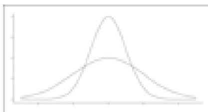
Probabilidad y Estadística 3

Actividades de Aprendizaje



Conceptos y definiciones de esta clase:

- Moda
- Mediana
- Medidas de Variabilidad o Dispersión
- Desvío medio
- Varianza
- Desvío estándar
- Coeficiente de variación
- Medidas de posición
- Cuartiles y percentiles

1. Medidas de tendencia central: moda y mediana

- Modo o Moda:

Es el valor de la variable que se repite con mayor frecuencia es decir que se repite la mayor cantidad de veces.
Si la variable es cuantitativa discreta bastará con inspeccionar la columna de las frecuencias y ubicar el valor de la variable a la cual le corresponde la mayor.

Ejemplo 9: Supongamos que la variable es la edad de 40 personas.

Edad X_i	Frecuencia
19	1
22	6
24	9
28	4
29	2
33	3
35	4
38	7
45	4
Total	40

La frecuencia es la cantidad de personas y el modo es 24 años, que se repite 9 veces.

Observación: Muchas veces se confunde el modo colocando el valor que tiene la frecuencia, recordar que el modo como cualquier otra medida ***son valores de la variable***.

Si la variable es cuantitativa continua al intervalo al cual le corresponde la mayor frecuencia se lo denomina *intervalo modal.* Dentro de este intervalo se encontrará el valor del modo y se calcula con la siguiente expresión:

$$M_0 = L_i + \frac{d_i}{d_i + d_{i+1}}.a_i$$

$M_0 = Modo$

$L_i =$ *Límite inferior del intervalo modal*

$d_i =$ *Diferencia entre la frecuencia absoluta del intervalo modal y la frecuencia absoluta del intervalo anterior* $[f_i - f_{i-1}]$

$d_{i+1} =$ *Diferencia entre la frecuencia absoluta del intervalo modal y la frecuencia absoluta del intervalo posterior* $[f_i - f_{i+1}]$

$a_i =$ *Amplitud del intervalo modal.*

Desarrollemos estos pasos en un ejemplo

Ejemplo 10: Supongamos que la variable es la altura de 30 personas. Entonces obtenemos:

Altura X_i	Frecuencia
[1.50; 1.60)	7
[1.60; 1.70)	18
[1.70; 1.80)	5
Total	30

Como estamos trabajando con intervalo modal para obtener el modo tenemos que calcular la siguiente fórmula:

$$M_0 = L_i + \frac{d_i}{d_i + d_{i+1}}.a_i$$

Donde:

$M_0 = Modo$

$L_i =$ es 1.60

$d_i =$ es 18-7=11

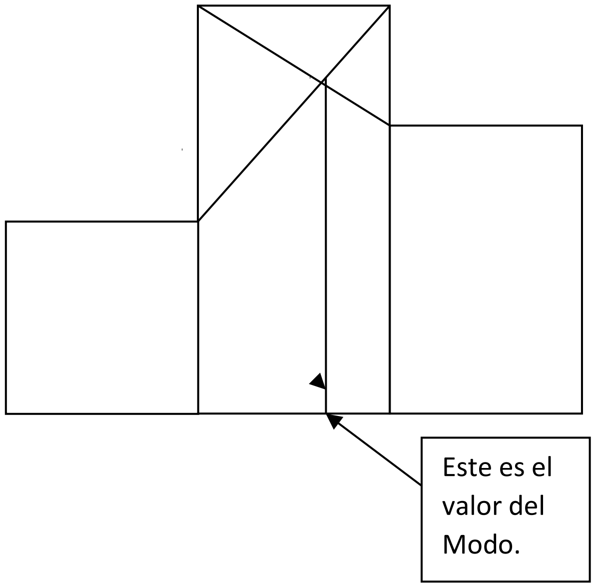
$d_{i+1} =$ *es 18-5=13*

$a_i =$ es 0.10

Entonces:

$$M_0 = 1.60 + \frac{11}{11+13}.0.10 = 1.64$$

Podemos calcular el modo gráficamente, construyendo un histograma de frecuencia, por ejemplo, supongamos que al rectángulo del centro le corresponde la mayor frecuencia. Entonces unimos los vértices de la base superior del rectángulo con los vértices de los rectángulos adyacentes al mismo.
Es importante aplicar en el estudio de la matemática, el sentido común. Es decir, si el Mo hubiera sido de 1.74 seguramente está mal, ya que el intervalo modal a simple vista es [1.60; 1.70).



Actividad para la Facilitación de los Aprendizajes

Situación Problemática 14:
El gráfico anterior está incompleto. Dibuje los ejes y ubique la frecuencia y la variable en ambos ejes.

El modo no se ve afectado por valores extremos de la variable, pero sí se ve afectado por su agrupamiento. Si la variable es cualitativa, se puede determinar cualquiera sea la escala en que esté medida. El valor del modo corresponde a aquella clase que presente mayor frecuencia. Puede haber más de una moda. Si existe una sola moda, la distribución se denomina *unimodal*, si hay dos *bimodal* y si hay más de dos *multimodal o polimodal.*
La Moda presenta algunas limitaciones como medida de tendencia central. Por ejemplo:

- No es eficaz si las frecuencias se concentran fuertemente alrededor de algunos valores de la variable
- Una misma distribución con los valores agrupados en clases distintas puede dar distinta moda.

Ejemplo 11: En una distribución de datos puede haber modo, o este no existir.

i) Datos: 1 2 3 4 5
ii) Datos: 1 2 2 3 3
iii) Datos: 1 2 2 3 4

En el i) el Modo no existe, en ii) el Modo es 2 y 3, por último, en iii) el Modo es 2

Actividad para la Facilitación de los Aprendizajes

Situación Problemática 15:

Obtener el Modo de la situación problemática 8

Situación Problemática 16:

Con la distribución de frecuencia del Ejemplo 1, donde la variable es la cantidad de hijos por mujer, determinar el Modo.

Situación Problemática 17:

Con la distribución de frecuencia del Ejemplo 7, donde la variable es monto de las ventas.
a- Calcular el Modo.
b- Determinar gráficamente el Modo.

Veamos ahora el desarrollo de otra Medida de Tendencia Central la **Mediana**

- Mediana (Me)

La Mediana (Me (x)) es el valor que supera y es superado por, a lo sumo, igual cantidad de observaciones. Para calcularla, se efectúan los siguientes pasos:

Paso 1: Se ordenan los datos de menor a mayor.

Paso 2: La Mediana es el valor central si n es impar ó el promedio entre los dos valores centrales si n es par.

Desarrollemos estos pasos en un ejemplo

Ejemplo 12:

Supongamos los siguientes valores: 1 2 4 6 9 $\Rightarrow n = 5$. Como 5 es impar, la mediana es 4 ya que es el valor central.

Supongamos ahora los siguientes valores: 1 2 4 6 8 9 $\Rightarrow n = 6$. Como 6 es par, la mediana será $\frac{4+6}{2} = 5$. O sea, es cinco.
Si los valores se presentan en una tabla de frecuencias, es útil calcular las frecuencias acumuladas para hallar la mediana.

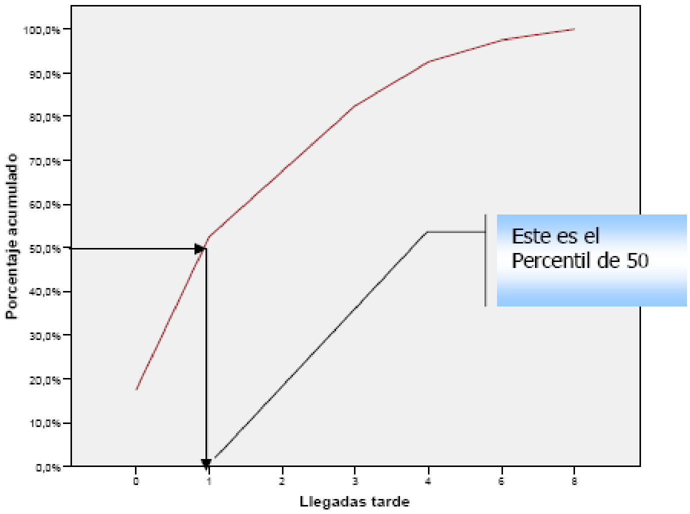
Supongamos que tenemos una tabla donde se registran las llegadas tardes de los empleados de una empresa. La variable x es "cantidad de llegadas tarde" y la frecuencia es "cantidad de personas".

Cuadro 3.5: Llegadas tardes de los empleados

X_i	Frecuencia h_i	Frecuencia Porcentual h_i %	Frecuencia Acumulada Porcentual H_i%
0	7	17,5	17,5
1	14	35,0	52,5
2	6	15,0	67,5
3	6	15,0	82,5
4	4	10,0	92,5
6	2	5,0	97,5
8	1	2,5	100,0
Total	40	100,0	

El gráfico que sigue es un gráfico de líneas y se ha marcado en el eje de las ordenas el 50% que corresponde a $\frac{n}{2}$. Es decir, en nuestro ejemplo $n = 40 \Rightarrow \frac{n}{2} = 20$. Si tomáramos en el eje de las ordenadas en lugar del porcentaje acumulado la frecuencia acumulada 20 obtendríamos el mismo valor que para el 50% y al proyectar el punto desde la curva hacia el eje de las abscisas obtendríamos el valor de la Mediana en el mismo punto que se obtiene el percentil 50.

De esto podemos deducir que el percentil 50 tiene el mismo valor que la Mediana.



Los ejes son: Porcentaje acumulado (H_i%) y llegadas tardes (X_i)

Resumiendo, para encontrar la Mediana en forma gráfica, se ubica en el eje de las ordenadas el valor $\frac{n}{2}$. Luego desde este punto $\frac{n}{2}$ se traza una paralela al eje x hasta intersecar a la curva del gráfico y desde ese punto de intersección se traza una perpendicular hasta el eje de abscisas. Este punto sobre el eje de abscisas es el valor de la Mediana. Podemos decir que el 50% de los empleados llegaron como mucho 1 vez tarde.

Si los datos están agrupados en intervalos como en la siguiente tabla.

Monto de Ventas	Punto Medio (x')	Cantidad de ventas (f_i)	$x' \cdot f_i$	F_i
[100, 105)	102,5	3	307,5	3
[105, 110)	107,5	4	430	7
[110, 115)	112,5	9	1012,5	16
[115, 120)	117,5	6	705	22
[120, 125)	122,5	2	245	24
[125, 130]	127,5	1	127,5	25
		$n = \sum f_i = 25$	2827,5	

Primero calculamos $\frac{n}{2} = \frac{25}{2} = 12{,}5$

Luego, voy a la columna de la frecuencia acumulada y como 12,5 está incluido en $F_3 = 16$, entonces el límite inferior del intervalo mediano es $L_i = 110$. Entonces ya obtuvimos: $\frac{n}{2} = 12{,}5$, $L_i = L_3 = 110$. Además, sabemos que $F_3 = 16$. El número 3 corresponde al subíndice i, ósea $F_i = F_3 = 16$. ¿Quién será entonces F_{i-1}? Sabemos que i=3, entonces i-1= 3-1, por lo tanto $F_{i-1} = F_{3-1} = F_2$, ahora veo cuánto vale F_2. Entonces $F_2 = F_{i-1} = 7$.

Todos estos cálculos los realizamos para poder aplicar la siguiente fórmula:

$$Me = L_i + \frac{\frac{n}{2} - F_{i-1}}{f_i}.a_i$$

En la cual

$L_i =$ es el extremo inferior del intervalo mediano

$F_{i-1} =$ es la frecuencia absoluta acumulada anterior a la que le corresponde al intervalo

$f_i =$ es la frecuencia absoluta del intervalo mediano

$a_i =$ es la amplitud del intervalo mediano

Esta fórmula nos permite calcular la mediana cuando tenemos una distribución de frecuencia.

Reemplazando $a_i = a_3 = 5$ y $f_i = f_3 = 9$ obtenemos:

$$Me = L_i + \frac{\frac{n}{2} - F_{i-1}}{f_i}.a_i = 110 + \frac{12{,}5 - 7}{9}.5 = \mathbf{113{,}05}$$

Entonces $Me = 113{,}05$.

Propiedades de la Mediana

- Cuando calculamos la Mediana no utilizamos todos los valores observados lo cual la limita como medida de tendencia central
- No se puede aplicar a distribuciones de variables cualitativas
- Presenta ciertas ventajas ya que no se ve afectada por valores extremos. Es invariante si aumentamos o disminuimos una observación superior o inferior a ella, dado que solo tenemos en cuenta los valores centrales de la variable. Este es uno de los motivos por los cuales se calcula en distribuciones asimétricas, o cuando hay valores no típicos.
- Si restamos, sumamos, multiplicamos o dividimos cada uno de los datos por un mismo número, las operaciones se trasladan a la mediana. O sea, es *invariante a los cambios de escala o de origen.*
- Es una medida *resistente* es decir si la muestra sufre fluctuaciones pequeñas, no cambia su valor.
- Si tenemos datos *ordinales* tiene sentido calcular la mediana, sin embargo, no tiene sentido calcular la media, ya que se basa en valores solamente numéricos
- En una distribución con intervalos de extremos abiertos se puede calcular la Mediana, en cambio no se puede calcular la media ya que faltaría una marca de clase o punto medio, justamente el del intervalo con extremos abiertos.

Actividad para la Facilitación de los Aprendizajes

Situación Problemática 18:

Con la distribución de frecuencias del Ejemplo 1, donde la variable es la cantidad de hijos por mujer, determinar la Mediana.

Situación Problemática 19:

La siguiente distribución de frecuencias corresponde a la cantidad de personas que se presentaron por día en la oficina de reclamos de una empresa durante los últimos 40 días hábiles.

Cuadro 3.6: Distribución de frecuencias

Cantidad de Personas x_i	Cantidad de Días f_i
2	4
3	7
5	10
6	8
9	5
8	4
11	2

a- Calcular la Mediana

Situación Problemática 20:

Con la distribución de frecuencia de la Situación problemática 11, donde la variable es ingreso por familias, calcular la Mediana.

2.1. Medidas de Variabilidad o Dispersión

Una vez que se ha localizado el "centro" de la distribución con las medidas de tendencia central, la investigación en busca de más información a partir de los conjuntos de datos se dirigirá a la búsqueda de medidas de dispersión o variabilidad. Las **medidas de variabilidad** son aquellas que permiten estudiar, cómo se desvían o varían, en su conjunto, los valores observados de una variable, con respecto a alguna medida de tendencia central.

Desvío Medio:

i) Si la medida de tendencia central es la mediana entonces el *desvío medio* de la variable X, DM (X), será:

$$\text{DM (X)} = \frac{\sum |x_i - Me|}{n}$$

Si los datos están agrupados (véase, Modulo I, unidad I, 3.2) en una distribución de frecuencia, el desvío medio será:

$$\text{DM (X)} = \frac{\sum |x_i - Me| f_i}{n}$$

ii) Si la medida de tendencia central es la media aritmética entonces el *desvío medio*, de la variable X, DM (X), será:

$$\text{DM (X)} = \frac{\sum |x_i - \bar{x}|}{n}$$

Si los datos están agrupados en una distribución de frecuencia, el desvío medio será así tanto con respecto a la mediana como con respecto a la media:

$$\text{DM (X)} = \frac{\sum |x_i - \bar{x}| f_i}{n}$$

Varianza:

Supongamos que tenemos un conjunto de datos {5, 9, 12, 14, 6, 8} donde los datos representar la edad de un grupo de niños.

1- Calculamos el promedio o media aritmética $\bar{x}$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

En este caso, n=6 porque hay seis datos:

$x_1 = 5$

$x_2 = 9$

$x_3 = 12$

$x_4 = 14$

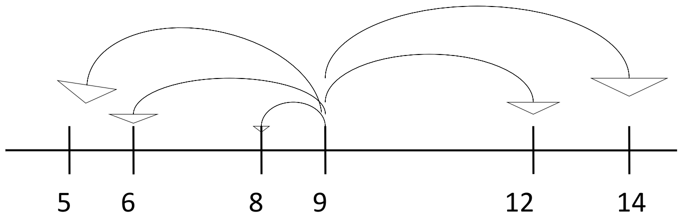
$x_5 = 6$

$x_6 = 8$

$$\overline{x} = \frac{\sum_{i=1}^{6} x_i}{6}$$ Sustituyendo n por 6

$$\overline{x} = \frac{(x_1 + x_2 + x_3 + x_4 + x_5 + x_6)}{6} \Rightarrow \overline{x} = \frac{(5+9+12+14+6+8)}{6} \Rightarrow \overline{x} = 9$$ Este es el promedio

2- Cálculo la distancia que tengo entre los datos y el promedio y la sumo.



↓ ↓ ↓ ↓ ↓ ↓

x_5 x_6 x_2 x_3

$$\begin{aligned}&(x_1 - \overline{x}) + (x_5 - \overline{x}) + (x_6 - \overline{x}) + (x_2 - \overline{x}) + (x_3 - \overline{x}) + (x_4 - \overline{x})\\ &= (5-9) + (6-9) + (8-9) + (9-9) + (12-9) + (14-9)\\ &= (-4) + (-3) + (-1) + 0 + 3 + 5\end{aligned}$$

Como tengo términos negativos necesito elevar cada término al cuadrado, dado que una distancia no podría ser negativa, es siempre mayor o igual que 0.

$$(-4)^2 + (-3)^2 + (-1)^2 + (0)^2 + (3)^2 + (5)^2 = 16 + 9 + 1 + 0 + 9 + 25 = 60$$

Luego lo divido por n que es el tamaño de la muestra y como n=6. Entonces

$\frac{60}{6} = 10$ De este modo obtuvimos la varianza. Luego $S^2(x) = 10$

Los datos anteriores están sin agrupar. Ahora veamos un ejemplo donde los datos estén agrupados con el mismo conjunto de datos anteriores, pero suponiendo que existen más niños. Entonces podremos armar una tabla de f_i como la que sigue:

x	f_i
5	2
6	1
8	3
9	2
12	1
14	2

El cálculo de las distancias quedaría

$$(5-9)^2 \cdot 2 + (6-9)^2 \cdot 1 + (8-9)^2 \cdot 3 + (9-9)^2 \cdot 2 + (12-9)^2 \cdot 1 + (14-9)^2 \cdot 2 = 103$$

Luego dividiendo por n y calculando obtenemos que $S^2(x) = 17{,}16$.

Anteriormente hemos sumado término a término cada una de las distancias, pero la matemática cuenta con fórmulas que sintetizan lo que hemos escrito anteriormente. Luego la varianza de la variable X, S^2 (X), es el promedio aritmético del cuadrado de las desviaciones con respecto a la media aritmética escrita como fórmula es:

$$S^2\text{(X)} = \frac{\sum (x_i - \overline{x})^2}{n} \qquad (*)$$

Si se tienen los datos agrupados la fórmula será:

$$S^2\text{(X)}= \frac{\sum (x_i - \overline{x})^2 f_i}{n} \qquad (**)$$

Para poder facilitar, y en algunos casos acelerar el cálculo de la varianza se utilizan las siguientes "fórmulas de trabajo":

-Datos sin Agrupar. Fórmula de Trabajo[1]

$$S^2(X) = \frac{\sum x_i^2}{n} - \overline{x}^2$$

-Datos Agrupados. Fórmula de Trabajo[2].

$$S^2(X) = \frac{\sum x_i^2 f_i}{n} - \overline{x}^2$$

La varianza es una medida de variación porque está midiendo, en su conjunto, las diferencias entre cada valor individual observado de la variable y la media aritmética.
Cuanto mayor es el valor numérico de la varianza, mayor es la variabilidad de los datos y, consecuentemente, menor la representatividad de la media aritmética.

Ejemplo 12:
Los siguientes datos corresponden al contenido, en gramos, de 5 envases de galletas.

850; 845; 865; 863

Calcular la varianza, utilizando

a) La definición
b) La fórmula de trabajo

Solución:

La variable es,

X: Contenidos de los envases, en gramos.

a) Dado que los datos están sin agrupar en una distribución de frecuencia, por definición, hay que calcular

[1] Esta fórmula se deduce matemáticamente de la fórmula (*) para datos sin agrupar.
[2] Esta fórmula se deduce matemáticamente de la fórmula (**) para datos agrupar

$$S^2(X) = \frac{\sum (x_i - \bar{x})^2}{n}$$

Primero hay que calcular la media aritmética, segundo, las respectivas desviaciones, cuya suma debe ser igual a cero, y, en tercer lugar, el cuadrado de cada desviación, cuya suma es la suma de cuadrados.
Para facilitar las operaciones se construye el siguiente cuadro

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
845	-10,4	108,16
850	-5,4	29,16
854	-1,4	1,96
863	7,6	57,76
865	9,6	92,16
$\sum x_i$ = 4277	$\sum (x_i - \bar{x})$ = 0,0	$\sum (x_i - \bar{x})^2$ =289,20

$$\bar{x} = \frac{\sum x_i}{n} = \frac{4277}{5} = 855{,}4$$

La suma de cuadrados es

SCx = $\sum (x_i - \bar{x})^2 = 289{,}20$

Luego, la varianza es

$$S^2(X) = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{289{,}20}{5} = 57{,}84$$

b) Dado que los datos están sin agrupar en una distribución de frecuencia, la fórmula de trabajo a utilizar es

$$S^2(X) = \frac{\sum x_i^2}{n} - \bar{x}^2$$

Para facilitar los cálculos se construye el siguiente cuadro

x_i	x_i^2
845	714025
850	722500
854	729316
863	744769
865	748225

$\sum x_i$ = 4277	$\sum x_i^{\,2}$ = 3658835

El contenido medio ya calculado, es

$$\bar{x} = 855{,}4$$

Reemplazo en la fórmula de trabajo

$$S^2(X) = \frac{\sum x_i^2}{n} - \bar{x}^2 = \frac{3658835}{5} - (855{,}4)^2 = 57{,}84$$

De esta manera, se comprueba empíricamente que el valor de la varianza se mantiene inalterable cualquiera sea la fórmula que se utilice

Algunas propiedades de la Varianza

- La varianza es, necesariamente, un número real no negativo

$$S^2\text{(X)} = \frac{\sum (x_i - \bar{x})^2}{n} \geq 0$$

La suma de cuadrados del numerador no puede ser negativa, y el número de observaciones del denominador necesariamente es positivo. Luego el cociente de dos números positivo es positivo.

- La varianza de una constante es nula

Si Y = constante entonces S^2(Y)=0

Si todos los valores observados de la variable son iguales, entonces no hay variabilidad, luego la varianza es cero.

Actividad para la Facilitación de los Aprendizajes

Situación Problemática 21: Con la distribución de frecuencia del Ejemplo 1, donde la variable es la cantidad de hijos por mujer, calcular la Varianza utilizando

a) La definición
b) La fórmula de trabajo

Situación Problemática 22: Con la distribución de frecuencia del Ejemplo 4, donde la variable es monto de las ventas, en pesos, calcular la varianza utilizando

a) La fórmula de trabajo.

Desvío Estándar

El valor numérico de la *varianza* de la variable X queda expresado en otra dimensión, el cuadrado de la magnitud de la variable. Esto hace que su interpretación sea dificultosa y que sea necesario tener una medida de variabilidad cuya magnitud esté en la misma dimensión que la de la variable. Para ello aplicaré la raíz cuadrada neutralizando de esta manera el cuadrado. El *desvío estándar* de la variable X, es la raíz cuadrada positiva de la varianza. Datos sin agrupar.

$$S(X) = +\sqrt{S^2(X)}$$

$$S(X)=+\sqrt{\frac{\sum(x_i-\overline{x})^2}{n}}=+\sqrt{\frac{\sum x_i^2}{n}-\overline{x}^2}$$

Observación: Por este motivo a la varianza se la simboliza $S^2(X)$

Datos agrupados $S(X)=+\sqrt{\frac{\sum(x_i-\overline{x})^2 f_i}{n}}=+\sqrt{\frac{\sum x_i^2 f_i}{n}-\overline{x}^2}$

El *desvío estándar*, es una medida de *variabilidad absoluta*, porque su valor numérico está expresado en la misma dimensión de la variable manteniendo la magnitud.
Esta medida es la adecuada para establecer la variabilidad que presentan los valores observados de la variable, en su conjunto, con respecto a la media aritmética.
En estadística es importante interpretar las medidas que se van obteniendo.
En este caso, interpretando el desvío estándar podemos decir que:
"Los contenidos en gramos de todos los envases de galletitas tienen en promedio un desvío de 7,6 gramos respecto de la media aritmética."

Ejemplo 13:
Con el ejemplo 12, donde la variable es contenido de los envases de galletas, en gramos, calcular el desvío estándar e interpretar.

Solución:
X: Contenido de los envases de galletas, en gramos.
En el ejemplo 12 habíamos calculado la varianza

$$S^2(\text{X}) = \frac{\sum(x_i-\overline{x})^2}{n}=\frac{289{,}20}{5}=57{,}84$$

Luego, el desvío estándar es

$$S(X)=\sqrt{S^2}=\sqrt{57{,}84}=7{,}605261337\approx 7{,}6$$

La unidad de medida de la variable es gramos, por lo tanto, el desvío estándar, por ser una medida de variabilidad absoluta, también está expresado en gramos.

S(X)=7,6 gramos

Y podemos decir que el contenido de todos los envases de galleta, en promedio, tiene una desviación respecto de la media aritmética de 7,6 gramos

Propiedades del desvío estándar

Se deducen fácilmente de las propiedades de la varianza, dado que el desvío estándar es la raíz cuadrada positiva de la varianza.

- El desvío estándar es siempre un valor no negativo.

  En símbolos: $S(x)\geq 0$
- El desvío estándar toma en cuenta las desviaciones de todos los valores de la variable.
- Si a todos los valores de la variable se le suma una misma constante el desvío estándar no varía.
- Si a todos los valores de la variable se multiplican por una misma constante, el desvío estándar queda multiplicada por el valor absoluto de dicha constante.

Situación Problemática 23: Con la distribución de frecuencia del Ejemplo 4, donde la variable es monto de las ventas, en pesos, calcular el desvío estándar e interpretar.

Coeficiente de Variación

Para poder establecer si la dispersión alrededor de la media aritmética que presenta una variable es baja o no, es necesario que el valor numérico del desvío estándar (la medida de variabilidad absoluta con la que habitualmente se mide la dispersión) sea comparada o relacionada con ella (la media aritmética).
El *coeficiente de variación* de la variable X, CV(X), es el cociente entre el desvío estándar y la media aritmética de dicha variable.

$$CV(X) = \frac{S(X)}{\bar{x}}$$

El coeficiente de variación se suele expresar en porcentaje:

$$CV(X) = \frac{S(X)}{\bar{x}} \cdot 100$$

El coeficiente de variación es un número puro, desprovisto de magnitud. Es una medida de variabilidad relativa. Relaciona el desvío estándar con la media aritmética. Su valor numérico permite establecer criterios generales acerca de la homogeneidad de los datos, de la representatividad de la media aritmética y la comparación con la variabilidad de otras variables, aunque las unidades de medida o las magnitudes sean distintas.
Un criterio generalmente aceptado es:
Si el CV(X) $\leq 0{,}10$ se puede considerar que los datos son homogéneos y consecuentemente la media aritmética representativa.
El coeficiente de variación permite comparar las dispersiones de dos distribuciones distintas, siempre que sus medias sean positivas. Se calcula para cada una de las distribuciones y los valores que se obtienen se comparan entre sí.
La mayor dispersión corresponderá al valor del coeficiente de variación mayor.

Propiedades del coeficiente de variación

- Sólo se debe calcular para variables con todos los valores positivos. Todo índice de variabilidad es esencialmente no negativo. Las observaciones pueden ser positivas o nulas, pero su variabilidad debe ser siempre positiva. De ahí que sólo debemos trabajar con variables positivas, para la que tenemos con seguridad que $\bar{x} > 0$
- Si a todos los valores de la variable se le suma una misma constante el coeficiente de variación queda alterado.

Ejemplo 14:

Supongamos que una distribución tiene $\bar{x} = 140$ y $S(x) = 28{,}28$ y otra $\bar{x} = 150$ y $S(x) = 24$. ¿Cuál de las dos presenta mayor dispersión?

$$CV_1 = \frac{28{,}28}{140} \cdot 100 = 20{,}2\%$$

$$CV_2 = \frac{24}{150} \cdot 100 = 16\%$$

En conclusión, la primera distribución presenta mayor dispersión.

Actividad para la Facilitación de los Aprendizajes

Situación Problemática 24: Con la distribución de frecuencia del Ejemplo 4, donde la variable es monto de las ventas, en pesos, calcular e interpretar la variabilidad relativa.

Situación Problemática 25: En la materia de estadística, un alumno obtuvo las siguientes notas:

10, 9, 7, 7

a) Calcule el promedio e interpretar
b) Calcule el modo e interpretar
c) Encuentre la mediana e interpretar
d) Calcule la varianza y el desvío estándar e interpretar el desvío
e) Halle el coeficiente de variación ¿Son los datos homogéneos?

Situación Problemática 26:
Los siguientes datos corresponden a la edad (en años cumplidos) de 87 alumnos de la cátedra de estadística:

31 37 24 27 25 30 24 24 33 24 30 31 27 31 26
27 26 26 28 27 32 27 27 31 33 34 24 32 30 23
28 26 25 26 33 33 29 32 35 26 22 24 33 32 30
34 30 37 44 28 29 32 27 22 25 24 47 28 23 26
31 34 37 24 25 22 25 27 26 27 24 33 28 47 26
30 33 27 28 23 29 21 22 31 23 26 27

a- Hacer la tabla de distribución de frecuencia
b- Calcular el promedio e interpretar
c- Calcular el modo e interpretar
d- Encuentre la mediana e interpretar
e- Calcular la varianza
f- Calcular el desvío estándar e interpretar
g- Hallar el coeficiente de variación ¿Los datos son homogéneos?
h- Realizar un gráfico de barras

Situación Problemática 27: La evolución del precio de la nafta común (en dólares por litro) en la República Argentina para el periodo Ago-88 a Abr-89 fue:

Ago-88	Sep-88	Oct-88	Nov-88	Dic-88	Ene-89	Feb-89	Mar-89	Abr-89
0,29	0,30	0,32	0,33	0,34	0,35	0,36	0,37	0,39

a) ¿Cuál es la variable?
b) Calcular el promedio e interpretar
c) Calcular la varianza
d) Calcular el desvío estándar e interpretar
e) Hallar el coeficiente de variación ¿Los datos son homogéneos?

Situación Problemática 28: Un comercio de venta de calzado ha vendido durante los últimos tres años 4.500 pares de zapatos para mujer. EL siguiente cuadro muestra la distribución de las ventas de calzado de mujer por medida.

x_i	f_i

35	10
36	60
37	200
38	910
39	1.720
40	1.170
41	350
42	70
43	10

a- Calcular el promedio
b- Calcular el modo e interpretar
c- Encuentre la mediana e interpretar
d- Calcular la varianza
e- Calcular el desvío estándar e interpretar
f- Hallar el coeficiente de variación ¿Los datos son homogéneos?

Situación Problemática 29:
Con un micrómetro, se realizan mediciones del diámetro de un balero, que tienen una media de 4,03 mm y una desviación estándar de 0,012 mm; con otro micrómetro se toman mediciones de la longitud de un tornillo que tiene una media de 1,76 pulgadas y una desviación estándar de 0,0075 pulgadas. ¿Cuál de los dos micrómetros presenta una variabilidad relativamente menor?

Situación Problemática 30:
Dos profesores que imparten diferentes materias a un mismo grupo deciden investigar como es el coeficiente de variación de una y otra materia, para lo cual se obtiene la media y la desviación estándar respectivamente, por lo que:
Resultado de la materia A: $\bar{x} = 6.3$ y $S(x) = 1{,}2$
Resultado de la materia B: $\bar{x} = 8$ y $S(x) = 3$
¿Cuál de las dos materias presenta una variabilidad relativamente mayor?

2.2. Medidas de Posición

Las *medidas de posición* se usan para describir la posición que tiene el valor de un dato específico en relación con el resto de los datos. Dos de las medidas de posición más conocidas son los *cuartiles* y los *percentiles*.

Cuartiles:

Son los valores de la variable que divide en cuartos a los *datos ordenados;* cada conjunto de datos posee tres cuartiles. El *primer cuartil, Q_1,* es un número tal que cuando mucho el 25% de los datos es menor en valor que Q_1 y cuando mucho el 75% de los datos en mayor que Q_1. El *segundo cuartil* es la *mediana*, la interpretación es similar a la del Q_1. O sea Q_2 es un número tal que cuando mucho el 50% de los datos es menor en valor que Q_2 y como mucho el 50% de los datos es mayor que Q_2. El *tercer cuartil, Q_3*, es un número tal que cuando mucho el 75% de los datos es menor en valor que Q_3 y cuando mucho el 25% de los datos es mayor que Q_3.

25%	75%

Q_1

50%	50%

Q_2

75%	25%

Q_3

$$Q_j = L_i + \frac{\left(j \cdot \frac{n}{4} - F_{i-1} \right)}{f_i} \cdot a_i \qquad j = 1,2,3$$

$L_i =$ es el extremo inferior del intervalo mediano

$F_{i-1} =$ es la frecuencia absoluta acumulada anterior a la que le corresponde al intervalo

$f_i =$ es la frecuencia absoluta del intervalo mediano

$a_i =$ es la amplitud del intervalo mediano

Si j = 1 se denomina *Primer Cuartil.*

Si j = 2 se denomina *Segundo Cuartil.* Se observa que $Q_2 = Me$

Si j = 3 se denomina *Tercer Cuartil.*

<u>Ejemplo 15</u>:

Calcular los cuartiles de la siguiente tabla:

	f_i	F_i
[50; 60)	7	7
[60; 70)	11	18
[70; 80)	16	34
[80; 90)	15	49
[90; 100)	9	58
[100, 110)	4	62
[110; 120)	3	65
	65	

<u>Solución</u>:

<u>Cálculo del primer cuartil</u>:

- Primero calculamos $j \cdot \frac{n}{4} = 1 \cdot \frac{65}{4} = 16{,}25$

Luego, voy a la columna de la frecuencia acumulada y como 16,25 está incluido en $F_2 = 18$, entonces el límite inferior del intervalo mediano es $L_i = 60$. Entonces

ya obtuvimos: $\frac{n}{4} = 16{,}25$, $L_i = L_2 = 60$. Además, sabemos que $F_2 = 18$. El número 2 corresponde al subíndice i, ósea $F_i = F_2 = 18$. ¿Quien será entonces F_{i-1}? Sabemos que i=2, entonces i-1= 2-1, por lo tanto $F_{i-1} = F_{2-1} = F_1$, ahora veo cuánto vale F_1. Entonces $F_1 = F_{i-1} = 7$.
Todos estos cálculos los realizamos para poder aplicar la siguiente fórmula

- $$Q_1 = 60 + \frac{16{,}25 - 7}{11} \cdot 10 = 68{,}40$$

Cálculo del segundo cuartil:

- Primero calculamos $j \cdot \frac{n}{4} = 2 \cdot \frac{65}{4} = 32{,}5$

Luego, voy a la columna de la frecuencia acumulada y como 32,5 está incluido en $F_3 = 34$, entonces el límite inferior del intervalo mediano es $L_i = 70$. Entonces ya obtuvimos: $j \cdot \frac{n}{4} = 32{,}5$, $L_i = L_3 = 70$. Además, sabemos que $F_3 = 34$. El número 3 corresponde al subíndice i, ósea $F_i = F_3 = 34$. ¿Quien será entonces F_{i-1}? Sabemos que i=3, entonces i-1= 3-1, por lo tanto $F_{i-1} = F_{3-1} = F_2$, ahora veo cuánto vale F_2. Entonces $F_2 = F_{i-1} = 18$.
Todos estos cálculos los realizamos para poder aplicar la siguiente fórmula

- $$Q_2 = 70 + \frac{32{,}5 - 18}{16} \cdot 10 = 79{,}06$$

Cálculo del tercer cuartil:

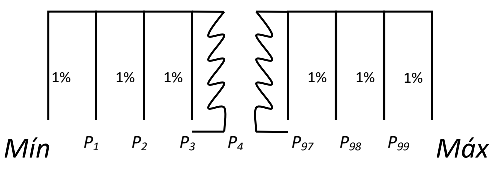
- Primero calculamos $j \cdot \frac{n}{4} = 3 \cdot \frac{65}{4} = 48{,}75$

Luego, voy a la columna de la frecuencia acumulada y como 48,75 está incluido en $F_4 = 49$, entonces el límite inferior del intervalo mediano es $L_i = 80$. Entonces ya obtuvimos: $j \cdot \frac{n}{4} = 48{,}75$, $L_i = L_4 = 80$. Además, sabemos que $F_4 = 49$. El número 4 corresponde al subíndice i, ósea $F_i = F_4 = 49$. ¿Quién será entonces F_{i-1}? Sabemos que i=4, entonces i-1= 4-1, por lo tanto $F_{i-1} = F_{4-1} = F_3$, ahora veo cuánto vale F_3. Entonces $F_3 = F_{i-1} = 34$.
Todos estos cálculos los realizamos para poder aplicar la siguiente fórmula

- $$Q_3 = 80 + \frac{48{,}75 - 34}{15} \cdot 10 = 89{,}83$$

Percentiles:

Son los valores de la variable que dividen a un conjunto de *datos ordenados* en 100 subconjuntos iguales; cada conjunto de datos tiene 99 percentiles. El *k-ésimo percentil, P_k,* es un valor que P_k y cuando mucho (100 - K) % de los datos es mayor.



$$P_k = L_i + \frac{\left(\frac{k \cdot n}{100} - F_{i-1}\right)}{f_i} \cdot a_i \qquad k = 1,2,\ 3\ldots 99$$

<u>Ejemplo 16</u>:
Calcular los percentiles 35 y 60 de la tabla del ejemplo 15.

<u>Cálculo del percentil 35:</u>

- Primero calculamos $\frac{k \cdot n}{100} = \frac{35 \cdot 65}{100} = 22{,}75$
- $P_{35} = 70 + \frac{22{,}75 - 18}{16} \cdot 10 = 72{,}97$

<u>Cálculo del percentil 60:</u>

- Primero calculamos $\frac{k \cdot n}{100} = \frac{60 \cdot 65}{100} = 39$
- $P_{60} = 80 + \frac{39 - 34}{15} \cdot 10 = 83{,}33$

Actividad para la Facilitación de los Aprendizajes

<u>Situación Problemática 31</u>:

Calcular la Mediana de la Situación Problemática 8.

<u>Situación Problemática 32:</u>

Con una puntuación de 100 José se situó en el percentil del 80% respecto al total de alumnos de su clase. Supongamos que el profesor decide subir 5 puntos a todos los alumnos. ¿En qué percentil estaría José?

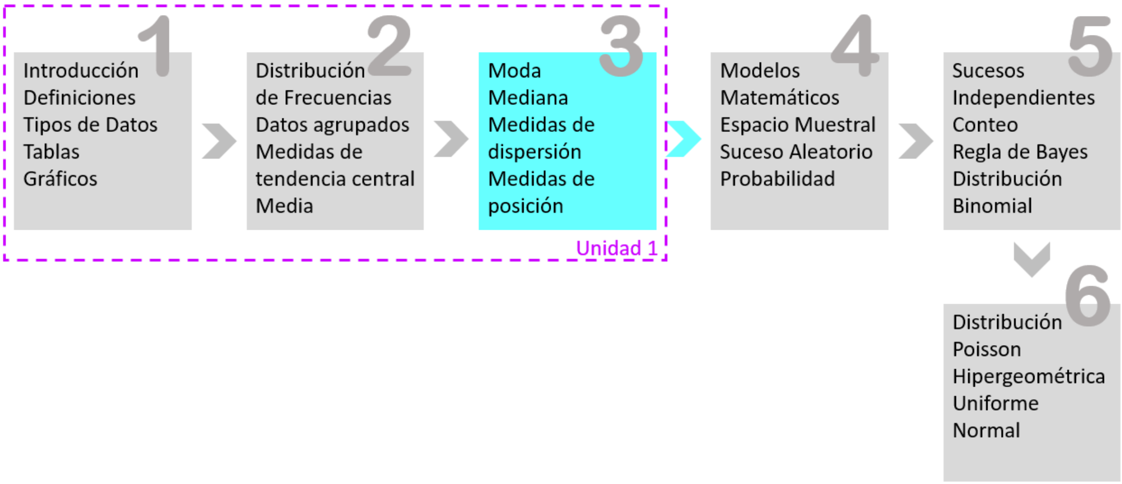
<u>Situación Problemática 33:</u>

Los siguientes datos corresponden a los pesos de 21 personas.

	f_i	F_i
[38; 45)	4	
[45; 52)	1	
[52; 59)	8	
[59; 66)	2	
[66; 73)	6	
	21	

Calcular:

a- La frecuencia acumulada
b- La mediana
c- Los tres cuartiles


1
Introducción
Definiciones
Tipos de Datos
Tablas
Gráficos
2
Distribución
de Frecuencias
Datos agrupados
Medidas de
tendencia central
Media
3
Moda
Mediana
Medidas de
dispersión
Medidas de
posición
Unidad 1
4
Modelos
Matemáticos
Espacio Muestral
Suceso Aleatorio
Probabilidad
5
Sucesos
Independientes
Conteo
Regla de Bayes
Distribución
Binomial
6
Distribución
Poisson
Hipergeométrica
Uniforme
Normal