SUMA Y SIGUE

MATEMÁTICA EN LÍNEA

MATERIAL PEDAGÓGICO COMPLEMENTARIO

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

CMM Centro de Modelamiento Matemático

Ministerio de Educación

Gobierno de Chile

MATERIAL PEDAGÓGICO
COMPLEMENTARIO

FICHAS TALLER 4:
MEDIDAS DE RESUMEN

INTRODUCCIÓN

A partir de diversas situaciones, en este taller hemos estudiado las medidas de tendencia central y de posición relativa, con énfasis en su interpretación y en su utilidad para describir la distribución de los datos.

Los contenidos abordados en las fichas son los siguientes:

- Medidas de tendencia central: moda, media y mediana.
  - Definiciones.
  - Interpretaciones.
  - Cálculo en datos no agrupados.
  - Estimación en datos agrupados.
  - Comportamiento de la media y la mediana frente a datos extremos.

- Medidas de posición relativa: cuartiles, quintiles, deciles y percentiles.
  - Definiciones.
  - Interpretación y uso de medidas de posición relativa para datos no agrupados.
  - Construcción de diagrama de caja y bigotes o boxplot.
  - Cálculo de percentiles para datos no agrupados.
  - Estimación de percentiles para datos agrupados.

1. Moda

La ***moda*** de un conjunto de datos es el valor que se presenta con mayor frecuencia.

Por ejemplo, en la edad de ingreso a la universidad en una carrera determinada la moda es 18, edad que se observó con mayor frecuencia.

Edad	Frecuencia
17	7
18	13
19	11
20	2

Si en una distribución de datos hay dos valores cuya frecuencia es máxima, decimos que la distribución es ***bimodal*** y a cada una de ellas la llamamos moda. En caso de que existan más de dos modas, hablaremos de una distribución ***multimodal***.

Cuando todos los valores de una distribución tienen la misma frecuencia, no tiene sentido hablar de moda.

Comentarios

- La moda puede utilizarse tanto para variables cuantitativas como cualitativas, y si bien es considerada una medida de tendencia central, no necesariamente responde a la idea de "centro". Esto significa que es posible que la moda de una distribución se encuentre en alguno de los extremos.

- Notemos, además, que la moda siempre coincide con uno de los datos, y que, como regla general, es útil reportarla cuando su frecuencia dista notablemente de la del resto de los valores.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: La campaña por el aire puro.

2. Media o promedio

La ***media*** o ***promedio*** de un conjunto de datos es una medida de tendencia central que admite las siguientes interpretaciones:

- En términos de un ***reparto equitativo***: Esta interpretación sugiere juntar (sumar) los valores de las observaciones como si fueran elementos, y luego repartirlos equitativamente (dividirlos) entre todas ellas.
- En términos de ***equilibrio de la suma de distancias***: Esta interpretación sugiere considerar el promedio como el único valor tal que la suma de las distancias desde sí mismo a los datos menores que él es igual a la suma de las distancias desde sí mismo a los datos que son mayores.

Su valor se obtiene al sumar todos los datos y luego dividir esa cantidad por el número de ellos. Es decir, el promedio de un conjunto de n datos x_1 , x_2 , ..., x_n es:

$$\frac{x_1 + \ldots + x_n}{n}$$

Comentarios

- El promedio solo tiene sentido para variables cuantitativas y no necesariamente coincide con alguno de los datos.
- Es importante considerar que cuando se trabaja con elementos no particionables, la interpretación del promedio como reparto equitativo puede no ser tan directa, ya que podrían resultar cantidades no enteras.
- Por el contrario, la interpretación del promedio como igualdad de suma de distancias es útil en cualquier contexto, puesto que no requiere que los elementos a promediar sean particionables ni que el valor que toma el promedio sea un número entero.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: La campaña por el aire puro.

3. Utilidad del promedio

Conocer el promedio de un conjunto de datos puede entregar información valiosa sobre ellos. Algunas de sus utilidades son:

- **Resume información**. Por ejemplo, el promedio de notas de un curso en una prueba contribuye a caracterizar el rendimiento de sus estudiantes.
- **Constituye una referencia**, pues al compararlo con un dato particular, permite caracterizar a este último. Por ejemplo, podemos comparar la nota de un estudiante con el promedio del curso para establecer cómo le fue respecto de sus compañeros.
- **Permite realizar comparaciones**. Podemos, por ejemplo, comparar los promedios de notas de dos cursos en una misma prueba para establecer en cuál de ellos los estudiantes tuvieron mejor desempeño.

Comentarios

- Es importante que los docentes planteen situaciones en las que sus estudiantes identifiquen las distintas formas de utilizar el promedio.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: La campaña por el aire puro.

4. Mediana

La ***mediana*** de un grupo de datos es una medida de tendencia central correspondiente al valor que representa la posición central en la lista ordenada de estos.

Para calcular su valor cuando hay n datos, estos se deben ordenar de manera creciente o decreciente. Se distinguen dos casos:

- Si n es impar: La mediana es el valor del dato que queda exactamente al medio, es decir, es el que ocupa la posición $\frac{(n+1)}{2}$ en la lista ordenada.
- Si n es par: En este caso no existe una posición central en la lista y, por convención, se considera como mediana el promedio de los dos datos centrales, es decir, aquellos que ocupan la posición $\frac{n}{2}$ y la posición $\frac{n}{2}+1$.

Por ejemplo:

- En el caso de los datos ordenados 1 - 4 - 5 - 6 - 8 - 8 - 10, se tienen 7 datos, por lo tanto, el valor que está justamente al centro, es decir, el de la posición $\frac{(7+1)}{2} = 4$, corresponde al valor 6.
- En el caso de los datos ordenados 1 - 1 - 4 - 5 - 6 - 8 - 8 - 10, se tienen 8 datos, por lo que la mediana se obtiene al promediar los valores de las posiciones $\frac{8}{2} = 4$ y $\frac{8}{2} + 1 = 5$, es decir, al promediar los valores 5 y 6. Así la mediana es igual a 5,5.

Comentarios

- La mediana tiene sentido solo en el caso de variables cuantitativas, y su valor no siempre coincide con alguno de los datos.
- Basta con ordenar la lista de datos hasta la posición $\frac{n}{2}+1$ para determinar el valor de la mediana en cualquier conjunto de datos.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: Terminando el sudoku.

5. Caracterización de la mediana

De acuerdo al uso que se le quiera dar a la mediana, surgen definiciones alternativas que podrían arrojar valores levemente diferentes. Por ejemplo:

- Valor que satisface la condición de que al menos el 50% de los datos es menor o igual que él, y al menos el 50% de los datos es mayor o igual que él.

- Valor que divide al conjunto de datos, ordenados en forma creciente o decreciente, en dos grupos de igual tamaño: los listados antes de la mediana, y los listados después de la mediana.

Comentarios

- Ambas maneras de definir la mediana tienen como propósito dividir el conjunto de datos en dos grupos de tamaño similar. En la primera de ellas, la mediana se considera como parte de ambos grupos, por lo que no necesariamente estos son del mismo tamaño. En cambio, en la segunda, los grupos no incluyen a la mediana y, por lo tanto, son del mismo tamaño.

- Cuando la cantidad de datos es par, existe más de un valor que satisface la primera definición de mediana. En tal caso, se suele adoptar la convención de tomar el promedio de los valores de las posiciones centrales para elegir un valor para la mediana.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: Terminando el sudoku.

6. Datos agrupados

Se dice que los datos están ***agrupados*** cuando no se observan uno a uno los valores de estos, pero sí las clases o intervalos asociados a la variable y la frecuencia de cada intervalo. El largo de un intervalo se conoce como ***amplitud***.

Comentarios

- En general, disponer de datos agrupados resulta especialmente útil cuando se trabaja con variables cuantitativas que pueden tomar muchos valores.
- Una vez que se decide agrupar datos, definir la amplitud de los intervalos puede depender del contexto y de la distribución de los datos. Algunos criterios para determinarla pueden estar asociados a la cantidad de intervalos con los que se desea trabajar, a la habilidad de manejar gran cantidad de datos, o al interés por distinguir determinados valores en la distribución.
- En muchos casos conviene definir la misma amplitud para todos los intervalos, ya que de esa manera se evitan distorsiones en la percepción de la distribución de los datos representados.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: ¿Cuánto duran las baterías?

7. Clase modal

Dado que en el caso de datos agrupados se desconoce la distribución de los datos dentro de cada intervalo, no es posible identificar un dato en particular que se repita más que los demás. Lo que sí se puede determinar es el o los intervalos que presentan la mayor frecuencia, los que se conocen como ***intervalos modales*** o ***clases modales***.

Comentarios

- Por lo general, la clase modal suele usarse como representante de los datos cuando su frecuencia dista notablemente de la del resto de las clases.
- El valor que presenta la mayor frecuencia de los datos no agrupados, es decir, la moda, no necesariamente se encuentra en la clase modal de los mismos datos cuando estos se han agrupado.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: ¿Cuánto duran las baterías?

8. Clase mediana

El ***intervalo mediano*** o ***clase mediana*** es aquel en el que se encuentra el valor de la mediana. Este corresponde al primer intervalo en el que la frecuencia acumulada iguala o supera la mitad del número total de datos.

Por ejemplo, en la cantidad de horas a la semana que 26 compañeros practican deportes, la clase mediana corresponde al intervalo “6 - 8”.

Cantidad de horas	Frecuencia	Frecuencia acumulada
0 - 2	8	8
2 - 4	3	11
4 - 6	0	11
6 - 8	10	21
8 - 10	5	26

Clase en la que la frecuencia acumulada iguala o supera a $\frac{26}{2} = 13$

Comentarios

- La clase mediana nunca coincide con una clase que tiene frecuencia 0.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: ¿Cuánto duran las baterías?

9. Estimación de la mediana para datos agrupados

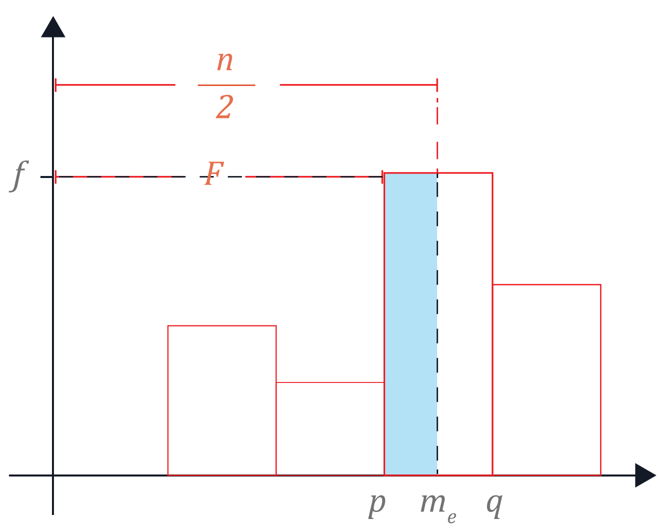
Cuando se trabaja con datos agrupados, no siempre es posible determinar el valor exacto de la mediana, por lo que debe ser estimado. Una manera usual de estimar la mediana es la que se describe en el siguiente procedimiento:

- Paso 1: determinar la clase mediana, cuyo extremo inferior es p.
- Paso 2: identificar la frecuencia f de la clase mediana y la frecuencia acumulada F hasta la clase anterior.
- Paso 3: evaluar la mediana a través de la siguiente expresión:

$$mediana = p + \left(\frac{\frac{n}{2} - F}{f} \right) \cdot a$$

donde a corresponde a la amplitud de a la clase mediana.

Lo anterior se puede observar gráficamente:



$$\left(\frac{m_e - p}{a} \right) f = \frac{n}{2} - F$$

Comentarios

- En esta forma de estimar la mediana se asume que los datos en cada clase están dispuestos de manera pareja, de modo que es posible aplicar un razonamiento proporcional que sustenta la expresión anterior.

- Lograr la comprensión de los procesos y razonamientos que hay detrás de esta expresión tiene más relevancia que memorizarla como fórmula. Para ello, puede ser útil ver diversos ejemplos numéricos y visualizar en ellos la relación de proporcionalidad.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: ¿Cuánto duran las baterías?

10. Marca de clase

Una manera usual de estimar algunas medidas de resumen a partir de datos agrupados se basa en tomar un valor que represente a cada clase, denominado ***marca de clase***.

Por convención, se suele considerar la marca de clase como el punto medio del intervalo.

Comentarios

- Una de las razones para elegir el punto medio es que este se puede interpretar como el valor que tomaría el promedio de los datos en ese intervalo si estos estuvieran dispuestos de forma pareja.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: ¿Cuánto duran las baterías?

11. Estimación del promedio para datos agrupados

El procedimiento usual para estimar el promedio a partir de datos agrupados es el siguiente:

- Paso 1: determinar la marca de clase de cada intervalo.
- Paso 2: calcular el producto entre la marca de cada clase y su respectiva frecuencia absoluta.
- Paso 3: sumar los productos.
- Paso 4: dividir el resultado por el número total de datos.

Lo anterior se resume en la siguiente expresión:

$$promedio = \frac{m_1 \cdot f_1 + \cdots + m_k \cdot f_k}{n}$$

donde m_1 , m_2 , ..., m_k son las marcas de clase, f_1 , f_2 , ..., f_k sus frecuencias absolutas y n es el total de datos.

Comentarios

- Nuevamente para estimar el valor de la media en datos agrupados se asume que los datos en cada clase están dispuestos de manera pareja.
- Notemos, además, que el promedio de datos agrupados puede interpretarse de dos maneras:
  - como un promedio en el que se ha reemplazado cada dato por su respectiva marca de clase, y por lo tanto, estas se repiten tantas veces como la frecuencia lo indica, o
  - como un promedio ponderado en el que se ponderan las marcas de clase por sus frecuencias respectivas.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: ¿Cuánto duran las baterías?

12. Datos extremos o atípicos

Cuando un valor es muy pequeño o muy grande en comparación con el grueso de los datos, decimos que se trata de un dato ***extremo*** o ***atípico***.

En general, la cantidad de datos extremos es pequeña comparada con el número total de ellos, y si bien en muchos casos aparecen debido a errores en la medición o en el registro, estos también se pueden deber simplemente a la distribución natural de los datos.

Comentarios

- Para determinar si un valor se puede considerar como dato extremo existen convenciones estadísticas. Sin embargo, durante este curso, solo se utilizaron nociones intuitivas que se apoyan en la observación de la distribución de los datos.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: ¿Moda, media o mediana?

13. Comportamiento de la media y mediana frente a datos extremos

Una de las diferencias entre la media y la mediana es su comportamiento frente a datos extremos o atípicos.

La media, en general, es más sensible a la presencia de datos extremos, esto es, su valor se altera significativamente al calcularla descartando o no estos valores. Si en el conjunto de datos existe un par de ellos cuyo valor es mucho más grande que los restantes, la media obtenida al incluir estos datos extremos puede ser significativamente mayor que gran parte de los datos, y por tanto deja de ser representativa de estos.

Por otro lado, la mediana no es tan sensible a los valores extremos, ya que depende de la posición y no del valor de los datos. No importa qué tan grandes o pequeños sean los datos extremos, la mediana no se verá alterada significativamente.

Por ejemplo, de los siguientes precios de un medicamento registrados en 15 farmacias, se pueden considerar como datos extremos los valores $31.590, $46.990 y $37.990.

Precio medicamento ($)				
17.390	31.590	18.550	17.190	37.990
17.990	17.290	17.590	19.590	17.490
18.050	17.690	46.990	17.690	17.190

Así, se obtienen las siguientes medidas:

	Todos los datos	**Sin datos extremos**
Media	$ 22.018	$ 17.808
Mediana	$ 17.690	$ 17.640

Comentarios

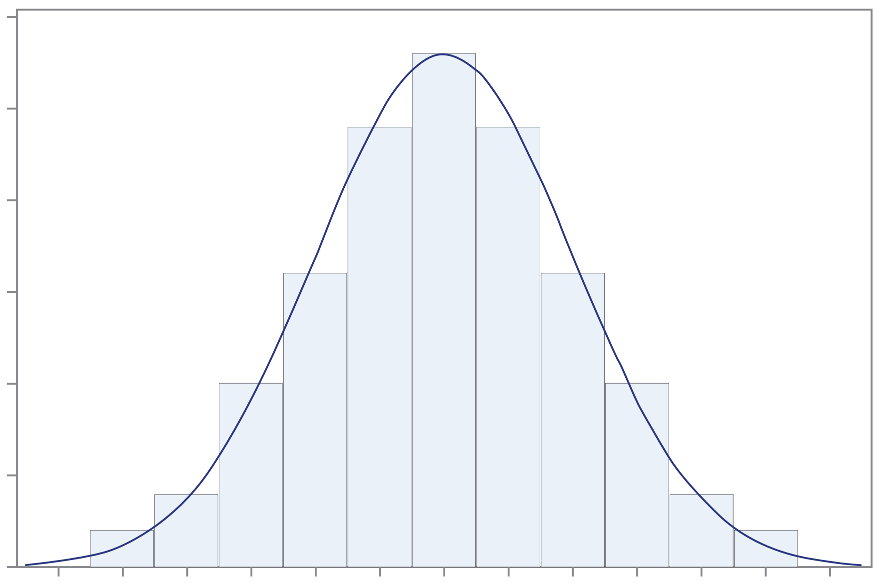
- Si bien la mediana es más robusta que la media frente a la presencia de datos extremos, ambas medidas reflejan características distintas de los datos y, por lo tanto, es posible que para satisfacer los objetivos del estudio, sea conveniente reportar la media.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: ¿Moda, media o mediana?

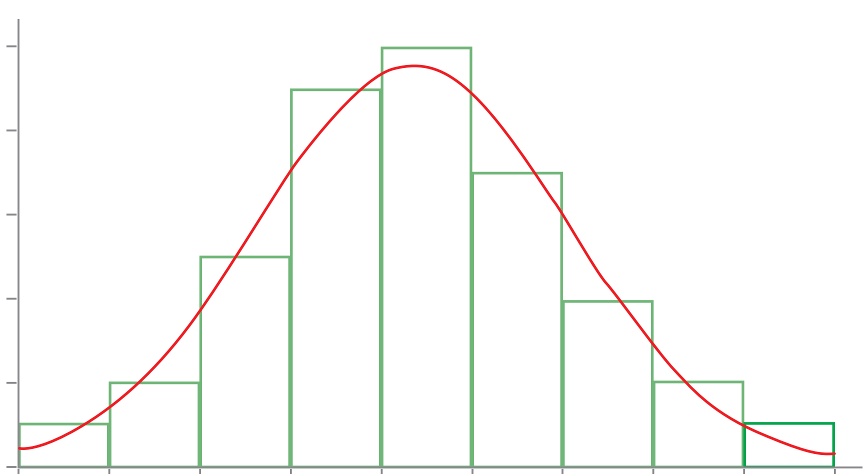
14. Distribución normal

Una de las distribuciones más comunes es la ***distribución normal***, que por su apariencia también se conoce como campana de Gauss. Su simetría con respecto a un eje vertical es una de las características principales de este tipo de distribución y, como consecuencia, la moda, la media y la mediana son iguales.



Comentarios

- Existen muchos fenómenos de la naturaleza en los que aparecen distribuciones que se asemejan a una normal. Por esta misma razón, esta distribución es muy usada para modelar distintas situaciones.



Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: ¿Moda, media o mediana?

15. Cuartiles

Los ***cuartiles***, denominados primer, segundo y tercer cuartil, son valores que, al tener los datos ordenados de menor a mayor, cumplen las siguientes condiciones:

- El ***primer cuartil*** (Q_1) es un valor tal que al menos el 25% de las observaciones son menores o iguales que él y al menos el 75% son mayores o iguales que él.
- El ***segundo cuartil*** (Q_2) es un valor tal que al menos el 50% de las observaciones son menores o iguales que él y al menos el 50% son mayores o iguales que él.
- El ***tercer cuartil*** (Q_3) es un valor tal que al menos el 75% de las observaciones son menores o iguales que él y al menos el 25% son mayores o iguales que él.

Los cuartiles son valores que, en general, dividen al conjunto de datos en 4 grupos con aproximadamente la misma cantidad de observaciones cada uno.

Comentarios

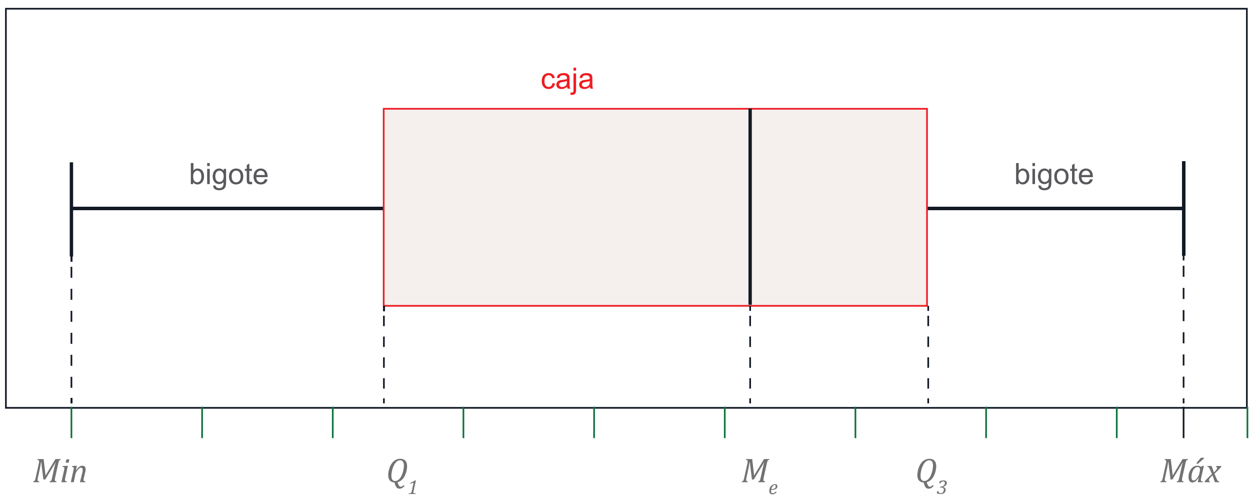
- La definición de cuartiles dada permite que, en algunos casos, exista más de un valor que cumple con las condiciones para un determinado cuartil. En esos casos, se adopta la convención de tomar el promedio de las observaciones entre las que se encuentran los valores que cumplen con la condición.
- En otras ocasiones, puede ser útil elegir como cuartil la primera observación en el conjunto de datos ordenados que cumple con tal condición.
- Notemos que el valor del segundo cuartil corresponde al de la mediana.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: ¿Dónde deben cargar bencina los vehículos de Rayotaxi?

16. Diagrama de caja y bigotes

El ***diagrama de caja y bigotes***, también conocido como ***boxplot***, es una representación gráfica que permite visualizar ciertas características de la distribución de los datos. Se construye a partir del valor mínimo, primer cuartil, mediana, tercer cuartil y el valor máximo, tal como se muestra en el siguiente ejemplo:



En este diagrama la caja representa al conjunto de datos comprendidos entre el primer y tercer cuartil, es decir, al 50% central de los datos. Los bigotes, por su parte, representan a los datos que están entre el valor mínimo y el primer cuartil, y entre el tercer cuartil y el valor máximo.

Comentarios

- La información que entrega el boxplot complementa a la que se obtiene de histogramas o gráficos de puntos, pues permite visualizar de forma resumida la distribución de los datos.
- Específicamente, el boxplot permite observar las siguientes características de la distribución de datos:

  - El lugar que ocupa la mediana dentro del rango de las observaciones.
  - Qué tan alejados están los cuartiles de la mediana, valor mínimo y valor máximo.
  - En qué rango de valores se encuentra el 50% central de las observaciones.

Ubicación: Módulo 2

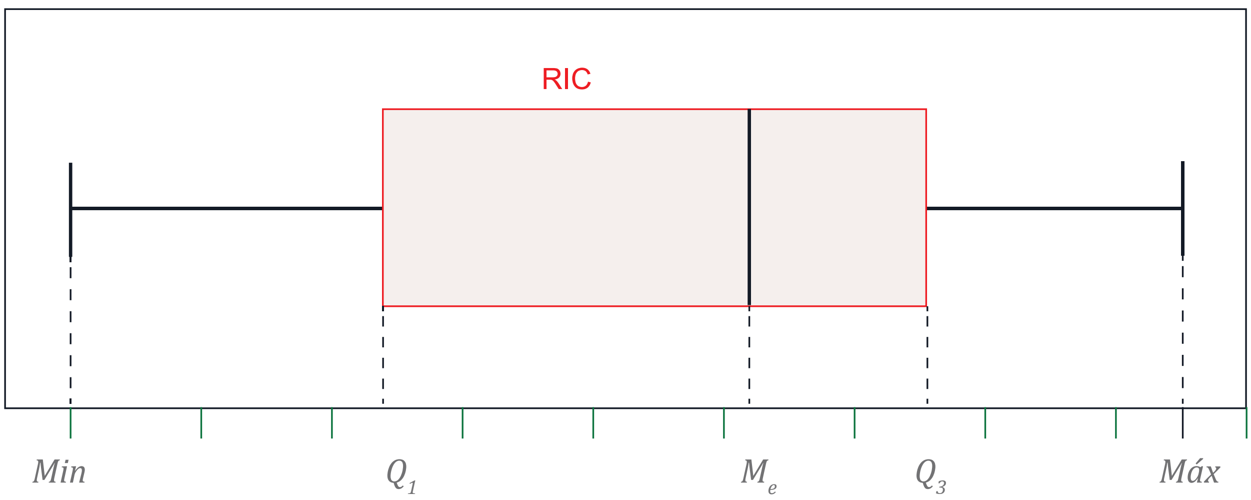
Taller: Medidas de resumen.
Actividad: ¿Dónde deben cargar bencina los vehículos de Rayotaxi?

17. Rango intercuartil

El ***rango intercuartil*** (RIC) es una medida de la variabilidad o dispersión de un conjunto de datos. Corresponde a la distancia entre el primer y el tercer cuartil, es decir:

$$\text{RIC} = Q_3 - Q_1$$

El RIC es el valor máximo que podría alcanzar la diferencia entre dos valores cualesquiera del 50% central de los datos. En el boxplot, el RIC es igual al largo de la caja.



Comentarios

- Existen muchas medidas que pueden ser usadas para indicar la dispersión de los datos. Además del RIC y del rango, existe otra medida conocida como desviación estándar, la que se estudia en cursos posteriores.
- El RIC tiene la ventaja de ser una medida más robusta que el rango frente a la presencia de datos extremos. Esto significa que no sufre modificaciones significativas ante la presencia o ausencia de este tipo de datos, al contrario del rango, el que sí se altera considerablemente.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: ¿Dónde deben cargar bencina los vehículos de Rayotaxi?

18. Procedimiento para el cálculo de cuartiles

Para calcular el primer cuartil de un conjunto ordenado de datos se puede realizar el siguiente procedimiento:

- Determinar el valor k, que corresponde al 25% del número de datos.

- Dependiendo de k, se tienen dos casos:
    - Si k es un número entero, entonces tomaremos como primer cuartil al promedio de los datos en las posiciones k y (k + 1).

    - Si k no es un número entero, entonces el primer cuartil es el dato en la posición que corresponde al entero inmediatamente superior a k .

El mismo procedimiento aplica para determinar el segundo y el tercer cuartil con los porcentajes respectivos.

Posición de cada dato	1	2	3	4	5	6	7	8	9	10
Datos	45	48	50	51	52	55	55	60	61	62

- El 25% de 10 es k = 2.5, por tanto Q_1 es el valor en la posición 3, es decir, Q_1 = 50.

- El 50% de 10 es k = 5, por tanto Q_2 es el promedio de los datos en las posiciones 5 y 6, es decir, Q_2 = 53.5.

- El 75% de 10 es k = 7.5, por tanto Q_3 es el valor en la posición 8, es decir, Q_3 = 60.

Comentarios

- Si bien es importante que los estudiantes ejerciten el procedimiento para el cálculo de los cuartiles en distintas situaciones, es necesario que el docente enfatice la comprensión conceptual de estas medidas y sus interpretaciones dentro del contexto.

- Cabe mencionar que el procedimiento acá descrito no es el único que puede ser usado para determinar los cuartiles. En efecto, existen variantes de él que dependen del orden en que se definen los cuartiles y de las convenciones que se adopten.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: ¿Dónde deben cargar bencina los vehículos de Rayotaxi?

19. Quintiles

En estadística, los quintiles corresponden a 4 valores que dividen a la lista de datos ordenados en 5 grupos de aproximadamente el mismo tamaño, es decir, con cerca del 20% de los datos cada uno.

Tal como en el caso de los cuartiles, se puede precisar que:

- El **primer quintil** satisface que al menos el 20% de los datos es menor o igual que él, y al menos el 80% de los datos es mayor o igual que él.
- El **segundo quintil** satisface que al menos el 40% de los datos es menor o igual que él, y al menos el 60% de los datos es mayor o igual que él.
- El **tercer quintil** satisface que al menos el 60% de los datos es menor o igual que él, y al menos el 40% de los datos es mayor o igual que él.
- El **cuarto quintil** satisface que al menos el 80% de los datos es menor o igual que él, y al menos el 20% de los datos es mayor o igual que él.

Es posible que haya más de un valor que satisfaga la definición anterior para algún quintil en particular. En tal caso, se debe adoptar una convención que permita elegir uno de ellos como quintil.

Comentarios

- Según la definición dada, los quintiles corresponden a los valores que permiten separar los datos en cinco grupos de aproximadamente el mismo tamaño. Sin embargo, hay que tener en cuenta que en economía es usual que el término “quintiles de ingreso” se refiera a los cinco grupos de la población determinados por los quintiles 1 a 4.
- Por ejemplo, es común escuchar en las noticias afirmaciones tales como *“En el mes de mayo, el segundo quintil de los chilenos tuvo un ingreso familiar per cápita promedio de $165.300”*, haciendo referencia al grupo de individuos y no a los valores de los quintiles propiamente tales.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: Posicionando el ingreso.

20. Deciles y percentiles

Existen otras medidas de posición que se usan bastante en Ciencias Sociales y que se conocen como deciles y percentiles.

Los deciles son 9 valores que permiten dividir a un grupo de datos, ordenados de menor a mayor, en 10 grupos de aproximadamente el mismo tamaño, es decir, con alrededor de un 10% de los datos cada uno.

De forma análoga, los percentiles son 99 valores que permiten dividir a un grupo de datos, ordenados de menor a mayor, en 100 grupos de aproximadamente el mismo tamaño, es decir, con alrededor de un 1% de los datos.

Comentarios

- Observemos que hay una correspondencia entre distintas medidas de posición. Por ejemplo, la mediana corresponde al segundo cuartil y, a su vez, al decil 5 y al percentil 50. De igual forma, el quintil 3 corresponde al decil 6 y al percentil 60.
- Es importante considerar que el uso de percentiles como medida de posición tiene sentido cuando se trabaja con grandes cantidades de datos.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: Posicionando el ingreso.

21. Cálculo de quintiles y deciles

Para calcular los quintiles y deciles se puede usar un procedimiento similar al utilizado para determinar los cuartiles. La diferencia está en que en el caso de quintiles se divide el número total de datos por 5, y en el caso de los deciles, por 10.

Luego de obtener los quintiles y deciles, y especialmente cuando hay datos repetidos, se debe adoptar una convención *ad hoc* que indique en cuál de los grupos se deben incluir los datos que coinciden con el quintil o decil respectivo.

Comentarios

- Para el cálculo de las medidas de posición relativa, es fundamental que los estudiantes comprendan los razonamientos que sustentan el procedimiento usado para el cálculo de los cuartiles. Esto les permitirá extender dichas ideas al resto de los cálculos.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: Posicionando el ingreso.

22. Procedimiento para el cálculo de percentiles para datos no agrupados

El procedimiento para el cálculo del percentil p puede resumirse en los siguientes pasos:

- Ordenar los n datos de menor a mayor.
- Determinar el p% del número de datos. Esto es, $k = \frac{p}{100} \cdot n$
- Se distinguen dos casos:
    - Si k es un número entero, entonces el percentil p es el dato en la posición k.
    - De lo contrario, el percentil se toma como el dato en la posición correspondiente al número entero siguiente a k.

Notemos que hemos utilizado un nueva convención para seleccionar el valor del percentil: tomar el valor de la primera observación que iguala o supera al p% de los datos.

Comentarios

- Los procedimientos de cálculo son importantes en el proceso de aprendizaje, ya que son necesarios para resolver problemas, pero además porque a menudo contribuyen a lograr una mayor comprensión de los conceptos. Debido a esto, es recomendable ejercitar los procedimientos, pero, a su vez, trabajar sobre los argumentos que los sustentan, para de esta forma evitar la mecanización de estos.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: Posicionando el ingreso.

23. Estimación de percentiles en datos agrupados

Cuando hay datos agrupados, en la mayoría de los casos no es posible determinar el valor exacto de un percentil, por lo que se hace necesario estimarlo. Para estimar el percentil p se puede proceder de la siguiente forma:

- Paso 1: se calculan las frecuencias porcentuales acumuladas de cada clase.
- Paso 2: se ubica la primera clase cuya frecuencia porcentual acumulada iguala o supera al p %.
- Paso 3: se procede a estimar el valor del percentil p dentro de los valores de esa clase.

En el último paso, la estimación del percentil p puede realizarse usando supuestos sobre la manera en que los datos están dispuestos en dicha clase.

Ejemplo: Estimación del percentil 75 de las edades de los empleados de una empresa agrupados en la siguiente tabla:

Edad (años)	Frecuencia	Frecuencia porcentual	Frecuencia porcentual acumulada
18 - 25	45	15%	15%
25 - 30	75	25%	40%
30 - 45	120	40%	80%
45 - 60	60	20%	100%
Total datos	**300**		

- La primera clase cuya frecuencia porcentual acumulada supera al 75% es la clase de 30 a 45 años.

- Si el criterio es estimar el percentil como el valor del extremo de la clase que esté más cercano al 75%, entonces el percentil 75 se estimaría en 45 años, pues 75% está más cerca de 80% que de 40%.

Comentarios

- Como ya se mencionó, cuando los datos están agrupados, solo se puede aspirar a lograr una aproximación de las medidas de posición. La estimación también puede realizarse, aunque los datos no estén agrupados, cuando no sea tan relevante determinar el valor exacto de una medida de posición. Debido a esto, es útil que los estudiantes manejen procedimientos o criterios que les permitan realizar buenas estimaciones.

Ubicación: Módulo 2

Taller: Medidas de resumen.
Actividad: Posicionando el ingreso.