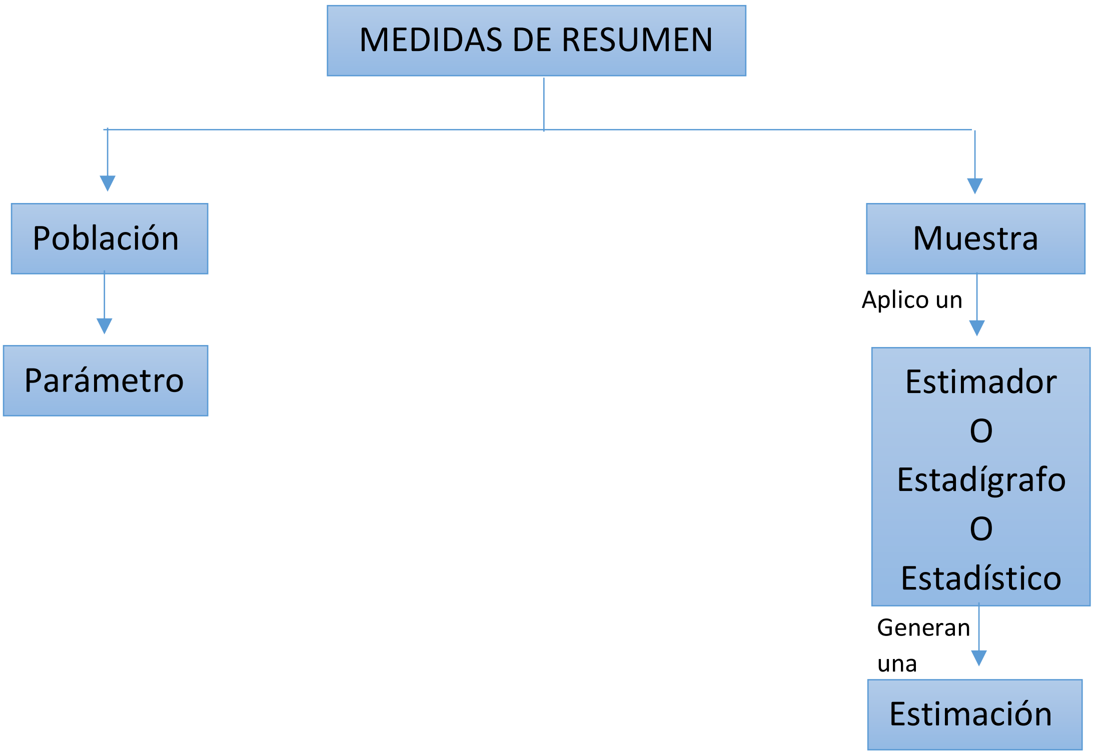
MEDIDAS DE RESUMEN

En estadística son valores numéricos que representan una característica determinada de la población (conjunto) o de la muestra (subconjunto aleatorio)
Cada unidad experimental tiene su característica, su valor, su dato; pero en el conjunto o subconjunto aleatorio comparten un valor que resume esa característica. Por ejemplo: Los alumnos de la clase tienen una edad promedio.



Parámetro: Es una medida que describe una variable que utiliza datos de una población

Estadígrafos / Estadísticos / Estimadores: Es una medida que describe una variable que utiliza datos de una muestra. Los estadígrafos, estadísticos o estimadores son el algoritmo matemático a partir del cual, al aplicarlo a una muestra, obtengo un valor que será denominado ***estimación.***

Por ejemplo: El promedio. La acción de sumar todo y dividirlo por la cantidad de elementos, cuando lo estoy realizando sobre una muestra, es el estimador; y el resultado que me devolverá ese algoritmo será la estimación. Si calculamos el promedio sobre la población, tanto el algoritmo matemático como el resultado son el parámetro. No hay diferencia para la población ya que la población es una, y cuando aplico el parámetro, ejemplo el concepto media poblacional, voy a tener un solo valor. No hay diferencia entre estimador y estimación.

Estas medidas de resumen, que explican determinada característica, van a tener distintos nombres dependiendo de lo que expliquen:

MEDIDAS DE TENDENCIA CENTRAL

Estas medidas indican como se agrupan los datos alrededor de un valor central

MEDIDAS DE VARIACION

Indican de qué manera se diseminan los datos a lo largo de la distribución

MEDIDAS DE POSICION

Estas medidas no toman en cuenta todos los valores de la variable en estudio sino que tienen en cuenta la posición de los mismos

MEDIDAS DE FORMA

Indican que tipo de concentración (sesgada o no) sigue los datos. Es el patrón de la distribución de valores, desde el más bajo hasta el más alto.

Todas esas medidas mencionadas, cuando las aplicamos a muestras, son estimadores, que nos van a arrojar un valor, ese valor va a ser la estimación.

Medidas de Tendencia Central

1. **Modo o Moda (Mo):** Indica el valor al que le corresponde la máxima frecuencia absoluta simple, es decir, el valor que más se repite.

Ejemplo: Realizo una encuesta: ¿Te gusto la clase? SI – NO – NO SE. Lo único que puedo hacer con ese tipo de encuesta es contar cual dato (si – no – no se) se repitió más. Es decir, lo único que puedo hacer es calcular la moda. Si la respuesta "si" fue la que más se repitió, entonces la moda va a ser "si".

Algunas consideraciones importantes:

- A la moda no le interesa cuantas veces se repite el valor que más se repite. No te da esa información.
- Las variables cualitativas nominales solo aceptan la moda como medida de tendencia central; pero la moda también se puede aplicar a variables cuantitativas.
- En ocasiones no existe una moda en un conjunto de datos, o bien tienen más de una moda (bimodal, multimodal). Ejemplo: considere los siguientes datos sobre el tiempo, en minutos, que le toma a una persona prepararse para salir: 29 31 35 39 39 40 43 44 44 52. Se observan dos modas: 39 minutos y 44 minutos, ya que cada uno de esos valores se presenta dos veces.

  Un conjunto de datos no tiene moda si ninguno de los valores es "el más común", es decir, si todos los valores se repiten la misma cantidad de veces.
- Los valores extremos no afectan a la moda.

2.1. **La Mediana (Me):** Es el valor que separa el 50% inferior de los datos del 50% superior de los datos.

Datos 50% ← Me → 50% datos

Ejemplo: Se realiza una encuesta en la que se debe seleccionar alguna de las siguientes opciones: Malo – Regular – Bueno

X_i	f_i
Malo	20
Regular	10
Bueno	20

Regular seria la mediana (vale lo mismo lo que esta abajo y lo que está arriba)

Algunas consideraciones importantes:

- Te da el valor de variable que acumula la misma cantidad por debajo y por encima, no te informa cuanto acumula el valor.
- Se puede utilizar a partir de escala ordinal.

2.2. La Mediana con Variables Numéricas: Es aquella posición, aquel valor, que acumula la misma cantidad de datos por un lado y la misma cantidad de datos por el otro, sin importar cuánto valen los datos.

Ejemplo: 1 1 1 2 3 4 5 6 7 n impar

En la posición 5 tengo 4 datos para un lado y 4 datos para el otro.

Mediana: 3 Posición Mediana: 5

1 1 1 1 2 3 4 5 6 7 n par

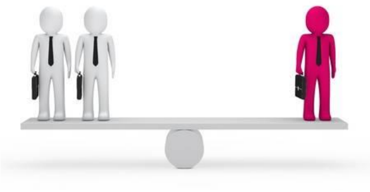
Cuando el conjunto o subconjunto es par, tomo esos dos valores, los sumo y los divido por dos para obtener la mediana.
Para obtener la posición mediana: $\frac{n+1}{2}$ Mediana: 2,5 Posición mediana: 5,5

Algunas consideraciones importantes:

- A la mediana, para variables cuantitativas, no le importa cuánto valen los valores que están por debajo o por encima de ella. Solo le importa cuántos son.
- Si el conjunto de datos contiene un número impar de valores, la mediana es el valor asociado con el dato ubicado a la mitad.
- Si el conjunto de datos contiene un número par de valores, la mediana es el valor asociado con el promedio de los datos ubicados a la mitad.

3.1. La Media Aritmética: La media aritmética (generalmente denominada media) sirve como un "punto de equilibrio" en un conjunto de datos (como el punto de apoyo en un sube y baja)

La media es el centro de gravedad
de la distribución



__Algunas consideraciones importantes:__

- La media se calcula sumando todos los valores en un conjunto de datos y luego dividiendo el resultado de esa suma por el número de valores en dicho conjunto.
- Solamente se puede utilizar para variables Numéricas.
- Debemos diferenciar si es un parámetro o un estadígrafo.

Media aritmética poblacional: μ_x

Media aritmética muestral: $\overline{X}$

Media Aritmética poblacional para datos desagrupados:

$$\mu_x = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N}$$

Referencias:

(Letra minúscula)

$\sum \longrightarrow$ Suma $i = 1 \longrightarrow$ desde la primer observación $X_i \longrightarrow$ observación i-esima

Lo resaltado con rosa me indica que estoy calculando la media de una población.

N $\longrightarrow$ Tamaño poblacional.

*En la parte superior del símbolo de suma va la cantidad de sumandos que tengo.

Media Aritmética Muestral para datos desagrupados:

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

Referencias:

$\sum$ ⟶ Suma $i = 1$ ⟶ desde la primer observación X_i ⟶ observación i-esima

Lo resaltado con amarillo me indica que estoy calculando la media de una muestra.

n ⟶ Tamaño muestral.

Media Aritmética poblacional para datos agrupados:

$$\mu = \frac{\sum_{i=1}^{k} x_i \times fi}{N} = \sum_{i=1}^{k} x_i \times fri$$

Media Aritmética muestral para datos agrupados:

$$\bar{X} = \frac{\sum_{i=1}^{k} x_i \times fi}{n} = \sum_{i=1}^{k} x_i \times fri$$

Referencias:

K ⟶ Clases

fi ⟶ Frec. Abs. Simple

fri ⟶ frec. Relativa simple

Me indica que los datos están agrupados.

Me indica que estoy trabajando con una población.

Me indica que estoy trabajando con una muestra.

3.2. Media Resumida:

Se utiliza con datos continuos. Se calcula como la sumatoria de las _Marcas de Clase_, donde M representa el intervalo, por (×) la frecuencia relativa simple.

$$Media\ Resumida = \sum_{i=1}^{M} \left(\frac{Miks + Miki}{2}\right) * fri$$

Recordatorio: No existe la exactitud en los datos continuos, entonces, necesariamente cuando tengo que agrupar los datos continuos debo hacerlo en distribuciones de frecuencia por intervalos de clase.

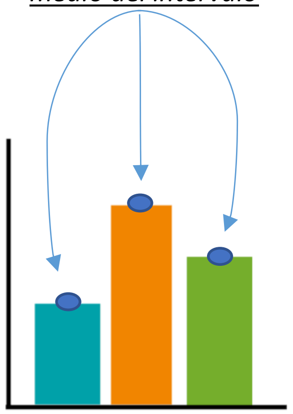
Miks → *M i-esima k superior* → *Es la clase superior del intervalo.*

Miki → *M i-esima k inferior* → *Es la clase inferior del intervalo.*

$$\frac{Miks+Miki}{2}$$

Esa suma está siendo dividida por 2 porque supone que todos los datos se concentran el valor medio del intervalo. Esa suposición es justamente el problema de la media resumida

Así calculo las _Marcas de Clase_ o _Punto medio del intervalo_



Histograma

Ejemplo: Tengo el intervalo [0;10) con una frecuencia relativa de 0,2. Esto solo me da información que entre 0 y 10 hay un 20% de las observaciones. No me dice dónde están dentro de ese intervalo. Lo que voy a suponer (ya que no tengo esa información) es que ese 0.2 sucede en la mitad del intervalo (en este caso ocurriría en 5).

A partir de esa media resumida generalizamos a una cantidad infinita de intervalos y suponemos que cada intervalo mide un infinitesimal. Al suponer eso suponemos que ese intervalo es el valor puntual de la variable continua.

$$\sum_{i=1}^{M}\left(\frac{Miks + Miki}{2}\right) \times fri =$$

$$= \int_{Kinf}^{Ksup} x_i \times fri \; dx = \underbrace{\sum_{i=1}^{M\to\infty}\left(\frac{Miks + Miki}{2}\right)}_{\text{La integral nos indica eso. La diferencia entre Miks y Miki tiende a cero.}} \times fri = \int_{0}^{100} x_i \times fri \; dx$$

__Referencias:__

Ksup ⟶ *El máximo valor de la distribución: El mayor valor del mayor intervalo.*

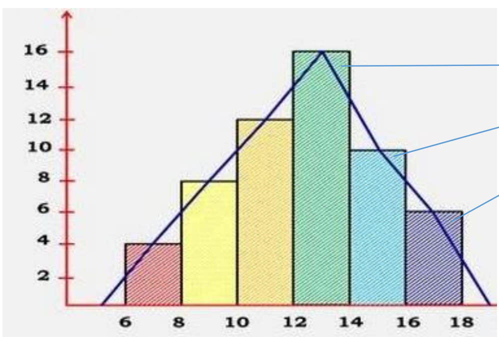
Kinf ⟶ *El minimo valor de la distribución: El menor valor del menor intervalo.*

Xi ⟶ *Valor de la clase: Son las marcas de clase de cada intervalo.*

fri ⟶ *Frecuencia Relativa Simple.*

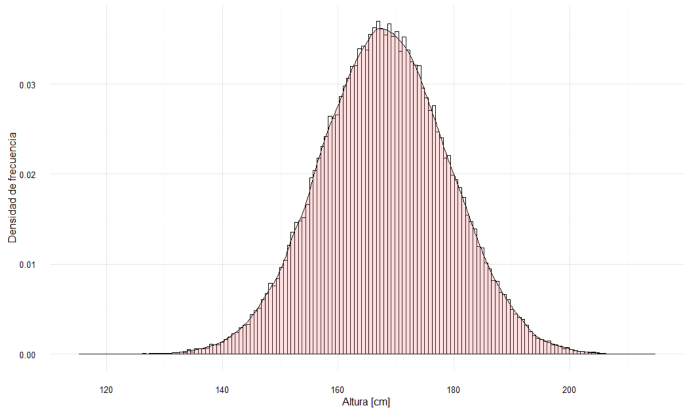
Esa integral definida la realizo con el objetivo de perder cada vez menos información.

Para perder cada vez menos información o, dicho de otra manera, tener cada vez más información de cada intercalo, tengo que lograr que estos sean cada vez más chicos. Podemos tomar infinitos intervalos para eso. Si los intervalos son cada vez más chicos también lo va a ser la diferencia entre la clase superior y la clase inferior: así nos vamos acercando a una medida más exacta de resumen.



Si trazamos una línea que pase por las marcas de clase podemos observar unos "triangulitos" que quedan por fuera de la línea, y que representan la información que perdemos.

Para solucionar ese problema, tendemos a infinito la cantidad de intervalos (haciéndolos muy muy chicos). De esa forma podemos observar que perdemos mucha menos información, prácticamente nula, y así minimizo la posibilidad de cometer errores al suponer que el valor se encuentra en la mitad del intervalo.



Haciendo los intervalos tan chicos como yo quiera puedo llegar al valor de la variable. Si hago tender a cero la diferencia que hay entre el límite inferior y el límite superior del intervalo, el resultado va a ser muy parecido a esos límites.

Por ejemplo:

$$\frac{10{,}001+10{,}002}{2}=10{,}0015$$

Podemos observar que el resultado es muy parecido al límite inferior y superior.

Entonces es por eso que en el límite en donde yo tengo infinitos intervalos, en lugar de poner las marcas de clase ((Miks + Miki)/2) pongo directamente el valor de la variable; es decir; puedo asociar a cada marca de clase con el punto en sí.

De esta forma si queremos calcular la media resumida de f(x) (que representa los datos de nuestra variable en observación) lo podríamos hacer a través del cálculo de áreas con la siguiente integral:

$$\int_{Kinf}^{Ksup} \underbrace{x_i \times fri}_{\text{Es f(x)}}\, dx$$

ACLARACIONES:

-Podemos tomar infinitos intervalos porque la variable es continua, y eso significa que entre un valor y el otro existen infinitos valores.

-Los limites van a depender del fenómeno que se esté estudiando. Pueden tomar cualquier valor desde -∞ a +∞

Medidas de Posición

1. **Los Cuartiles (Q):** Representan valores de la variable que dividen a la distribución en cuartos.

   Es una medida de posición ya que no toma en cuenta los valores de la variable en estudio sino que tienen en cuenta la posición de los mismos.

   -Cuartil 1 (Q1) Es el valor de variable hasta el cual se acumula la primer cuarta parte de las observaciones (25%)

   -Cuartil 2 (Q2) Es el valor de variable hasta el cual se acumula las dos cuartas partes de las observaciones (50%) (Coincide con la mediana).

   -Cuartil 3 (Q3) Es el valor de variable hasta el cual se acumula las tres cuartas partes de las observaciones (75%).

   -Cuartil 4 (Q4) Es el valor de variable hasta el cual se acumula el total de las observaciones (100%). Coincide con el último valor de variable.

   También hay **otras medidas de posición** como los **deciles (que dividen los datos en 10 partes iguales)** y los **percentiles** también conocidos como **centiles (que dividen los datos en 100 partes idénticas)**

Medidas de Dispersión o Variación

Las medidas de variación las utilizo para saber qué pasa con la distancia entre el menor valor y el mayor valor, con datos desagrupados. Es decir, la variación mide la dispersión de los valores en un conjunto de datos.

1. **Rango (R_x)** Es una medida de dispersión total (es decir, toma el total de la distribución y se fija cuanto varia punta a punta), medida en términos absolutos. El rango es igual al valor más grande menos el valor más pequeño.

$$Rango = x_{mas\ grande} - x_{mas\ pequeño}$$

$$Rango = Q_4 - Q_0$$

Problemas del Rango:

1) Ignora lo que pasa dentro de la distribución (porque solo toma en cuenta las puntas)

Ejemplo: 1 – 2 – 3 – 4 – 5 – 6 → Rango = 5
1 – 1 – 1 – 1 – 1 – 6 → Rango = 5

El rango es igual siendo distintas las distribuciones

2) Es sensible a grandes dispersiones en sus límites (es decir a grandes cambios en los valores de los limites).

Ejemplo: 1 – 1 – 1 – 1 – 1 – 106 → Rango = 105
1 – 1 – 1 – 1 – 1 – 6 → Rango = 5

El rango es muy distinto y las distribuciones son casi iguales. Solo cambia un valor

Estos problemas podemos solucionarlos construyendo medidas de variabilidad promedio, que evalúen el comportamiento respecto a la dispersión de todos los valores de la distribución. El problema va a ser que, como podremos observar, dichas medidas traerán aparejadas sus propios problemas.

Ejemplo disparador:

Se tienen dos distribuciones de datos, correspondientes a la variable “X” y a la variable “Y” respectivamente. Dichos valores se exponen a continuación: (Posterior a los cuadros se encuentran sus desarrollos paso a paso)

x_i	$(x_i - \mu_X)$	$\lvert x_i - \mu_X \rvert$	$(x_i - \mu_X)^2$
15	**-5**	**5**	**25**
18	**-2**	**2**	**4**
20	**0**	**0**	**0**
20	**0**	**0**	**0**
20	**0**	**0**	**0**
22	**2**	**2**	**4**
25	**5**	**5**	**25**
E(X) = 20	$E(x_i - \mu_X) = 0$	$E\lvert x_i - \mu_X \rvert = 2$	$E[(x_i - \mu_X)^2] = \frac{58}{7}$

$N_X = 7$
$R_X = 10$

$N_Y = 7$
$R_Y = 10$

y_i	$(y_i - \mu_Y)$	$\lvert y_i - \mu_Y \rvert$	$(y_i - \mu_Y)^2$
15	-5	5	25
19	-1	1	1
19	-1	1	1
20	0	0	0
21	1	1	1
21	1	1	1
25	5	5	25
E(Y) = 20	$E(y_i - \mu_Y) = 0$	$E\lvert y_i - \mu_Y \rvert = 2$	$E[(y_i - \mu_Y)^2] = \frac{54}{7}$

Objetivo: Quiero encontrar una medida de variabilidad que siempre, pase lo que pase, me muestre que esas dos distribuciones son diferentes (si es que lo son).

Procedimiento paso a paso del cuadro:

1. El Rango

Rango de X → 25 – 15 = 10
Rango de Y → 25 – 15 = 10

Mi medida de variabilidad en términos absolutos es 10.

Ahora voy a calcular distintos tipos de promedio para mensurar la variabilidad dentro de la distribución y no solo ver qué pasa con las puntas.

2. La media o Esperanza Matemática

E(X) = (15 + 18 + 20 + 20 + 20 + 22 + 25) / 7 = 20
E(Y) = 15 × 1/7 + 19 × 2/7 + 20 × 1/7 + 21 × 2/7 + 25 × 1/7 = 20

3. La Esperanza de las dispersiones respecto del valor medio:

Calculo el promedio de las comparaciones de cada observación con respecto a la media.

$(x_i - \mu_X)$ →

15 – 20 = -5
18 – 20 = -2
20 – 20 = 0
20 – 20 = 0
20 – 20 = 0
22 – 20 = 2
25 – 20 = 5

$(y_i - \mu_Y)$ →

15 – 20 = -5
19 – 20 = -1
19 – 20 = -1
20 – 20 = 0
21 – 20 = 1
21 – 20 = 1
25 – 20 = 5

$$E(x_i - \mu_X) = \frac{(-5) + (-2) + 0 + 0 + 0 + 2 + 5}{7} = 0$$

$$E(y_i - \mu_Y) = \frac{(-5) + (-1) + (-1) + 0 + 1 + 1 + 5}{7} = 0$$

La media es el valor que compensa las dispersiones menores a ese valor con las dispersiones mayores a ese valor. Por eso siempre va a dar cero, razón por la cual esta medida no me sirve como medida promedio de variabilidad.

Como el problema es que tengo valores negativos y positivos que se compensan y por eso me da cero, para solucionar ese problema voy a aplicar el modulo.

4. **La esperanza del módulo de las comparaciones:** El promedio de los valores absolutos de las dispersiones respecto del valor medio.

$|x_i - \mu_X|$ →

I 15 – 20 I = 5
I 18 – 20 I = 2
I 20 – 20 I = 0
I 20 – 20 I = 0
I 20 – 20 I = 0
I 22 – 20 I = 2
I 25 – 20 I = 5

$|y_i - \mu_Y|$ →

I 15 – 20 I = 5
I 19 – 20 I = 1
I 19 – 20 I = 1
I 20 – 20 I = 0
I 21 – 20 I = 1
I 21 – 20 I = 1
I 25 – 20 I = 5

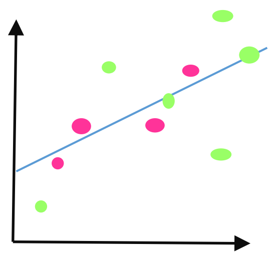
$$E|x_i - \mu_X| = \frac{5+2+0+0+0+2+5}{7} = 2 \qquad |E(y_i - \mu_Y)| = \frac{5+1+1+0+1+1+5}{7} = 2$$

Este problema surge <u>a veces:</u> cuando calculo la esperanza de los valores absolutos de las diferencias de las observaciones con respecto a la media <u>puede pasar</u> que siendo distintas distribuciones me den iguales. Entonces, esta medida tampoco me sirve, porque no me garantiza que siempre que sean diferentes las distribuciones esta medida me dé diferente también.

Esto ocurre porque no castiga a las variabilidades más grandes: para que no valga lo mismo una variabilidad de dos que dos variabilidades de uno. Por eso, para solucionar este problema, <u>voy a elevar al cuadrado las diferencias entre las observaciones y el valor medio.</u>

<u>NOTA:</u>

En estadística se prefiere tener más cantidad de variabilidades pero de una menor magnitud que menos cantidad de variabilidades de mayor magnitud.



En este caso, voy a preferir los puntos rosa.

2

5. La Varianza: El promedio de los valores cuadráticos de las dispersiones respecto al valor medio

$(x_i - \mu_X)^2$ →

15 – 20 = $(-5)^2$ = 25
18 – 20 = $(-2)^2$ = 4
20 – 20 = 0^2 = 0
20 – 20 = 0^2 = 0
20 – 20 = 0^2 = 0
22 – 20 = 2^2 = 4
25 – 20 = 5^2 = 25

$(y_i - \mu_Y)^2$ →

15 – 20 = $(-5)^2$ = 25
19 – 20 = $(-1)^2$ = 1
19 – 20 = $(-1)^2$ = 1
20 – 20 = 0^2 = 0
21 – 20 = 1^2 = 1
21 – 20 = 1^2 = 1
25 – 20 = 5^2 = 25

$$E(x_i - \mu_X)^2 = \frac{25+4+0+0+0+4+25}{7} = \frac{58}{7}$$

$$E(y_i - \mu_Y)^2 = \frac{25+1+1+0+1+1+25}{7} = \frac{54}{7}$$

Esta medida de variabilidad castiga a los valores que valen más con respecto a los que valen menos y me informa así que pasa con las variabilidades dentro de la distribución. Además, al elevar las comparaciones al cuadrado también impide que se compensen las variabilidades menores con las variabilidades mayores a la media.

Varianza poblacional $\rightarrow$ VAR (X) $\rightarrow$ ${\sigma^2}_X$

Varianza muestral $\rightarrow$ VAR (X) $\rightarrow$ ${S^2}_X$

Varianza poblacional para datos desagrupados:

$$\mathrm{VAR(X)} = {\sigma^2}_X = E(x_i - \mu_X)^2 = \frac{\sum_{i=1}^{N}(x_i - \mu_X)^2}{\mathrm{N}}$$

Varianza muestral para datos desagrupados:

$$VAR(X) = {S^2}_X = \frac{N}{n-1}E(x_i - \bar{X})^2 = \sum_{i=1}^{n}\frac{(x_i - \bar{X})^2}{N} \times \frac{N}{n-1} = \sum_{i=1}^{n}\frac{(x_i - \bar{X})^2}{n-1}$$

En lugar de comparar las observaciones con respecto a la media poblacional se compara las observaciones de la muestra con respecto a la media de esa muestra

Factor de corrección por finitud

Siempre que utilizo un estimador para reemplazar a un parámetro desconocido pierdo un grado de libertad. Como yo no conozco el valor de la media poblacional y la reemplazo por el valor de esa media muestral estoy utilizando en la construcción de la varianza un estimador reemplazando a un parámetro. En ese caso en el denominador le tengo que restar 1.

Hay medidas en las que utilizo más de un estimador para reemplazar a más de un parámetro. En ese caso por cada estimador que uso pierdo un grado de libertad.

Si el tamaño muestral es grande entonces el grado de libertad que se pierde se vuelve “despreciable” (que reste o no ese 1 no genera cambios grandes en el resultado). Entonces si n tiende a un número grande voy a calcular la varianza muestral de la siguiente manera:

$$VAR(X) = S^2{}_X = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$$

<u>Varianza poblacional para datos agrupados:</u>

$$\mathrm{VAR(X)} = \sigma^2{}_X = \sum_{i=1}^{K} (x_i - \mu_X)^2 \times fr(x_i)$$

Esto se lee: La suma de las diferencias entre las observaciones con respecto al valor medio elevados al cuadrado, multiplicado por la frecuencia relativa simple.

5.1. Varianza para variables continuas:

Voy a dejar de comparar clases para pasar a comparar <u>marcas de clase</u> (el supuesto de que todos los datos del intervalo se concentran en la mitad del intervalo)

$$VAR(X) = \sum_{i=1}^{K} (x_i - \mu_X)^2 \times fr(x_i) \rightarrow \int_{-\infty}^{+\infty} (x_i - \mu_X)^2 \times fr(x_i)\, dx$$

Media resumida que calcule utilizando la integral

La marca de clase

$$\frac{Miks + Miki}{2}$$

Va a ser el valor de la variable continua cuando tengo infinitos intervalos.

Problema de la varianza:

Su gran problema es que **esta expresada en medidas cuadráticas,** por lo tanto, no puedo compararla con otras unidades de medida como la media, ya que se encuentran en distintas dimensiones. Ejemplo: No puedo comparar un metro cuadrático con un metro lineal. De la misma forma la varianza esta en unidades cuadráticas (pertenecientes al espacio bidimensional) mientras que la media está en unidades lineales (pertenecientes al espacio unidimensional).

Para solucionar este problema utilizo el…

6. **Desvío estándar** Solucionamos el problema poniéndole la raíz cuadrada a la varianza.

$$\sigma_x = \sqrt{\sum_{i=1}^{K} (x_i - \mu_X)^2 \times fr(x_i)}$$

7. **Coeficiente de variabilidad** Lo que hace es comparar el desvío con respecto al valor medio. Es una medida de variabilidad relativa con respecto al valor medio y es adimensional.

$$\boldsymbol{poblacional \rightarrow CV = \frac{\sigma_x}{\mu_x}}$$

$$\boldsymbol{muestral \rightarrow CV = \frac{S_x}{\bar{x}}}$$

Ejemplo: Tengo una distribución que tiene un desvío igual a 3 años y una media igual a 1 año. Por otro lado tengo otra distribución que tiene un desvío igual a 12 meses y una media igual a 6 meses.

$$\sigma_x = 3\ años \ \wedge \ \mu_x = 1\ año \ \rightarrow CV = 3$$

$$\sigma_y = 12\ meses \ \wedge \ \mu_y = 6\ meses \ \rightarrow CV = 2$$

Esos valores no usan una unidad de medida de dimensión, entonces los puedo comparar: 3 tiene mayor variabilidad relativa con respecto a la media.

Momentos Estadísticos

Pueden utilizarse tanto para datos discretos como para datos continuos.

Momentos Estadísticos → **Absolutos**

→ **Centrales o Centrados**

Para variables discretas de orden K

- Un momento para variables discretas es un promedio de cosas.
- Los momentos tienen orden 1, orden 2, orden 3, orden k. Son discretos, es decir, no existe un orden ½.
- No todos los momentos son asociables a medidas de estadística.

Momentos Absolutos: El momento de orden k absoluto es la suma de los valores de variable elevado a la k sobre N:

$$\sum_{i=1}^{N} \frac{{X_i}^k}{N} \rightarrow \sum_{i=1}^{N} \frac{X_i}{N}$$

Momento absoluto de orden 1 es la media (poblacional en este caso)

Momentos Centrales:

$$\sum_{i=1}^{N} \frac{(x_i - \mu_x)^K}{N} \rightarrow \sum_{i=1}^{N} \frac{(x_i - \mu_x)^2}{N}$$

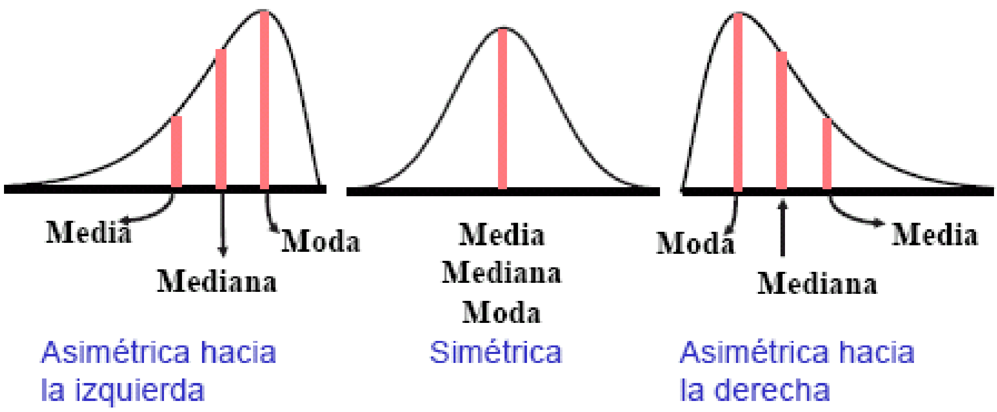
Momento central de orden 2 es la varianza.

$$\sum_{i=1}^{N} \frac{(x_i - \mu_x)^1}{N} = 0$$

SIEMPRE

Medidas de Forma

1. Simetría: Coeficiente de asimetría de Fisher.



En una distribución simétrica los valores que están por debajo de la media se distribuyen exactamente de la misma forma que los valores que están por arriba de esta. Así los valores altos y bajos se neutralizan.

En una distribución sesgada o asimétrica se produce un desequilibrio entre los valores altos y los bajos. Los valores no se distribuyen de manera simétrica alrededor de la media.

Sesgados hacia la izquierda: La mayoría de los valores se encuentran en la parte superior de la distribución. Estos valores hacen que la media se deslice hacia abajo, provocando que esta sea menor que la mediana.

Sesgados hacia la derecha: La mayoría de los valores se encuentran en la parte inferior de la distribución. Estos valores hacen que la media se deslice hacia arriba, provocando que esta sea mayor que la mediana.

$$A_F = \frac{1}{N}\sum_{i=1}^{N}\frac{(x_i-\mu_x)^3}{\sigma^3} \Longrightarrow A_F = \sum_{i=1}^{N}\frac{1}{N}\frac{(x_i-\mu_x)^3}{\sigma^3} = \sum_{i=1}^{N}\frac{(x_i-\mu_x)^3}{N}\times\frac{1}{\sigma^3} = \boxed{\frac{m_c(3)}{\sigma^3}}$$

La asimetría de Fisher es un promedio de comparaciones con respecto a la media elevado al cubo.

Momento Centrado de Orden 3 dividido el desvío al cubo

2. Coeficiente de Curtosis Compara la altura de la función con respecto a la distribución normal.

$$C = \sum_{i=1}^{N} \frac{(x_i - \mu_x)^4}{N} \times \frac{1}{\sigma^4} = \boxed{\frac{m_c(4)}{\sigma^4}}$$

Momento Centrado de Orden 4 dividido el desvío a la cuarta (curtosis poblacional)

$$C = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^4}{n} \times \frac{1}{s^4} = \boxed{\frac{m_c(4)}{S^4}}$$

Momento Centrado de Orden 4 dividido el desvío a la cuarta (curtosis muestral)