

Ejercicio 1 – Tabla de frecuencias

Se pregunta a 20 estudiantes de primero de una facultad de Málaga cuántos hermanos tienen. Los resultados son: 4 estudiantes con 0 hermanos, 8 con 1, 5 con 2, 2 con 3 y 1 con 4.

a. Clasifica la variable.
b. Construye la tabla con frecuencias absolutas, relativas y acumuladas.
c. ¿Qué porcentaje de estudiantes tiene como mucho un hermano?
d. ¿Qué porcentaje tiene al menos dos hermanos?

Ejercicio 2

Ejercicio 2 – Media, mediana y moda

Con los datos del ejercicio anterior sobre el número de hermanos:

a. Calcula la media.
b. Calcula la mediana.
c. Indica la moda.
d. ¿Qué medida describe mejor el «estudiante típico»? Razona.

Ejercicio 3

Ejercicio 3 – Varianza y desviación típica

Sigue con los datos del número de hermanos ($\bar{x} = 1{,}4$).

a. Calcula $\sum f_i x_i^2$.
b. Calcula la varianza con la fórmula $\sigma^2 = \frac{\sum f_i x_i^2}{n} - \bar{x}^2$.
c. Calcula la desviación típica y el coeficiente de variación.
d. ¿Qué porcentaje de datos está en el intervalo $\bar{x} \pm \sigma$?

Ejercicio 4

Ejercicio 4 – Datos agrupados en intervalos

Las estaturas (en cm) de 40 jugadores de un club de baloncesto de Badalona se agrupan en la tabla.

Estatura (cm)	[150,160)	[160,170)	[170,180)	[180,190)
Jugadores	6	14	12	8

a. Calcula las marcas de clase y la estatura media.
b. Calcula la desviación típica.
c. Estima la mediana por interpolación en el intervalo mediano.
d. ¿Qué intervalo es el modal?

Ejercicio 5

Ejercicio 5 – Cuartiles y diagrama de caja

Los minutos que tardan 14 estudiantes en llegar a la facultad son: 12, 15, 17, 18, 18, 20, 21, 22, 23, 25, 26, 28, 30, 55. (Toma como Q_1 y Q_3 las medianas de la mitad inferior y de la superior.)

a. Calcula la mediana y los cuartiles.
b. Calcula el rango intercuartílico.
c. Aplica el criterio de Tukey: ¿hay valores atípicos?
d. Describe el diagrama de caja y comenta la asimetría.

Ejercicio 6

Ejercicio 6 – Coeficiente de variación

Un nutricionista compara la dispersión del peso de recién nacidos ($\bar{x} = 3{,}3$ kg, $\sigma = 0{,}45$ kg) y de adultos ($\bar{x} = 72$ kg, $\sigma = 11$ kg).

a. ¿Qué grupo tiene mayor desviación típica?
b. Calcula los coeficientes de variación.
c. ¿Qué grupo es relativamente más disperso?
d. ¿Por qué no se pueden comparar directamente las desviaciones típicas?

Ejercicio 7

Ejercicio 7 – Dos grupos de Estadística

Las notas de un examen de Estadística en dos grupos de la Universidad de Sevilla son: grupo A: 4, 5, 5, 6, 6, 6, 7, 7, 8, 9; grupo B: 2, 3, 4, 5, 6, 7, 8, 9, 9, 10.

a. Calcula la media de cada grupo.
b. Calcula la desviación típica de cada grupo.
c. ¿Qué grupo es más homogéneo?
d. Calcula la mediana de cada grupo y comenta qué grupo tiene más alumnos con 8 o más.

Ejercicio 8

Ejercicio 8 – La espera en una parada de autobús

La EMT de Valencia registra el tiempo de espera de 60 viajeros en una parada. La tabla recoge los datos agrupados (en minutos).

Espera (min)	[0,4)	[4,8)	[8,12)	[12,16)
Viajeros	18	22	12	8

a. Calcula el tiempo medio de espera.
b. Calcula la desviación típica.
c. ¿Qué porcentaje de viajeros espera 8 minutos o más?
d. La empresa promete que el 60 % espera menos de 6 minutos. Estima el percentil 60 y decide si cumple.

Ejercicio 9

Ejercicio 9 – El alquiler en un barrio de Madrid

Los alquileres mensuales (en euros) de 8 pisos anunciados en un barrio de Madrid son: 700, 750, 800, 820, 850, 900, 950 y 3500 (un ático de lujo).

a. Calcula la media y la mediana.
b. Calcula de nuevo ambas sin el ático.
c. ¿Qué medida es más robusta frente a valores atípicos?
d. Un portal inmobiliario anuncia «alquiler medio de 1159 €». ¿Es una información engañosa? Razona.

Ejercicio 10

Ejercicio 10 – Temperaturas en Burgos

Las temperaturas mínimas de cinco días de enero en Burgos fueron 2, 5, 8, 10 y 13 °C. Un turista quiere los datos en grados Fahrenheit: $F = 1{,}8C + 32$.

a. Calcula la media y la desviación típica en °C.
b. Calcula la media y la desviación típica en °F sin transformar cada dato.
c. Comprueba la media transformando los datos.
d. ¿Cambia el coeficiente de variación? ¿Tiene sentido compararlo?

Ejercicio 11

Ejercicio 11 – Horas de estudio y nota

Seis estudiantes de una academia de Zaragoza anotan las horas de estudio semanales x y la nota del examen y: $(2, 4)$, $(4, 5)$, $(5, 6)$, $(7, 7)$, $(8, 8)$, $(10, 9)$.

a. Calcula las medias y la covarianza.

b. Calcula el coeficiente de correlación e interprétalo.
c. Escribe la recta de regresión de y sobre x.
d. Predice la nota con 6 horas de estudio. ¿Es fiable la predicción con 20 horas?

Ejercicio 12

Ejercicio 12 – Publicidad y ventas

Una tienda de deportes de Pamplona anota durante 5 meses el gasto en publicidad x (miles de euros) y las ventas y (miles de euros): $(1, 12)$, $(2, 15)$, $(3, 19)$, $(4, 20)$, $(5, 24)$.

a. Calcula la recta de regresión de y sobre x.
b. Calcula el coeficiente de correlación y el de determinación r^2.
c. Interpreta la pendiente.
d. ¿Demuestran estos datos que la publicidad causa las ventas? Razona.

Ejercicio 13

Ejercicio 13 – Comparar con puntuaciones tipificadas

Lucía sacó un 7,5 en Cálculo (media de la clase 6, desviación típica 1,2) y Pablo un 8 en Álgebra (media 7, desviación típica 1,6). Ambos están en la misma facultad de Salamanca.

a. Calcula la puntuación tipificada de cada uno.
b. ¿Quién obtuvo un resultado relativamente mejor?
c. ¿Qué nota en Álgebra equivaldría al resultado de Lucía?
d. Si las notas fueran aproximadamente normales, ¿qué porcentaje de la clase superó aproximadamente a Lucía?

Ejercicio 14

Ejercicio 14 – Consumo de agua de los hogares

Una compañía de aguas de Murcia agrupa el consumo mensual de 100 hogares (en litros por día).

Consumo (L/día)	[0,100)	[100,200)	[200,300)	[300,400)
Hogares	10	35	40	15

a. Construye las frecuencias acumuladas.
b. Estima la mediana.
c. Estima el percentil 90.
d. La compañía quiere avisar al 10 % de hogares que más consume. ¿A partir de qué consumo enviará el aviso?

Ejercicio 15

Ejercicio 15 – Problema: encuesta de satisfacción

Un hospital de Pamplona pide a 100 pacientes que valoren la atención de 1 (muy mala) a 5 (excelente). La tabla recoge los resultados. La dirección quiere un informe para decidir si mantiene el nuevo protocolo de acogida, que exigía una valoración media de al menos 3,8 y que al menos el 70 % de los pacientes diera 4 o 5.

Valoración	1	2	3	4	5
Pacientes	3	7	15	45	30

a. Calcula la media y la desviación típica.
b. Calcula la mediana y la moda.
c. ¿Se cumplen los dos objetivos?
d. ¿Es adecuada la media para una variable de este tipo? Comenta otras medidas útiles.
e. Redacta una conclusión breve para la dirección.

Ejercicio 16

Ejercicio 16 – Problema: crecimiento de un cultivo

En el laboratorio de microbiología de la Universidad de Navarra se cuentan las bacterias (en miles) de un cultivo cada hora: 100, 150, 220, 330 y 500 para $t = 0, 1, 2, 3, 4$. Se sospecha un crecimiento exponencial $y = ae^{bt}$, que no es una recta, pero se puede linealizar.

a. Toma logaritmos: comprueba que $\ln y = \ln a + bt$ es una relación lineal entre t y $\ln y$.
b. Calcula $\ln y$ para cada dato.
c. Calcula la recta de regresión de $\ln y$ sobre t y el coeficiente de correlación.
d. Deduce a y b y predice el número de bacterias a las 5 horas.
e. ¿Cuánto tarda la población en duplicarse según el modelo?

Ejercicio 17

Ejercicio 17 – Problema: la desigualdad de Chebyshev

Un fabricante de bombillas de Sabadell sabe que la duración media es de 1000 h con desviación típica de 50 h, pero desconoce la forma de la distribución. La desigualdad de Chebyshev asegura que al menos una proporción $1 - \frac{1}{k^2}$ de los datos está a menos de k desviaciones de la media.

a. ¿Qué proporción mínima de bombillas dura entre 900 y 1100 h?
b. ¿Y entre 850 y 1150 h?
c. ¿Qué intervalo simétrico garantiza al menos el 96 %?
d. Si la distribución fuera normal, la proporción del apartado a) sería del 95 %. ¿Por qué Chebyshev da un valor menor?

Ejercicio 18

Ejercicio 18 – Problema: unir dos grupos

En una asignatura de una universidad de Valladolid hay dos grupos. El grupo de mañana tiene 30 estudiantes con media 6,2 y desviación típica 1,5; el de tarde, 20 estudiantes con media 7,0 y desviación típica 1,0. El coordinador solo tiene estos resúmenes y quiere los de toda la asignatura.

a. Calcula la media conjunta.
b. Calcula $\sum x_i^2$ de cada grupo usando $\sigma^2 = \frac{\sum x_i^2}{n} - \bar{x}^2$.
c. Calcula la varianza y la desviación típica conjuntas.
d. ¿Por qué la desviación conjunta no es la media ponderada de las desviaciones?

Ejercicio 19

Ejercicio 19 – Problema: un dato atípico y la correlación

Un estudio sobre 6 comercios de un barrio de Gijón relaciona los metros de escaparate x con las ventas diarias y (cientos de euros): $(1, 2)$, $(2, 3)$, $(3, 3)$, $(4, 5)$, $(5, 4)$, $(6, 20)$. El último es un gran almacén.

a. Calcula el coeficiente de correlación con los 6 datos.
b. Calcúlalo sin el gran almacén.
c. Explica la diferencia.
d. ¿Qué recomendarías hacer antes de calcular una correlación?

Ejercicio 20

Ejercicio 20 – Desafío: propiedades de la media y la mediana

Demuestra las siguientes propiedades para datos $x_1, \ldots, x_n$.

a. $\sum (x_i - \bar{x}) = 0$.
b. $\frac{1}{n}\sum (x_i - \bar{x})^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2$.
c. La función $S(a) = \sum (x_i - a)^2$ es mínima para $a = \bar{x}$.
d. La función $T(a) = \sum |x_i - a|$ es mínima en la mediana (razónalo con 3 datos).