

Análisis Cuantitativo I

Diseño, Medición y Medidas de Tendencia Central

Carlos Cardona

Universidad del Rosario

3 de febrero de 2017

Quiz

Indique si la siguiente afirmación es verdadera o falsa. Si es falsa justifique su respuesta.

- **Tema A:** El diferencial de salarios entre nativos e inmigrantes es un constructo de la definición operacional *prejuicio*.
- **Tema B:** La estadística descriptiva busca alcanzar un estadístico a partir de un parámetro.

Repaso

- El número diario promedio de mensajes de texto que envían los colombianos es un parámetro o un estadístico?
- Si cada uno elige su programa favorito, esta variable sería nominal, ordinal o de razón?
- La edad es una variable discreta o continua?
- A la hora de graficar una variable de escala ordinal, cuál(es) de los tipos de gráfica vistos en la clase anterior sería(n) viables?

Diseño de Investigación

- La mayoría de fenómenos que estudian las ciencias sociales tienen múltiples causas.
- Las teorías, en cambio, se enfocan en una de las causas e ignoran las demás.
- La naturaleza multivariada de los conceptos puede llevar a conclusiones engañosas.
- Por ejemplo, algunos estudios han mostrado que la raza (X) en EE.UU está relacionada con la participación política (Y). Donde los anglosajones participan más que las demás razas.

- La conclusión podría ser errónea si se lleva a cabo una comparación más compleja. Personas de diferentes razas (X) tienen condiciones socioeconómicas diferentes (Z), las cuales se asocian tanto con la raza y también afectan la participación política.
- Las *comparaciones* están en la esencia de la ciencia. Si evaluamos una teoría que relaciona X y Y, la labor científica es hacer todo lo posible para asegurarse que no existen otras influencias (Z) que interfieran con las comparaciones realizadas.
- La variable X se conoce como la *variable independiente*, aquella que es denotada como causa. Por otro lado, la variable Y es la *dependiente*.

- Existen diferentes estrategias, o diseños investigativos, que pueden usar para encontrar conclusiones, creíbles y no influenciadas por factores externos, de ciertas teorías.
- Las dos técnicas más usadas y efectivas son:
 - 1 Los experimentos.
 - 2 Los estudios observacionales.

El Diseño Experimental

- Supongan que son un candidato a un cargo político. Además, tienen aún presupuesto para usarlo en publicidad que contraste sus habilidades con las del oponente.
- La pregunta sería: ¿La publicidad funcionará con los votantes?¹
- La aproximación estándar para una situación como esta es conducir un *experimento*.

¹Wattenberg, Martin P. & Craig Leonard Brians. 1999. "Negative Campaign Advertising: Demobilizer or Mobilizer?" *American Political Science Review*.

El Diseño Experimental

Definición

Un experimento es un diseño investigativo en el cual, el investigador controla y asigna aleatoriamente valores de la variable independiente a los participantes.

- Los participantes del experimento serán divididos de manera *aleatoria* a un **grupo tratamiento** y a un **grupo control**.
- ¿Por qué es importante la asignación aleatoria?

El Diseño Experimental

Definición

Un experimento es un diseño investigativo en el cual, el investigador controla y asigna aleatoriamente valores de la variable independiente a los participantes.

- Los participantes del experimento serán divididos de manera *aleatoria* a un **grupo tratamiento** y a un **grupo control**.
- ¿Por qué es importante la asignación aleatoria?
- La asignación aleatoria descontamina la comparación entre ambos grupos de otras influencias.

Características e Inconvenientes de un Diseño Experimental

- Validez Interna-¿Qué tan confiables son las conclusiones?
- Tipos de Experimentos:
 - 1 Encuesta Experimental.
 - 2 Experimento de Campo.
 - 3 Experimentos Naturales.
- Desventajas:
 - 1 No toda variable independiente es manipulable.
 - 2 Bajos niveles de validez externa- Muestreo por conveniencia y el estímulo.
 - 3 Aspectos éticos.
 - 4 X es la principal causa?

Estudios Observacionales

- La implementación de un experimento a menudo resulta ser inviable, y otras veces imposible.
- Como resultado de esto, la experimentación no es el diseño más común entre las ciencias sociales.
- En casos donde no se puede controlar la exposición a la variable de interés, la única elección es observar al mundo como si ya existiera y hacer las comparaciones entre *individuos* o entre cantidades agregadas que varían en el tiempo.
- Los estudios observacionales no son experimentos, pero tratan de emularlos.

Estudios Observacionales

Definición

Un estudio observacional es un diseño donde el investigador no tiene control sobre los valores de la variable independiente, la cual ocurre naturalmente.

- Sin embargo, es importante que exista variabilidad tanto en la variable dependiente como la independiente.
- Debido a la incapacidad de manipular la variable independiente, algunos académicos los denominan estudios correlacionales.
- Tipos de estudios observacionales:
 - 1 Corte Transversal
 - 2 Corte Longitudinal
 - 3 Series de Tiempo

La medición

- La medición es un “problema” para todas las ciencias.
- Para las ciencias físicas, el problema de la medición se reduce a la instrumentación, bajo el cual los científicos desarrollan protocolos para medir ciertos fenómenos.
- Las ciencias sociales, por el contrario, rara vez tienen consenso sobre cómo medir sus conceptos importantes.
- Más importante es que las ciencias sociales lidian con un objeto difícil de predecir: el ser humano.

- Es un error pensar que para todas las disciplinas de las ciencias sociales, la medición es igual de problemática.
- Consideremos el objeto de estudio en la mayoría de investigaciones de economía: el dólar, el euro o el peso.
- Si el concepto de interés es el “producto económico”, resulta bastante sencillo obtener una observación empírica que sea consistente → el PIB.
- Luego de tener clara esta medida, los economistas pueden ir al siguiente (y más interesante) paso en el proceso científico: preguntarse qué causa mayor o menor crecimiento en el producto.

- No todo concepto en economía es medido con tal facilidad.
- ¿Por qué algunos son pobres mientras que otros no?
- Aunque todos sabemos que la pobreza existe, medirla es complicado.
- Los economistas rara vez (pero ocasionalmente) tienen obstáculos con la medición. El extremo opuesto del espectro sería la psicología.

- El comportamiento humano, el proceso cognitivo y las emociones son conceptos extremadamente complicados de medir.
- La depresión y la ansiedad existen pero ¿cómo medirlas?
- Es relevante tener una medición precisa de ellas para poder evaluar si la terapia clínica o los antidepresivos son efectivos.
- La sociología y la ciencia política se localizan entre los extremos que son la economía y la psicología.
- Los siguientes conceptos críticos son ejemplos que tienen complicaciones a la hora de ser cuantificados:
 - 1 Democracia.
 - 2 Capital Social.
 - 3 Prejuicio.

Claridad Conceptual

- El primer paso para medir cualquier fenómeno es tener un sentido claro del concepto de interés.
- En muchos casos, es un trabajo difícil y revelador.
- Aún en ejemplos que parecen sencillos, el tener claro lo que queremos decir con los conceptos es más complejo de lo que parece.
- La *mejor* medida de un concepto depende de cuáles son nuestros objetivos teóricos.

Confiabilidad

- Una definición operacional es confiable si es repetible y consistente.
- Si al aplicar la misma regla de medida a una observación produce los mismos resultados, la medida es consistente.
- Confiabilidad vs. Variabilidad.
- Es sumamente importante al momento de codificar eventos o textos para una análisis cuantitativo.

Sesgo de Medida y Confiabilidad

- Una preocupación que surge con cualquier técnica de medición es el sesgo de medida.
- El sesgo de medida es el sobre-reporte o sub-reporte sistemático de los valores de una variable.
- Imaginemos que estamos frente a la elección de dos definiciones operacionales de una misma variable.
- La definición operacional A es confiable y sesgada, y la definición B es insesgada y no confiable. ¿Cuál sería la definición a usar?

- Aunque el sesgo de medida es un problema para alguien que quiere saber los “verdaderos” valores de una variable, no es un gran inconveniente a la hora de evaluar teorías.
- Al indagar teorías, estamos buscando patrones generales entre dos variables.
- ¿Con valores más altos de X , tendemos a ver valores mas altos de Y ? O valores más bajos?
- Si la medida de X está sesgada hacia arriba, el mismo patrón será visible en la asociación con Y .
- Si X no es confiable, la relación entre Y y X no será clara.

Validez


- La validez es la característica más importante de una medida.
- Esta se refiere a cuan adecuadamente una medida refleja el concepto en consideración.
- No existe una fórmula simple que la validez de una medida en una escala de 0 a 100.
- En cambio, nos basamos en la superposición de varios criterios:
 - ① Validez Aparente.
 - ② Validez Predictiva.
 - ③ Validez de Constructo.
 - ④ Validez de Contenido.

Confiabilidad vs. Validez

- Claramente, queremos medidas que sean tanto confiables como válidas.
- Sin embargo, existe una disyuntiva entre ambas características.
- Entre mayor variación le permitamos tener a un concepto, es más probable crear desacuerdos en relación a si este aplica a una situación particular.
- En cierta medida, este dilema explica la persistencia de dos enfoques diferente de investigación social: el cualitativo y el cuantitativo.
- En la generalización más simple, los métodos cualitativos tienden a ser más validos y los cuantitativos más confiables.

Controversia: Midiendo la Democracia

- Se podría estar tentado a evaluar la democracia como una condición absoluta, pero podríamos obviar el hecho de que hay países más democráticos que otros.
- Por lo tanto, se podría pensar la democracia como un *continuum*².
- Definir un continuo de valores entre democracia y totalitarismo no es sencillo.
- ¿A qué nos referimos con democracia? ¿Cuáles son los elementos básicos que conforman un gobierno más o menos democrático?

²Elkins, Zachary. 2000. "Gradations of Democracy? Empirical Tests of Alternative Conceptualizations." *American Journal of Political Science* 

Polity IV

- La medida más conocido -no necesariamente más aceptada- es la del proyecto Polity IV.
- Miden la democracia anualmente mediante valores que oscilan de -10 (fuertemente autocrático) a 10 (fuertemente democrático) para cada país de 1800 a 2012.
- Esta definición operacional tiene cuatro componentes:
 - 1 Regulación en la elección del ejecutivo.
 - 2 Competencia en la elección del ejecutivo.
 - 3 Que tan abierta es la elección del ejecutivo.
 - 4 Restricciones al ejecutivo.

- Para cada una de esas dimensiones, expertos valoran cada país en una escala.
- Por ejemplo, para el primer criterio utilizan los siguiente valores:
 - +3 = competición regular entre grupos reconocidos
 - +2 = competencia transicional.
 - +1 = patrones restringidos de competencia o de facciones.
 - 0 = no hay competencia.

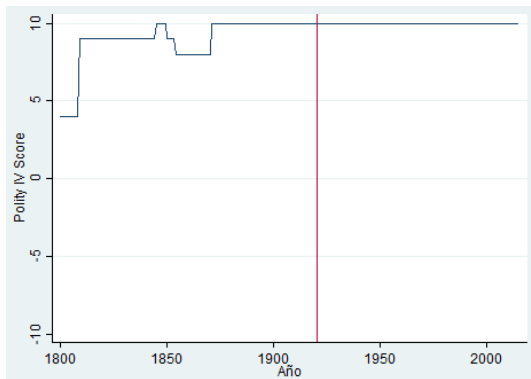
Polity IV: Colombia

Figura 1



Polity IV: Estados Unidos

Figura 2

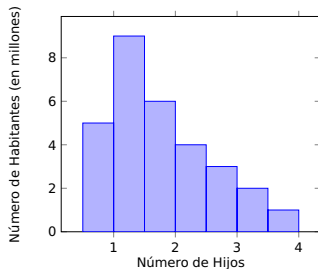


Distribución de Frecuencia

- Retomemos el concepto de distribución de frecuencia de la clase pasada...

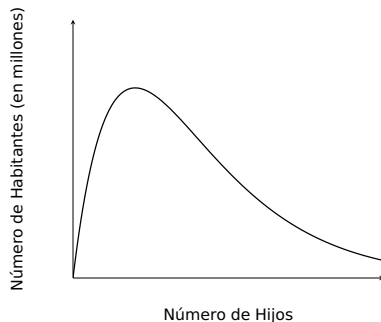
Distribución de Frecuencia

- Retomemos el concepto de distribución de frecuencia de la clase pasada...
- El histograma es una herramienta útil para presentar la distribución de una variable de razón.
- Por ejemplo: la distribución del número de hijos de un país.



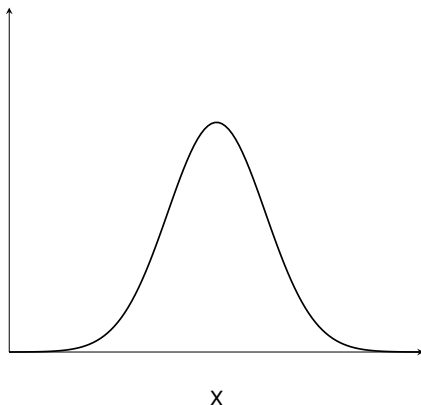
Curva de Distribución de Frecuencia

- Cuando una población consiste en valores numéricos de una escala de razón, se acostumbra a dibujar la distribución con una curva suavizada.



- Muchos textos definen la curva de distribución de frecuencias como un sustituto del histograma o del polígono.
- La sustitución es apropiada porque la curva suavizada se presenta más como una estimación de la distribución de los valores en la población.
- La distribución poblacional más común es *la curva normal*.
- La distribución normal es *simétrica*, es decir, equilibrada a ambos lados.

La distribución normal



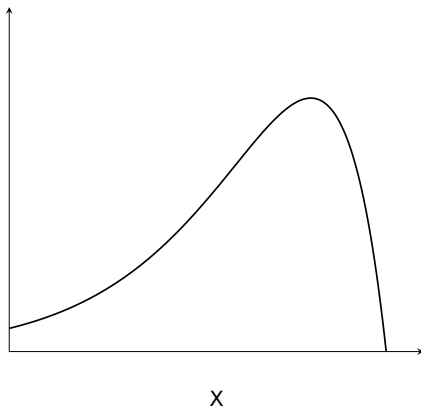
- A partir de acá, siempre que aparezca el término *distribución*, imaginen una gráfica de distribución de frecuencia.

La Forma de una Distribución de Frecuencia

- En vez de presentar la figura de la distribución de frecuencia, se puede describir al enumerar sus características.
- Existen tres características que describen completamente una distribución:
 - 1 Forma.
 - 2 Tendencia Central.
 - 3 Variabilidad.
- La forma de una distribución puede ser clasificada en *simétrica* o *sesgada*.

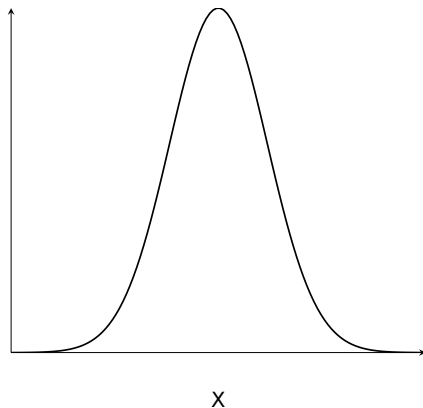
La Forma de una Distribución de Frecuencia

Distribución Sesgada Negativamente



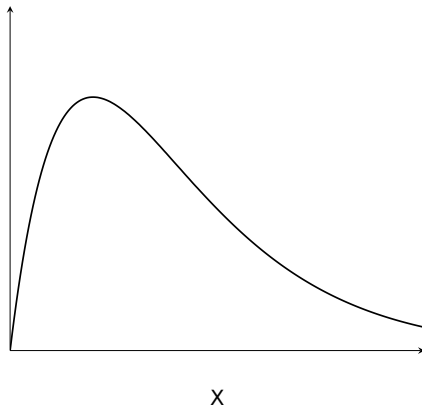
La Forma de una Distribución de Frecuencia

Distribución Simétrica



La Forma de una Distribución de Frecuencia

Distribución Sesgada Positivamente



Tendencia Central

- El propósito principal de la estadística descriptiva es organizar y resumir un conjunto de valores.
- El método más común de hacer esto es encontrar un valor *puntual* que defina el valor promedio y sea representativo a toda la distribución.
- Usualmente, este valor identifica el centro de la distribución.

Definición

Un estadístico de tendencia central proporciona una estimación de la puntuación más representativa para un grupo de valores.

La Media

La Media

Es la suma de todos los valores dividida entre el número de valores observados.

- También es conocido como el promedio aritmético.
- La media de una población se identifica por la letra griega μ . Por otro lado, \bar{X} es la notación acostumbrada para la media muestral.

$$\bar{X} = \frac{\sum X}{n}$$

Definiciones Alternativas para la Media

- La primera alternativa es pensar la media como una medición de “partes iguales”. Es decir, la cantidad que cada *individuo* recibe cuando el total ($\sum X$) es dividido entre todos los individuos (n).
- 6 estudiantes se encuentran 180 mil pesos en la calle. Si quisieran dividir el total de la plata equitativamente, ¿cuánto recibiría cada estudiante?
- Ahora supongamos que los 6 estudiantes invirtieron su dinero obteniendo una ganancia de $\bar{X} = \$500$ por estudiante. ¿Cuánto fue el total de ganancia para todo el grupo?

Definiciones Alternativas para la Media

- La segunda alternativa es considerar a la media como un punto de equilibrio.
- Imaginemos una muestra constituida por $n=5$ valores (1,2,6,6,10). Además, para esta muestra $\bar{X} = 5$

Valor	Distancia de la Media
1	4 abajo de la media
2	3 abajo de la media
6	1 arriba de la media
6	1 arriba de la media
10	5 arriba de la media

- Abajo de la media: $4+3=7$
- Arriba de la media: $1+1+5=7$
- La media equilibra las distancias de la distribución.

La Media de Muestras Combinadas

- En ocasiones es necesario combinar dos conjuntos o más de valores para encontrar la media total.
- Supongamos que tenemos dos muestras separadas. La primera tiene $n_1 = 12$ y $\bar{X}_1 = 6$. La segunda muestra tiene $n_2 = 8$ y $\bar{X}_2 = 7$.
- ¿Cuál es la media para el total del grupo?

$$\bar{X}_{total} = \frac{\sum X_1 + \sum X_2}{n_1 + n_2} = \frac{72 + 56}{12 + 8} = \frac{128}{20} = 6,4$$

- Sería un error calcular la media de las medias!

$$\frac{\bar{X}_1 + \bar{X}_2}{2} = \frac{6 + 7}{2} = \frac{13}{2} = 6,5$$

Características de la Media

- Cambiar cualquiera de los valores cambia la media.
- Adicionar un nuevo valor o sustraer un valor existente, cambia la media salvo que el valor sea igual a la media.
- Si una constante se le suma (o resta) a cada valor en una distribución, la misma constante se le suma (o resta) a la media.
- Si cada valor en una distribución es multiplicado (o dividido) por una constante, la media cambia de la misma manera.

Debilidades de la Media

- Cuando la distribución tiene valores extremos, la media no es una buena medida representativa.
- Los valores extremos pueden tener una influencia grande sobre el valor de la media.

X	1
11	4
12	3
13	1
100	1

$$\bar{X} = \frac{\sum X}{n} = \frac{203}{10} = 20,3$$

La Mediana


La Mediana

Es el valor que denota el punto medio en una distribución ordenada. En otras palabras, 50 % de los valores están por debajo de este valor.


- Para calcular la mediana, lo primero es ordenar los datos de menor a mayor.
- Luego se divide entre 2 el tamaño de la muestra n y se suma 0.5 al valor obtenido. Esto nos dará la **posición** de la mediana.
- Si el n es impar, la mediana será el valor de la mitad de la distribución.
- Si el n es par, la mediana será el promedio entre los dos valores de la mitad.

¿Cómo obtener la Mediana?

- Consideremos el siguiente conjunto de $n = 5$ ingresos familiares:

Orden	1	2	3	4	5
Valores	\$3540	\$4675	\$7350	\$9860	\$19000
					
			<i>Mediana</i>		

- Ahora, si una sexta familia con ingreso de \$20000 se añade a la muestra anterior:

Orden	1	2	3	4	5	6
Valores	\$3540	\$4675	\$7350	\$9860	\$19000	\$20000
						
			<i>Mediana=\$8605</i>			

Debilidades de la Mediana

- Dos distribuciones pueden tener la misma mediana aun cuando estén compuestas de puntuaciones muy diferentes.

$$\text{Notas}_1 = 39 \quad 51 \quad 77 \quad 78 \quad 81$$

$$\text{Notas}_2 = 74 \quad 75 \quad 77 \quad 94 \quad 98$$

- A pesar de que la mediana es insensible a los valores, es muy sensible al tamaño de la muestra.
- Por ejemplo, si dos notas se añaden al conjunto 1:

$$\text{Notas}_1 = 34 \quad 36 \quad 39 \quad 51 \quad 77 \quad 78 \quad 81$$

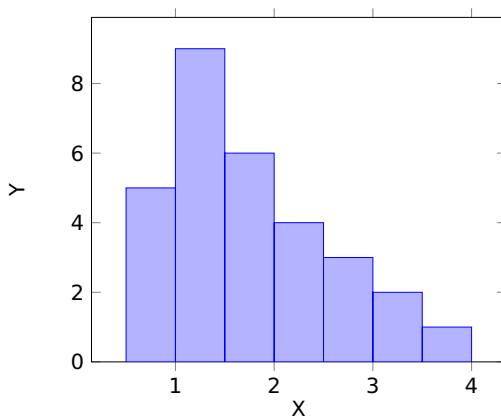
La Moda

La Moda

En una distribución de frecuencia, la moda es el valor o categoría que más se repite.

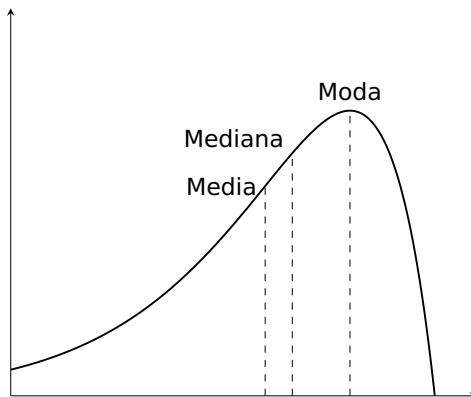
- Aunque una distribución sólo puede tener una media y una mediana, es posible que tenga más de dos modas.
- La existencia de dos modas, a menudo indica que dos grupos diferentes existen dentro de la misma población.
- La moda es representativa para variables nominales y ordinales. Con variables de razón, es útil acompañada de la media y la mediana.

La Moda en una Gráfica



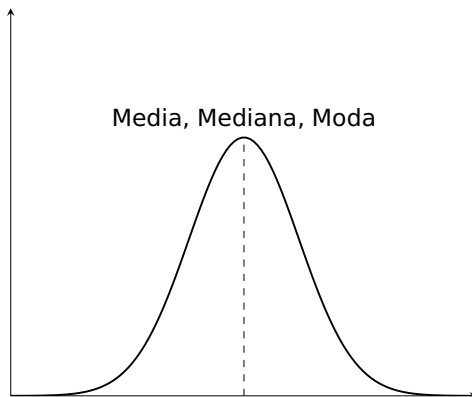
Tendencia Central y la Forma de la Distribución

Sesgada a la Izquierda



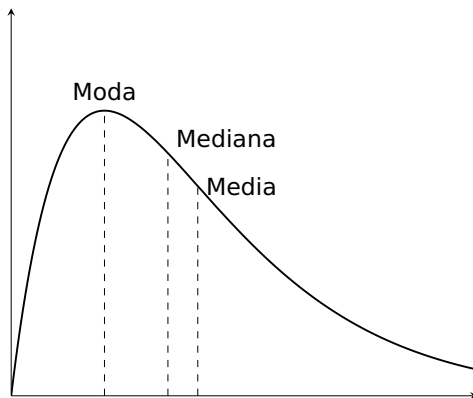
Tendencia Central y la Forma de la Distribución

Simétrica



Tendencia Central y la Forma de la Distribución

Sesgada a la Derecha



Ejercicio

X	f
4	1
3	4
2	3
1	2

- ¿Cuál es la media, la mediana y la moda?

Ejercicio

X	f
4	1
3	4
2	3
1	2

- ¿Cuál es la media, la mediana y la moda?
- $\bar{X} = \frac{\sum fX}{\sum f} = \frac{24}{10} = 2,4$

Ejercicio

X	f
4	1
3	4
2	3
1	2

- ¿Cuál es la media, la mediana y la moda?
- $\bar{X} = \frac{\sum fX}{\sum f} = \frac{24}{10} = 2,4$
- Mediana = $\frac{2 + 3}{2} = \frac{5}{2} = 2,5$

Ejercicio

X	f
4	1
3	4
2	3
1	2

- ¿Cuál es la media, la mediana y la moda?
- $\bar{X} = \frac{\sum fX}{\sum f} = \frac{24}{10} = 2,4$
- Mediana = $\frac{2 + 3}{2} = \frac{5}{2} = 2,5$
- Moda = $X = 3$