

Apuntes Unidad 1

Moda, media y mediana

Curso: Probabilidad y estadística descriptiva

Unidad 1 : ¿Qué dicen los gráficos? Análisis crítico de la información

Tema: Medidas de centralidad.

Contenido: Moda, media y mediana

Comencemos revisando el concepto de **diagrama o gráfico de puntos**. El gráfico de puntos es una representación en la que se utilizan **puntos** para representar los datos para **variables numéricas**. Este tipo de gráficos **ayuda a visualizar directamente la frecuencia** de datos de cada valor.

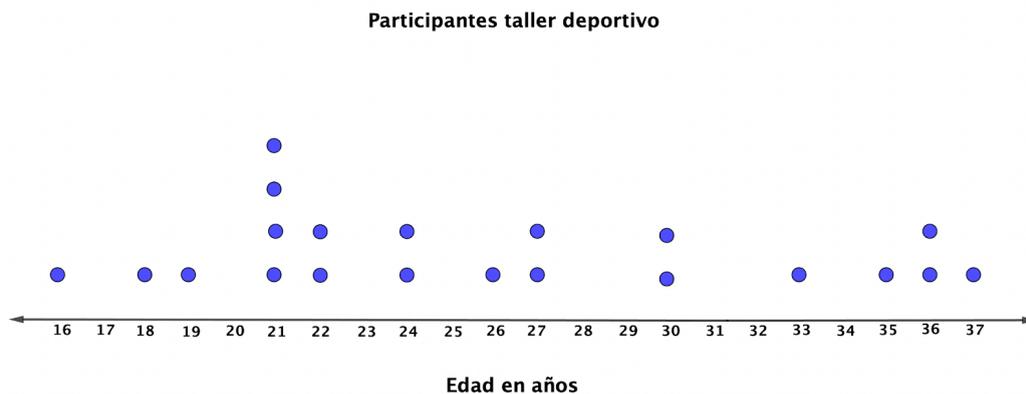


Figura 1: Ejemplo de gráfico de puntos, relacionado a participantes de un taller deportivo y las edades de quienes asisten.

En cuanto a los datos, recordemos que existen valores que nos permiten conocer más sobre la distribución de estos. Esto se conoce como **medidas de tendencia central**, las cuales buscan representar el “centro” de una distribución de datos. Estas medidas son tres: **la moda, la media y la mediana**. En el caso de datos agrupados, se considera un intervalo modal en lugar de la moda.

Cuando los datos están representados en un histograma, la media podemos pensarla como el punto de equilibrio de los datos, en este caso representados por las barras del histograma.

Una característica de la media como medida de tendencia central es que es más sensible a datos que están alejados del centro de la distribución.

Recordemos que la media o promedio, se calcula al **sumar todos los datos y luego dividir esa cantidad por el número de datos**, matemáticamente se tiene la siguiente fórmula.

$$Media = \frac{\text{suma de datos}}{\text{Número de datos}}$$

Curso: Probabilidad y estadística descriptiva

Unidad 1 : ¿Qué dicen los gráficos? Análisis crítico de la información

Tema: Medidas de centralidad.

Contenido: Moda, media y mediana

Por su parte, la **mediana es el valor que se encuentra en la posición central** de un conjunto de datos ordenados de manera creciente. Una interpretación común de la mediana es que corresponde al valor que **divide al conjunto de datos**, ordenados en forma creciente, en dos grupos con igual cantidad de datos. Por tanto, **al menos un 50% de los datos por debajo y al menos el 50% por encima de este valor**

De la misma forma, **cuando los datos están representados en un histograma**, la mediana divide a estos en dos conjuntos con aproximadamente la misma cantidad de datos. Esto se puede interpretar en términos del área del histograma: **la mediana separa las barras del histograma en dos regiones que tienen aproximadamente la misma área, equivalente a la mitad del área de todas las barras.**

La mediana, a diferencia de la media, no se ve afectada por los valores que están más alejados del centro de la distribución.

FORMAS DE DISTRIBUCIONES

Tal y como hemos visto en lecciones anteriores, los histogramas permiten mostrar la distribución que sigue cierta muestra de datos. Por ello, revisaremos algunas formas de distribuciones y la relación que tienen estas con las medidas de tendencia central.

Distribución simétrica

Una distribución es simétrica si la distribución de los datos al lado derecho de la mediana es igual a la del lado izquierdo, pero reflejada con respecto a un eje que pasa por la mediana, denotado en la figura 1 con una línea punteada.

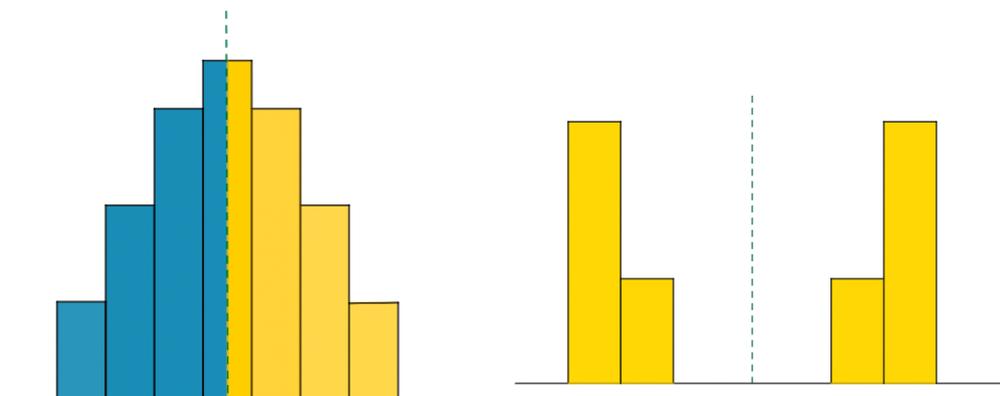


Figura 1: Distribuciones simétricas.

Curso: Probabilidad y estadística descriptiva

Unidad 1 : ¿Qué dicen los gráficos? Análisis crítico de la información

Tema: Medidas de centralidad.

Contenido: Moda, media y mediana

Distribución simétrica con los datos agrupados al centro

Los datos se concentran en torno a un valor central, es decir, las mayores frecuencias se encuentran cerca de ese valor.

Las distribuciones de datos reales que son aproximadamente simétricas también las consideraremos como simétricas.

Por ejemplo, el peso de los estudiantes de un curso expresado en kilogramos.

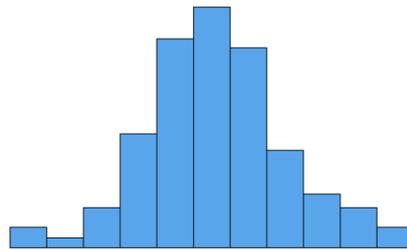


Figura 2: Distribución de datos simétrica con datos concentrados en el centro.

Conociendo estas dos formas de distribuciones, podemos notar que en las simétricas la media y la mediana son iguales. En las distribuciones aproximadamente simétricas, la media y la mediana tienen valores similares. Esto ocurre usualmente cuando se trabaja con datos reales,

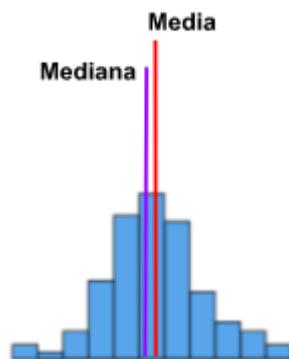


Figura 3: Distribución de datos simétrica con datos concentrados en el centro y la ubicación de la media y mediana.

Curso: Probabilidad y estadística descriptiva

Unidad 1 : ¿Qué dicen los gráficos? Análisis crítico de la información

Tema: Medidas de centralidad.

Contenido: Moda, media y mediana

Distribución asimétrica hacia la izquierda

Diremos que una distribución tiene esta forma cuando las mayores frecuencias están hacia la derecha, pero existen datos alejados del centro que están hacia la izquierda, lo que genera la asimetría con respecto al centro.

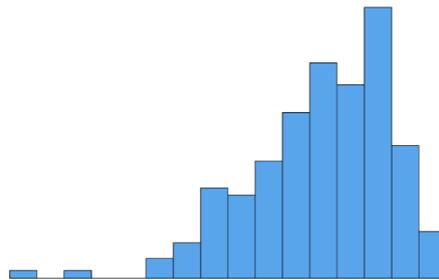


Figura 4: Distribución de datos asimétrica hacia la izquierda

Un ejemplo de situación que puede presentar esta distribución de datos es la estatura de los recién nacidos, por la existencia de bebés que nacen prematuros.

Además, en las distribuciones asimétricas hacia la izquierda la media es menor que la mediana.

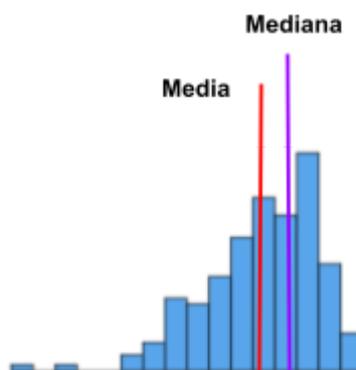


Figura 5: Distribución de datos asimétrica hacia la izquierda con la ubicación de la media y la mediana.

Distribución asimétrica hacia la derecha

Diremos que una distribución tiene esta forma cuando las mayores frecuencias están hacia la izquierda, pero existen datos alejados del centro que están hacia la derecha, lo que genera la asimetría con respecto al centro

Curso: Probabilidad y estadística descriptiva

Unidad 1 : ¿Qué dicen los gráficos? Análisis crítico de la información

Tema: Medidas de centralidad.

Contenido: Moda, media y mediana

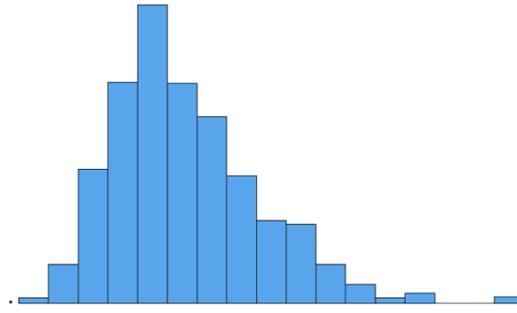


Figura 6: Distribución de datos asimétrica hacia la derecha

Una situación que describe este tipo de distribución son las edades de las mujeres embarazadas. Además, en las distribuciones asimétricas hacia la derecha la media es mayor que la mediana, tal como se muestra en la siguiente imagen.

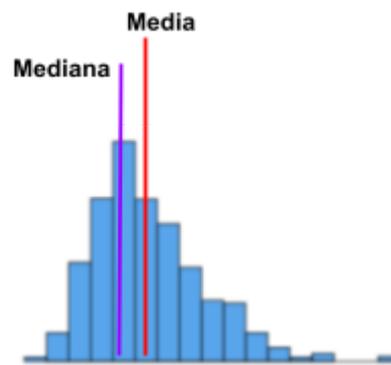


Figura 7: Distribución de datos asimétrica hacia la derecha y la ubicación de la mediana y la media.

Distribución bimodal

Diremos que una distribución tiene esta forma cuando las mayores frecuencias están en torno a dos valores distintos.

Curso: Probabilidad y estadística descriptiva

Unidad 1 : ¿Qué dicen los gráficos? Análisis crítico de la información

Tema: Medidas de centralidad.

Contenido: Moda, media y mediana

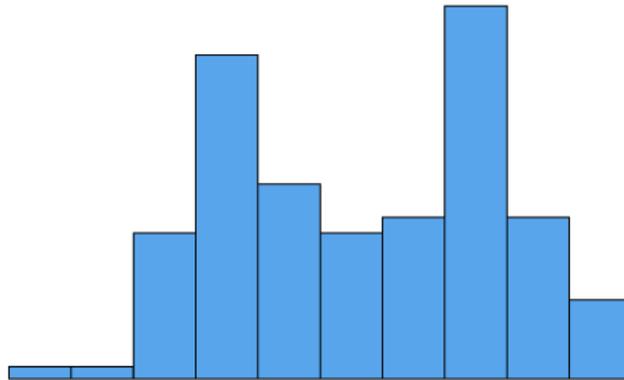


Figura 8: Distribución de datos bimodal.

Esta distribución se puede observar al analizar el flujo de pasajeros en el metro durante el día, ya que hay dos horarios puntas: en la mañana y luego en la salida del trabajo. Finalmente, en las distribuciones bimodales la media y la mediana no se encuentran donde se concentran los datos, sino que entre ambos intervalos modales.

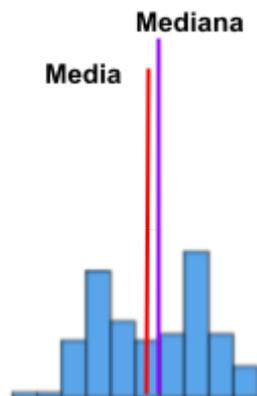


Figura 9: Distribución de datos bimodal junto con la ubicación de la media y la mediana.

Curso: Probabilidad y estadística descriptiva

Unidad 1 : ¿Qué dicen los gráficos? Análisis crítico de la información

Tema: Medidas de centralidad.

Contenido: Moda, media y mediana

SÍNTESIS

- Si se dispone de los datos ya agrupados en intervalos, sólo es posible estimar los valores de la media y la mediana. Si se cuenta con los datos sin agrupar, es preferible calcular la media y la mediana de manera directa y así obtener valores más exactos.
- Cuando los datos están representados en un histograma:
 - La media o promedio podemos pensarla como el punto de equilibrio de los datos, representados por las barras del histograma.
 - La mediana separa las barras del histograma en dos regiones que tienen aproximadamente la misma área, equivalente a la mitad del área de todas las barras.
- La mediana, a diferencia de la media, no se ve afectada por los valores que están más alejados del centro de la distribución.
- Las medidas de tendencia central resultan útiles para responder muchas preguntas estadísticas, pues representan el “centro” de una distribución de datos.Cuál de ellas representa de mejor manera los datos depende del contexto de la situación y también de la forma de la distribución.
- Una **distribución simétrica con datos concentrados al centro** es aquella en que los datos se concentran en torno a un valor central, es decir, las mayores frecuencias se encuentran hacia este valor. En este caso, la media y la mediana son valores muy similares o iguales.
- Una **distribución asimétrica hacia la izquierda** es aquella en que los datos alejados del centro están hacia la izquierda y las mayores frecuencias se encuentran hacia los valores de la derecha. En este caso, en general, la media es menor a la mediana.
- Una **distribución asimétrica hacia la derecha** es aquella en que los datos alejados del centro están hacia la derecha y las mayores frecuencias se encuentran hacia los valores de izquierda. En este caso, en general, la media es mayor a la mediana.
- Una distribución **bimodal** es aquella en que los valores están en torno a dos valores distintos.
- Es importante tener en cuenta que al trabajar con datos reales en general las distribuciones son aproximadamente simétricas, asimétricas o bimodales