

PROBLEMAS. RELACIÓN 1: Estadística Descriptiva Univariante.

FUNDAMENTOS DE PROBABILIDAD Y ANÁLISIS EXPLORATORIO DE DATOS.

GRADO EN CIENCIA E INGENIERÍA DE DATOS.

1. Los siguientes datos corresponden al nivel de glucosa en sangre de diez niños;

56	62	63	65	65	65	65	68	70	72
----	----	----	----	----	----	----	----	----	----

- Calcula la media, la mediana y la moda. ¿Son parecidas? ¿Por qué?
- Dibuja el diagrama de barras y el gráfico acumulativo de frecuencias.
- Calcula el rango, la varianza y la desviación típica.
- Calcula el coeficiente de variación, el coeficiente de asimetría y el coeficiente de apuntamiento.

Interpreta sus resultados.

2. Con el fin de analizar los sueldos de los habitantes de una comarca, se entrevistó a 20 trabajadores en activo que proporcionaron los siguientes valores de sus sueldos brutos anuales (en miles de euros):

7	9	7	158	11	37	37	31	89	60
98	26	69	41	16	22	1	7	29	125

- Para los datos SIN agrupar calcula la media aritmética, la mediana, los cuartiles, la desviación típica, la varianza, el coeficiente de variación y el rango intercuartílico. Compara los valores de la media y la mediana, ¿son parecidos o distintos?, ¿a qué se debe?
- Dibuja de forma detallada el diagrama de caja y bigotes del conjunto de datos. Comentar las características más relevantes: ¿los datos se distribuyen de forma simétrica o asimétrica?, ¿cuál es el intervalo de valores admisibles, es decir, dónde se encuentran los no atípicos?, ¿existen valores atípicos?
- Si queremos agrupar los datos en clases, ¿cuántas clases deberíamos realizar?, ¿de qué amplitud?
- Realizar la agrupación de los datos en 5 clases y determinar la tabla de frecuencias absolutas, relativas y sus acumuladas. ¿Qué porcentaje de individuos de la muestra cobran un salario menor o igual a 96000 euros brutos anuales?, ¿y entre 64000 y 128000?
- Calcular la media aritmética y la varianza para los datos agrupados en el apartado anterior. Determinar la clase que contiene a la mediana, la clase modal y proporciona sus marcas de clase. Compara la magnitud de las tres medidas de centralización (media, mediana y moda)
- Representar el histograma que resulta de los datos agrupados y comentar sus características más relevantes (unimodal o plurimodal, simétrico o asimétrico, mucha o poca dispersión, posibles atípicos).
- Atendiendo a la forma del diagrama de caja-bigotes y al histograma, ¿qué medidas de centralización y de dispersión te parecen más representativas para este conjunto de datos?

3. a) Sea $\{x_1, x_2, \dots, x_{100}\}$ una muestra de media aritmética 21,5 y sean $x_{101} = 22$, $x_{102} = 19$, $x_{103} = 20,5$ tres observaciones más. Calcular la media aritmética de la nueva muestra $\{x_1, x_2, \dots, x_{100}, x_{101}, x_{102}, x_{103}\}$.

b) Sea \bar{x} la media aritmética de la muestra $\{x_1, x_2, \dots, x_n\}$ y sea \bar{y} la media aritmética de la muestra $\{y_1, y_2, \dots, y_m\}$. ¿Cuál será la media aritmética de la unión de ambas muestras?

4. Probar que la media aritmética y la mediana son operadores lineales, es decir si tenemos una muestra $\{x_1, x_2, \dots, x_n\}$ con media \bar{x} y cada una de las observaciones x_i la multiplicamos por una constante “a” y le sumamos otra constante “b”, es decir obtenemos una nueva muestra $\{y_1, y_2, \dots, y_n\}$ donde $y_i = ax_i + b \quad \forall i = 1, \dots, n$, entonces la media de la nueva muestra \bar{y} , verifica $\bar{y} = a\bar{x} + b$ (igual con la mediana).

5. Sea $\{x_1, x_2, \dots, x_n\}$ una muestra con media \bar{x} y varianza s_x^2 . Si cada una de las observaciones x_i la multiplicamos por una constante “a” (con $a \neq 0$) y le sumamos otra constante “b”, es decir obtenemos una nueva muestra $\{y_1, y_2, \dots, y_n\}$ donde $y_i = ax_i + b \quad \forall i = 1, \dots, n$:

a) ¿Qué relación existe entre la varianza de la $Y (s_y^2)$ y la varianza de la $X (s_x^2)$?

b) ¿Qué relación existe entre la desviación típica de la $Y (s_y)$ y la desviación típica de la $X (s_x)$?