

Matemáticas II

Medidas de dispersión y variabilidad. Las medidas de variabilidad o dispersión tienden a dar una idea de la mayor o menor concentración de los valores de una distribución alrededor de los valores centrales. Consecuentemente, la dispersión de un conjunto de datos hace referencia a la variabilidad entre estos. Si todos los datos son iguales entre sí no existe dispersión, si no lo son hay dispersión. La dispersión es pequeña cuando los valores, aunque diferentes, están próximos entre sí, por el contrario, si están muy separados, la dispersión será grande. Hemos de distinguir entre: a) Medidas de dispersión absolutas. Dan la dispersión en términos absolutos (dependen de las unidades empleadas en la medida de las variables). b) Medidas de dispersión relativas. Dan la dispersión en términos relativos (se expresan como números adimensionales o coeficientes que permiten comparar entre sí índices de dispersión de diferentes distribuciones).

Medidas de dispersión absolutas.

Desviación de un dato respecto de la media. Para cada dato x_i se define como $d_i = x_i - \bar{x}$. Se verifica que $\sum d_i = 0$. Varianza y desviación típica. **Varianza** se define como “La media aritmética de los cuadrados de las desviaciones de los valores de la variable con respecto a la media de la muestra”. No tiene la misma unidad de medida que los datos, lo que es un inconveniente a la hora de comparar con los mismos. **Desviación típica** es “La raíz cuadrada positiva de la varianza”. Tiene la ventaja que tiene la misma unidad de medida que los datos. **Recorrido.** También denominado rango o amplitud total, es la medida de dispersión más simple. Se define como “la diferencia entre el valor máximo y el valor mínimo de la muestra”. $R = x_{\max} - x_{\min}$. **Recorrido intercuartílico.** “Es la diferencia entre el cuartil superior y el cuartil inferior”. Representa la amplitud del intervalo en el que se encuentra el 50% central de la distribución de frecuencias. Cuanto más pequeño sea RI menos dispersión hay respecto de la mediana.

Medidas de dispersión relativa. Coeficiente de variación de Pearson. Se define como “El cociente entre la desviación típica y la media aritmética”. Mide que tamaño tiene la desviación típica con respecto a la media. Importante es que el coeficiente de variación no tiene unidad de medida y permite comparar la variabilidad de distintas variables y la media será más representativa en la muestra en la que el valor de dicho coeficiente es menor. Se suele expresar en porcentaje.

MEDIDA DE SIMETRÍA. Es una medida que indica la mayor o menor simetría de la distribución. Una distribución de frecuencias es simétrica cuando valores de la variable equidistantes de un valor central tienen las mismas frecuencias. Podemos ver si una distribución de frecuencias es simétrica por: a) Histograma de frecuencias absolutas. b) Si los cuartiles inferior y superior equidistan de la mediana. c) Si la media y la mediana coinciden. En caso contrario diremos que es asimétrica, pudiendo ser la asimetría por la derecha o por la izquierda. Diremos que la distribución de frecuencias presenta una asimetría positiva o por la derecha si los valores pequeños de la variable son más frecuentes que los grandes, por lo que la correspondiente representación gráfica presenta una cola alargada hacia la derecha. De forma análoga, se dice que una variable es asimétrica negativa o por la izquierda si los valores más frecuentes son los mayores de la variable, y la representación gráfica tendrá una cola alargada hacia la izquierda. Para calcular la Asimetría se utilizan el coeficiente: $As = \frac{3(\bar{x} - Me)}{s}$ Si se verifica que: $As > 0$, entonces hay asimetría por la derecha o positiva $As = 0$, entonces hay simetría. $As < 0$, entonces hay asimetría por la izquierda o negativa.

GRÁFICO BOX-PLOT (CAJA-BIGOTE) Este gráfico ofrece un resumen de medidas descriptivas, permite estudiar la simetría de la distribución de frecuencias, la existencia de datos atípicos y comparar grupos de datos entre sí. Contiene los siguientes elementos: Un rectángulo o caja que se extiende del primer cuartil al tercer cuartil, por tanto la longitud del rectángulo es el recorrido intercuartílico. Dentro del mismo se marca la mediana. Se consideran dos bigotes, uno a cada lado de la caja, que se extienden desde los cuartiles hasta $Q1 - 1.5(Q3 - Q1)$ y $Q3 + 1.5(Q3 - Q1)$ respectivamente. a) Si la mediana está en el centro del rectángulo la distribución de frecuencias es simétrica. b) Se consideran datos atípicos aquellos que están fuera del intervalo $(Q1 - 1.5(Q3 - Q1), Q3 + 1.5(Q3 - Q1))$.