



INSTITUTO NACIONAL GENERAL JOSÉ  
MIGUEL CARRERA

DEPARTAMENTO DE MATEMÁTICAS

---

# Apunte Completo: Probabilidad y Estadística

---

*Probabilidad y Estadística - Sección 2025*

**Autor:**  
Felipe Colli Olea

**Profesor:**  
Sergio Díaz

2 de julio de 2025

## Resumen

### Resumen

Este apunte ofrece una introducción completa a los conceptos fundamentales de la probabilidad y la estadística. Comienza con la definición de términos esenciales como población, muestra, parámetro y estadígrafo, y la clasificación de variables en cualitativas y cuantitativas. Posteriormente, se profundiza en la estadística descriptiva, abordando las medidas de tendencia central (media, mediana, moda) y de dispersión (rango, varianza, desviación estándar), incluyendo sus propiedades y demostraciones matemáticas. El documento también explora diversas técnicas de muestreo, diferenciando entre métodos probabilísticos (aleatorio simple, estratificado, por conglomerados, sistemático) y no probabilísticos. Finalmente, se introduce la combinatoria mediante los principios aditivo y multiplicativo como base para el cálculo de probabilidades.



Figura 1: Gatito para hacer bonito el apunte.

# Índice

<b>1. 14/03</b>	<b>4</b>
1.1. Materia . . . . .	4
1.2. Ejercicios: . . . . .	5
<b>2. 17/03</b>	<b>7</b>
2.1. Materia . . . . .	7
<b>3. 26/03</b>	<b>9</b>
3.1. Ejercicios: . . . . .	9
<b>4. 04/04</b>	<b>12</b>
4.1. Población y Muestra . . . . .	12
4.2. Muestreo . . . . .	12
4.2.1. Representatividad . . . . .	12
4.3. Tipos de Muestreo Probabilístico . . . . .	12
4.3.1. Muestreo Aleatorio Simple (M.A.S) . . . . .	13
4.3.2. Muestreo Estratificado . . . . .	13
4.3.3. Muestreo por Conglomerados (Clusters) . . . . .	13
4.3.4. Muestreo Aleatorio Sistemático . . . . .	13
4.4. Muestreos No Probabilísticos . . . . .	13
4.4.1. Muestreo por Cuotas . . . . .	13
4.4.2. Muestreo Bola de Nieve . . . . .	13
4.4.3. Muestreo por Juicio o Conveniencia . . . . .	14
<b>5. 10/04</b>	<b>15</b>
5.1. Medidas de Dispersión . . . . .	15
5.1.1. Rango (o Amplitud Total): . . . . .	15
5.1.2. Desviación Media (DM): . . . . .	15
5.1.3. Varianza ( $\sigma^2$ para población, $s^2$ para muestra): . . . . .	16
5.1.4. Desviación Estándar (o Típica) ( $\sigma$ para población, $s$ para muestra): . . . . .	16
5.1.5. Propiedades de $\sigma$ y $\sigma^2$ (usando la definición con denominador $n$ ) . . . . .	16
<b>6. 16/04</b>	<b>18</b>
6.1. Demostración Propiedad 4 (Multiplicación por una constante)	18

6.2. Ejercicio . . . . .	19
6.3. Demostración Propiedad 5 (Fórmula Computacional de la Varianza) . . . . .	20
<b>7. 23/04</b>	<b>21</b>
7.1. Demostraciones de Propiedades Relacionadas con el Valor de $\sigma$	21
7.1.1. Propiedad 6: $\sigma^2 = \sigma \iff \sigma = 0 \vee \sigma = 1$ . . . . .	21
7.1.2. Propiedad 7: $\sigma^2 < \sigma \iff 0 < \sigma < 1$ . . . . .	21
7.1.3. Propiedad 8: $\sigma^2 > \sigma \iff \sigma > 1$ . . . . .	22
<b>8. 30/05</b>	<b>23</b>
8.1. Cuartiles . . . . .	23
8.2. Quintiles . . . . .	23
8.3. Deciles . . . . .	23
8.4. Percentiles . . . . .	23
<b>9. 11/06</b>	<b>24</b>
9.1. Combinatoria . . . . .	24
9.2. Ejercicios . . . . .	24
<b>10.07/07</b>	<b>26</b>
10.1. Combinatoria . . . . .	26
10.1.1. Permutacion simple . . . . .	26
10.1.2. Permutacion con repetición . . . . .	26
10.1.3. Permutacion Circular . . . . .	26
10.2. Ejercicios . . . . .	27

# 1. 14/03

## 1.1. Materia

**Población:** Conjunto de todos los elementos que se quieren estudiar. Cuando la información deseada está disponible para todos los objetos de la población, lo llamamos **censo**. En la práctica es muy difícil o casi imposible realizar un censo.

**Muestra:** Subconjunto de la población que se mide u observa.

**Parámetro:** Es una medición numérica que describe alguna característica de una población.

**Estadístico (o estadígrafo):** Es una medición numérica que describe alguna característica de la muestra.

**Variables cualitativas:**

- Se describen mediante palabras o categorías.
- Se usan para categorizar a los individuos o para identificar.
- Sirven para comprender aspectos subjetivos y complejos.
- Se pueden clasificar en nominales y ordinales.
- Ejemplos: el color del cabello, el deporte favorito, la comida favorita, el lugar de nacimiento.

**Variables cuantitativas:**

- Se expresan mediante números, es decir, se pueden contar o medir.
- Permiten más operaciones matemáticas.
- Se pueden usar para conocer fenómenos o situaciones a través de la recolección y generación de números y datos.
- Ejemplos: la edad, los ingresos, el peso, la altura, la presión, la humedad o cantidad de hermanos.

## 1.2. Ejercicios:

Para cada una de las siguientes situaciones, identifica la población de interés, la variable estadística, clasifícala, y entrega un ejemplo de cuál podría ser una posible muestra.

1. Un investigador universitario desea estimar la proporción de ciudadanos chilenos de la *GEN X* que están interesados en iniciar sus propios negocios.
  - a) **Población:** Chilenos de la Generación X.
  - b) **Muestra (ejemplo):** 500 chilenos de la Generación X seleccionados aleatoriamente del padrón electoral.
  - c) **Variable:** Interés en iniciar un negocio (Sí/No) (*cualitativa nominal*).
2. Durante más de un siglo, la temperatura corporal normal en seres humanos ha sido aceptada como 37°C. ¿Es así realmente? Los investigadores desean estimar el promedio de temperatura de adultos sanos en Chile.
  - a) **Población:** Adultos sanos en Chile.
  - b) **Muestra (ejemplo):** Adultos sanos de Santiago seleccionados de diversos centros de salud.
  - c) **Variable:** Temperatura corporal (*cuantitativa continua*).
3. Un ingeniero municipal desea estimar el promedio de consumo semanal de agua para unidades habitacionales unifamiliares en la ciudad.
  - a) **Población:** Unidades habitacionales unifamiliares de la ciudad.
  - b) **Muestra (ejemplo):** 200 unidades habitacionales unifamiliares de distintos barrios de la ciudad, seleccionadas aleatoriamente.
  - c) **Variable:** Consumo semanal de agua (en litros) (*cuantitativa continua*).
4. El National Highway Safety Council desea estimar la proporción de llantas para automóvil con dibujo o superficie de rodadura insegura, entre todas las llantas manufacturadas por una empresa específica durante el presente año de producción.

- a) **Población:** Todas las llantas para automóvil manufacturadas por la empresa específica durante el presente año de producción.
  - b) **Muestra (ejemplo):** Una selección aleatoria de llantas producidas en diferentes lotes o días del año.
  - c) **Variable:** Estado de la superficie de rodadura (segura/insegura) (*cualitativa nominal*).
- 5. Un politólogo desea estimar si la mayoría de los residentes adultos de una región están a favor de una legislatura unicameral.
  - a) **Población:** Residentes adultos de la región.
  - b) **Muestra (ejemplo):** Residentes adultos de varias comunas seleccionadas aleatoriamente de la región.
  - c) **Variable:** Opinión sobre la legislatura unicameral (a favor/en contra/indeciso) (*cualitativa nominal*).
- 6. Un científico del área médica desea determinar el tiempo promedio para que se vuelva a presentar cierta enfermedad infecciosa, una vez que las personas se recuperan de ella por primera vez.
  - a) **Población:** Personas que se han recuperado de la enfermedad infecciosa por primera vez.
  - b) **Muestra (ejemplo):** Pacientes recuperados seleccionados de registros médicos de diversos hospitales.
  - c) **Variable:** Tiempo hasta la recurrencia de la enfermedad (*cuantitativa continua*).
- 7. Un ingeniero electricista desea determinar si el promedio de vida útil de transistores de cierto tipo es mayor que 500 horas.
  - a) **Población:** Todos los transistores de cierto tipo.
  - b) **Muestra (ejemplo):** Una muestra de 100 transistores de ese tipo, seleccionados aleatoriamente de la producción.
  - c) **Variable:** Vida útil del transistor (en horas) (*cuantitativa continua*).

## 2. 17/03

### 2.1. Materia

Medidas de Tendencia Central

Medidas de Posición

Medidas de Dispersión

#### Tablas de Frecuencia: Conceptos Básicos

- **Dato o Intervalo:** Información (variable) que se estudia en estadística.
- **Marca de Clase ( $c_i$ ):** Promedio entre los extremos de un intervalo.
- **Amplitud de un intervalo:** Es la diferencia entre el límite superior y el límite inferior del intervalo.

#### Tipos de Frecuencia:

- **Frecuencia Absoluta ( $f_i$ ):** Cantidad de veces que se repite un dato o que los datos caen en un intervalo.
- **Frecuencia Absoluta Acumulada ( $F_i$ ):** Suma de las frecuencias absolutas hasta determinado dato o intervalo.  $F_i = \sum_{j=1}^i f_j$ .
- **Frecuencia Relativa ( $h_i$  o  $f_{ri}$ ):** Es la proporción (fracción, decimal o porcentaje) de observaciones que corresponden a cierto valor o intervalo. ( $h_i = \frac{f_i}{n}$ ), donde  $n$  es el número total de datos.
- **Frecuencia Relativa Acumulada ( $H_i$ ):** Es la proporción (fracción, decimal o porcentaje) de la frecuencia acumulada hasta cierto dato o intervalo. ( $H_i = \frac{F_i}{n} = \sum_{j=1}^i h_j$ ).



## Medidas de Tendencia Central:

- **Media Aritmética ( $\bar{x}$ ):** Es el cociente entre la suma de todos los datos y el número total de datos ( $n$ ). Si se tienen  $n$  datos  $x_1, x_2, \dots, x_n$ :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Para datos agrupados en una tabla de frecuencia con  $k$  clases:

Marca ( $c_i$ )	Frec. ( $f_i$ )
$c_1$	$f_1$
$c_2$	$f_2$
$\vdots$	$\vdots$
$c_k$	$f_k$

$$\bar{x} = \frac{c_1 f_1 + c_2 f_2 + \dots + c_k f_k}{f_1 + f_2 + \dots + f_k} = \frac{\sum_{i=1}^k c_i f_i}{n}$$

(donde  $n = \sum_{i=1}^k f_i$ )

- **Mediana ( $M_e$ ):** Es el valor que ocupa la posición central de la muestra cuando los datos se encuentran ordenados. **Si la muestra tiene un número par de datos, la mediana es la media aritmética de los dos términos centrales.**
  - Si  $n$  es impar, la posición es  $\frac{n+1}{2}$ .  $M_e = x_{(\frac{n+1}{2})}$ .
  - Si  $n$  es par, las posiciones son  $\frac{n}{2}$  y  $\frac{n}{2} + 1$ .  $M_e = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2} + 1)}}{2}$ .
- **Moda ( $M_o$ ):** Es el dato o intervalo con la mayor frecuencia absoluta. La muestra puede ser:
  - **Amodal:** No presenta moda.
  - **Unimodal:** Una sola moda.
  - **Bimodal:** Dos modas.
  - **Polimodal (o Multimodal):** Más de dos modas.

### 3. 26/03

#### 3.1. Ejercicios:

Si las notas de Esteban en una asignatura son 3, 4, 6, 3, 5, 5, 6, 3, 4 y de estas notas se cambian un 6 por un 7. ¿Cuál(es) de las siguientes medidas de tendencia central cambia(n)?

1. La moda
2. La mediana
3. La media aritmética

*Solución:* Notas originales (ordenadas): 3, 3, 3, 4, 4, 5, 5, 6, 6.  $n = 9$ .

- Moda original: 3 (frecuencia 3).
- Mediana original: El dato en la posición  $\frac{9+1}{2} = 5$ . Mediana = 4.
- Media original:  $\frac{3 \cdot 3 + 2 \cdot 4 + 2 \cdot 5 + 2 \cdot 6}{9} = \frac{39}{9} \approx 4,33$ .

Notas nuevas (se cambia un 6 por un 7), ordenadas: 3, 3, 3, 4, 4, 5, 5, 6, 7.  $n = 9$ .

- Moda nueva: 3 (sigue siendo la más frecuente).  $\rightarrow$  No cambia.
- Mediana nueva: El dato en la posición 5 sigue siendo 4.  $\rightarrow$  No cambia.
- Media nueva:  $\frac{39-6+7}{9} = \frac{40}{9} \approx 4,44$ .  $\rightarrow$  Cambia.

**Respuesta:** Solo la media aritmética cambia (opción 3).

La siguiente tabla muestra los valores de una variable  $X$  y sus respectivas frecuencias. ¿Cuál es el valor de la mediana?

$X_i$	Frecuencia ( $f_i$ )	Frecuencia Acumulada ( $F_i$ )
4	4	4
5	8	12
6	10	22
7	20	42
8	8	50
<b>Total</b>	<b>n=50</b>	

*Solución:* Total de datos  $n = 50$  (par). La mediana es el promedio de los datos en las posiciones  $\frac{50}{2} = 25$  y  $\frac{50}{2} + 1 = 26$ . Buscamos en la Frecuencia Acumulada ( $F_i$ ):

- Hasta  $X = 6$ , se acumulan 22 datos.
- Para  $X = 7$ , la frecuencia acumulada llega a 42. Esto significa que los datos desde la posición 23 hasta la 42 son iguales a 7.

Por lo tanto, el dato 25 y el dato 26 son ambos 7. La mediana es  $M_e = \frac{7+7}{2} = 7$ .

**Respuesta:** La mediana es 7.

De acuerdo a la siguiente muestra  $a + 2, a + 4, a + 6, a + 6, a + 6, a + 4, a + 2$ , la suma de la mediana y la moda es:

*Solución:* Muestra ordenada:  $a + 2, a + 2, a + 4, a + 4, a + 6, a + 6, a + 6$ .  $n = 7$ .

- **Moda ( $M_o$ ):** El dato más frecuente es  $a + 6$ .
- **Mediana ( $M_e$ ):** Como  $n = 7$  (impar), la mediana es el dato en la posición  $\frac{7+1}{2} = 4$ . El cuarto dato es  $a + 4$ .

**Suma:**  $M_o + M_e = (a + 6) + (a + 4) = 2a + 10$ .

**Respuesta:**  $2a + 10$ .

Los datos de una muestra son todos números naturales consecutivos, si no hay ningún dato repetido y la mediana de la muestra es 11.5, entonces ¿qué cantidad de datos no puede tener la muestra? *Solución:* Si la mediana es 11.5 (un número no entero), la cantidad de datos ( $n$ ) debe ser par. Si  $n$  fuera

impar, la mediana sería el dato central, que es un número natural, lo cual contradice el enunciado. Por lo tanto, la cantidad de datos  $n$  no puede ser un número impar.

**Respuesta:** La cantidad de datos no puede ser un número impar.

## 4. 04/04

### 4.1. Población y Muestra

¿Qué inconvenientes puede implicar realizar un censo?

- **Cardinalidad (tamaño):** La población puede ser demasiado grande o infinita.
- **Destrucción:** El proceso de medición puede destruir el elemento (ej. pruebas de vida útil).
- **Costos:** Implica altos costos en tiempo, dinero y recursos.
- **Acceso:** Puede ser logísticamente imposible acceder a todos los miembros.
- **Tiempo:** Puede tomar tanto tiempo que la información obtenida se vuelve obsoleta.

### 4.2. Muestreo

Proceso para escoger los elementos que conformarán la muestra.

**Es fundamental que la muestra sea representativa para realizar una inferencia estadística válida.**

#### 4.2.1. Representatividad

Debe reflejar las características de la población. Claves:

- **Tamaño ( $n$ ):** Debe ser suficientemente grande.
- **Aleatoriedad:** Cada elemento debe tener una probabilidad conocida de ser seleccionado para minimizar el sesgo.

Notación:  $N$  para el tamaño de la población y  $n$  para el tamaño de la muestra.

### 4.3. Tipos de Muestreo Probabilístico

Cada unidad tiene una probabilidad conocida y no nula de ser seleccionada.

#### **4.3.1. Muestreo Aleatorio Simple (M.A.S)**

Cada posible muestra de tamaño  $n$  tiene la misma probabilidad de ser elegida. Requiere un listado completo (marco muestral).

#### **4.3.2. Muestreo Estratificado**

Se usa cuando la población se puede dividir en subgrupos (estratos) internamente homogéneos. Se toma una muestra aleatoria de cada estrato. Es más preciso si hay baja variabilidad dentro de los estratos y alta variabilidad entre ellos.

#### **4.3.3. Muestreo por Conglomerados (Clusters)**

Se usa cuando la población está dividida en grupos naturales (conglomerados). Se selecciona una muestra aleatoria de conglomerados y se analizan **todos** los individuos de los conglomerados elegidos. Es eficiente en costos, especialmente si los conglomerados son internamente heterogéneos.

#### **4.3.4. Muestreo Aleatorio Sistemático**

Se elige un elemento al azar al principio y luego se seleccionan los demás a intervalos regulares ( $k = N/n$ ) de una lista ordenada.

### **4.4. Muestreos No Probabilísticos**

La selección es subjetiva y no permite inferencia estadística formal.

#### **4.4.1. Muestreo por Cuotas**

Se fijan cuotas de individuos con ciertas características y el entrevistador los selecciona por conveniencia hasta llenar la cuota.

#### **4.4.2. Muestreo Bola de Nieve**

Los primeros individuos contactados ayudan a encontrar a otros, útil para poblaciones difíciles de localizar.

#### 4.4.3. Muestreo por Juicio o Conveniencia

La selección se basa en el juicio del investigador o en la facilidad de acceso a los sujetos.



Figura 2: Patata para rellenar el espacio

## 5. 10/04

*Objetivo: Aplicar y comprender propiedades de las medidas de dispersión*

### 5.1. Medidas de Dispersión

Las medidas de tendencia central (como la media) no son suficientes por sí solas para describir un conjunto de datos, ya que no indican cuán dispersos o concentrados están los datos alrededor de ese centro. Consideremos dos conjuntos con la misma media  $\bar{x} = 0$ :

$$A = \{-4, 4, -4, 4\} \quad (\text{Media } \bar{x}_A = 0)$$

$$B = \{7, 1, -6, -2\} \quad (\text{Media } \bar{x}_B = 0)$$

Ambos tienen  $\bar{x} = 0$ , pero los datos en el conjunto  $A$  están menos dispersos (más concentrados alrededor de la media) que en el conjunto  $B$ . Las medidas de dispersión cuantifican esta variabilidad o "esparcimiento" de los datos.

#### 5.1.1. Rango (o Amplitud Total):

Se define como la diferencia entre el valor máximo y el valor mínimo de los datos.

$$Rango = x_{max} - x_{min}$$

Es una medida simple pero muy sensible a valores extremos y no considera la distribución de los datos intermedios.

#### 5.1.2. Desviación Media (DM):

Dada una variable  $X$ , con  $n$  datos  $x_1, x_2, \dots, x_n$  y media aritmética  $\bar{x}$ . Se define la desviación media como el promedio de las desviaciones absolutas de cada dato respecto a la media:

$$DM = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

Mide el promedio de cuánto se desvían los datos de la media, en valor absoluto.



### 5.1.3. Varianza ( $\sigma^2$ para población, $s^2$ para muestra):

Es el promedio de las desviaciones al cuadrado de cada dato respecto a la media. Es la medida de dispersión más utilizada junto con su raíz cuadrada (la desviación estándar). Para una **población** de  $N$  datos:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

Donde  $\mu$  es la media poblacional. Si los datos  $x_1, \dots, x_n$  constituyen toda la población (y  $\bar{x}$  es su media):

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Para una **muestra** de  $n$  datos, la varianza muestral *insesgada* (estimador de  $\sigma^2$ ) es:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

(En este curso, si no se especifica,  $\sigma^2$  con denominador  $n$  se refiere a la varianza de un conjunto de datos específico, sea este una población o una muestra descripta como tal).

### 5.1.4. Desviación Estándar (o Típica) ( $\sigma$ para población, $s$ para muestra):

Es la raíz cuadrada positiva de la varianza. Tiene la ventaja de estar expresada en las mismas unidades que los datos originales.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

### 5.1.5. Propiedades de $\sigma$ y $\sigma^2$ (usando la definición con denominador $n$ )

1.  $\sigma \geq 0$  y  $\sigma^2 \geq 0$ . Son siempre no negativas.

2.  $\sigma = 0 \iff \sigma^2 = 0 \iff x_i = \bar{x}$  para todo  $i \iff x_i = x_j$  para todo  $i, j \in \{1, \dots, n\}$ . La desviación estándar (y varianza) es cero si y sólo si todos los datos son iguales.
3. Si a todos los datos de un conjunto se les suma (o resta) una constante  $k$  (transformación  $y_i = x_i + k$ ), la nueva media es  $\bar{y} = \bar{x} + k$ , pero la varianza y la desviación estándar no cambian:  $\sigma_y^2 = \sigma_x^2$  y  $\sigma_y = \sigma_x$ .
4. Si todos los datos de un conjunto se multiplican (o dividen) por una constante  $k$  (transformación  $y_i = k \cdot x_i$ ), la nueva media es  $\bar{y} = k\bar{x}$ , la nueva varianza es  $\sigma_y^2 = k^2\sigma_x^2$ , y la nueva desviación estándar es  $\sigma_y = |k|\sigma_x$ .
5. Fórmula computacional (o abreviada) para la varianza:  $\sigma^2 = \frac{\sum x_i^2}{n} - (\bar{x})^2 = \overline{x^2} - (\bar{x})^2$ . Es decir, la varianza es la media de los cuadrados de los datos menos el cuadrado de la media de los datos.
6.  $\sigma^2 = \sigma \iff \sigma = 0 \vee \sigma = 1$ . (Asumiendo que  $\sigma$  es el valor numérico de la desviación estándar).
7.  $\sigma^2 < \sigma \iff 0 < \sigma < 1$ .
8.  $\sigma^2 > \sigma \iff \sigma > 1$ .

## 6. 16/04

### 6.1. Demostración Propiedad 4 (Multiplicación por una constante)

Sea la variable  $X$  con datos  $x_1, \dots, x_n$ , media  $\bar{x}$  y varianza  $\sigma_x^2$ . Sea  $Y$  una nueva variable tal que  $y_i = k \cdot x_i$  para cada  $i$ . Sabemos que la media de  $Y$  es  $\bar{y} = k \cdot \bar{x}$ . La varianza de  $Y$ ,  $\sigma_y^2$ , se define como:

$$\sigma_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$$

Sustituyendo  $y_i = kx_i$  y  $\bar{y} = k\bar{x}$ :

$$\sigma_y^2 = \frac{\sum_{i=1}^n (kx_i - k\bar{x})^2}{n}$$

Factorizando  $k$  dentro del paréntesis al cuadrado:

$$\sigma_y^2 = \frac{\sum_{i=1}^n [k(x_i - \bar{x})]^2}{n}$$

Aplicando la potencia al producto:

$$\sigma_y^2 = \frac{\sum_{i=1}^n k^2 (x_i - \bar{x})^2}{n}$$

Como  $k^2$  es una constante para la sumatoria, puede salir fuera:

$$\sigma_y^2 = k^2 \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Reconociendo que  $\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$  es la definición de  $\sigma_x^2$ :

$$\sigma_y^2 = k^2 \cdot \sigma_x^2$$

Tomando la raíz cuadrada positiva para obtener la desviación estándar (ya que  $\sigma_x \geq 0$ ):

$$\sigma_y = \sqrt{k^2 \cdot \sigma_x^2} = \sqrt{k^2} \cdot \sqrt{\sigma_x^2} = |k| \cdot \sigma_x$$

L.Q.Q.D. (Lo Que Queríamos Demostrar)

## 6.2. Ejercicio

Dados los datos: -2, 0, 2, 4, 6. ( $n = 5$ ). Determinar:

1.  $\bar{x}$

*Solución:*  $\bar{x} = \frac{-2+0+2+4+6}{5} = \frac{10}{5} = 2.$

2.  $\sigma$  (desviación estándar)

*Solución:* Primero calculamos la varianza  $\sigma^2$ :

$$\begin{aligned}\sigma^2 &= \frac{\sum (x_i - \bar{x})^2}{n} \\ &= \frac{(-2 - 2)^2 + (0 - 2)^2 + (2 - 2)^2 + (4 - 2)^2 + (6 - 2)^2}{5} \\ &= \frac{(-4)^2 + (-2)^2 + (0)^2 + (2)^2 + (4)^2}{5} \\ &= \frac{16 + 4 + 0 + 4 + 16}{5} = \frac{40}{5} = 8\end{aligned}$$

Ahora la desviación estándar:  $\sigma = \sqrt{\sigma^2} = \sqrt{8} = \sqrt{4 \cdot 2} = 2\sqrt{2} \approx 2,828.$

3.  $\overline{x^2}$  (el promedio de los cuadrados de los datos)

*Solución:* Los cuadrados de los datos son:  $(-2)^2 = 4, 0^2 = 0, 2^2 = 4, 4^2 = 16, 6^2 = 36.$

$$\overline{x^2} = \frac{4 + 0 + 4 + 16 + 36}{5} = \frac{60}{5} = 12$$

4. Calcular  $\overline{x^2} - (\bar{x})^2$  y comparar con  $\sigma^2$ .

*Solución:*  $\overline{x^2} - (\bar{x})^2 = 12 - (2)^2 = 12 - 4 = 8.$

Este resultado (8) es igual a la varianza  $\sigma^2$  calculada en el punto 2, lo cual verifica la propiedad 5 (fórmula computacional de la varianza).

### 6.3. Demostración Propiedad 5 (Fórmula Computacional de la Varianza)

Partimos de la definición de varianza (usando denominador  $n$ ):

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Expandimos el binomio al cuadrado  $(a - b)^2 = a^2 - 2ab + b^2$ :

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + (\bar{x})^2)}{n}$$

Distribuimos la sumatoria y el denominador  $n$ :

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

En el segundo término,  $2\bar{x}$  es una constante respecto a la suma  $\sum x_i$ . En el tercer término,  $(\bar{x})^2$  es una constante, y  $\sum_{i=1}^n (\bar{x})^2 = n(\bar{x})^2$ .

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

Reconocemos que  $\frac{\sum x_i^2}{n} = \overline{x^2}$  (la media de los cuadrados) y  $\frac{\sum x_i}{n} = \bar{x}$  (la media):

$$\sigma^2 = \overline{x^2} - 2\bar{x}(\bar{x}) + (\bar{x})^2$$

$$\sigma^2 = \overline{x^2} - 2(\bar{x})^2 + (\bar{x})^2$$

$$\sigma^2 = \overline{x^2} - (\bar{x})^2$$

L.Q.Q.D.

## 7. 23/04

### 7.1. Demostraciones de Propiedades Relacionadas con el Valor de $\sigma$

Recordar que  $\sigma \geq 0$  por definición (es una raíz cuadrada positiva o cero).

#### 7.1.1. Propiedad 6: $\sigma^2 = \sigma \iff \sigma = 0 \vee \sigma = 1$

Partimos de la ecuación:

$$\sigma^2 = \sigma$$

Reordenamos para formar una ecuación cuadrática en  $\sigma$ :

$$\sigma^2 - \sigma = 0$$

Factorizamos  $\sigma$ :

$$\sigma(\sigma - 1) = 0$$

Esto implica que uno de los factores debe ser cero:

$$\sigma = 0 \quad \text{o} \quad \sigma - 1 = 0$$

Por lo tanto:

$$\sigma = 0 \vee \sigma = 1$$

L.Q.Q.D.

#### 7.1.2. Propiedad 7: $\sigma^2 < \sigma \iff 0 < \sigma < 1$

Partimos de la desigualdad:

$$\sigma^2 < \sigma$$

Reordenamos:

$$\sigma^2 - \sigma < 0$$

Factorizamos:

$$\sigma(\sigma - 1) < 0$$

Para que el producto de dos factores sea negativo, uno debe ser positivo y el otro negativo. Analizamos los signos de  $\sigma$  y  $(\sigma - 1)$ :

Intervalo	$(-\infty, 0)$	0	$(0, 1)$	1	$(1, +\infty)$
Signo de $\sigma$	-	0	+	+	+
Signo de $(\sigma - 1)$	-	-	-	0	+
Signo de $\sigma(\sigma - 1)$	+	0	-	0	+

La desigualdad  $\sigma(\sigma - 1) < 0$  se cumple cuando  $\sigma \in (0, 1)$ . Dado que  $\sigma \geq 0$  por definición, el intervalo  $(-\infty, 0)$  no es relevante para la desviación estándar. Por lo tanto:

$$0 < \sigma < 1$$

L.Q.Q.D.

### 7.1.3. Propiedad 8: $\sigma^2 > \sigma \iff \sigma > 1$

Partimos de la desigualdad:

$$\sigma^2 > \sigma$$

Reordenamos:

$$\sigma^2 - \sigma > 0$$

Factorizamos:

$$\sigma(\sigma - 1) > 0$$

Para que el producto de dos factores sea positivo, ambos deben ser positivos o ambos deben ser negativos. Usando la tabla de signos anterior:

- Ambos negativos:  $\sigma < 0$  y  $\sigma - 1 < 0$  (es decir,  $\sigma < 0$ ). No es posible para  $\sigma$ .
- Ambos positivos:  $\sigma > 0$  y  $\sigma - 1 > 0$  (es decir,  $\sigma > 1$ ).

La desigualdad  $\sigma(\sigma - 1) > 0$  se cumple cuando  $\sigma \in (-\infty, 0) \cup (1, +\infty)$ . Considerando la restricción  $\sigma \geq 0$ :

- Si  $\sigma = 0$ , entonces  $\sigma(\sigma - 1) = 0$ , lo cual no satisface  $0 > 0$ .
- El intervalo  $(-\infty, 0)$  no es válido para  $\sigma$ .
- Nos queda el intervalo  $(1, +\infty)$ .

Por lo tanto:

$$\sigma > 1$$

L.Q.Q.D.

## 8. 30/05

Dado un Conjunto de datos Ordenados se define:

### 8.1. Cuartiles

Son tres valores que dividen en parte iguales a un conjunto de datos ordenados. Se denominan  $Q_1$ ,  $Q_2$  y  $Q_3$  **Observacion:**  $Q_2 = Me$

### 8.2. Quintiles

Son los 4 valores que dividen en partes iguales a los datos, que corresponden al 20 %, 40 %, 60 % y 80 %

### 8.3. Deciles

Son los 9 valores que dividen en partes iguales a los datos, que corresponden al 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 % y 90 %

### 8.4. Percentiles

Son los 99 valores que dividen en partes iguales a los datos, que corresponden al 1 %, 2 %, 3 %, ... ,99 %

$$P_i = \frac{i \cdot n}{100}$$

Corresponde a la posición del percentil.



## 9. 11/06

### 9.1. Combinatoria

Son técnicas de conteo. Sean A y B dos sucesos, que pueden ocurrir de  $a$  y  $b$  maneras respectivamente.

- **Principio Aditivo:** Si los sucesos no pueden ocurrir de manera simultánea (son mutuamente excluyentes), entonces hay  $a+b$  maneras posibles de que ocurra A o B.
- **Principio Multiplicativo:** Si los sucesos ocurren uno a continuación del otro o de manera simultánea, entonces hay  $a \cdot b$  formas de que ocurra A y B.

### 9.2. Ejercicios

1. Al lanzar una moneda y un dado, ¿cuántos resultados posibles hay?  
*Solución:* Lanzar una moneda y un dado son sucesos independientes que ocurren simultáneamente. Usamos el principio multiplicativo.

- Resultados de la moneda: 2 (cara, sello).
- Resultados del dado: 6 (1, 2, 3, 4, 5, 6).

Total de resultados:  $2 \times 6 = 12$ .

**Respuesta: 12 resultados posibles.**

2. Si Pedro tiene 5 lápices de pasta, 4 de tinta y 3 de grafito, ¿de cuántas maneras puede elegir un lápiz para escribir?

*Solución:* Pedro debe elegir un lápiz de pasta o uno de tinta o uno de grafito. Son elecciones mutuamente excluyentes. Usamos el principio aditivo.

- Maneras de elegir un lápiz de pasta: 5.
- Maneras de elegir un lápiz de tinta: 4.
- Maneras de elegir un lápiz de grafito: 3.

Total de maneras:  $5 + 4 + 3 = 12$ .

**Respuesta: 12 maneras distintas.**

3. En un local se puede elegir un combo con las siguientes opciones disponibles: 5 tipos de hamburguesas, 4 bebidas distintas o bien un jugo de 2 sabores distintos. Si el combo consiste en una hamburguesa y una bebida/jugo, ¿de cuántas maneras distintas se puede elegir?

*Solución:* La elección del combo tiene dos etapas: elegir una hamburguesa y elegir un líquido.

- **Etapas 1 (Hamburguesa):** Hay 5 opciones.
- **Etapas 2 (Líquido):** Se puede elegir una bebida o un jugo. Esta es una sub-decisión que usa el principio aditivo.
  - Opciones de bebida: 4.
  - Opciones de jugo: 2.
  - Total de opciones de líquido:  $4 + 2 = 6$ .

Para formar el combo, se aplica el principio multiplicativo entre la Etapa 1 y la Etapa 2. Total de combos:  $5 \text{ (hamburguesas)} \times 6 \text{ (líquidos)} = 30$ .

**Respuesta:** Se puede elegir de 30 maneras distintas.

## 10. 07/07

### *Diferenciar distintas Permutaciones*

#### 10.1. Combinatoria

Evento A: a maneras

Evento B: b maneras

- Principio Aditivo:  $a + b$  **Excluyentes**
- Principio Multiplicativo:  $a \cdot b$  **Simultaneos**
- *Ej:*
  - Evento A: Lanzar un dado
  - Evento B: lanzar un dado 2 veces

##### 10.1.1. Permutacion simple

Corresponde a la manera en que se pueden ordenar  $n$  objetos

$$P_n = n!$$

$$P_4 = 4! = 4 \cdot 3 \cdot 2 \cdot 1$$

##### 10.1.2. Permutacion con repetición

Si tienes  $n$  elementos y un elemento se repite  $r_1$  veces, otro  $r_2$  veces y así sucesivamente, entonces se pueden ordenar:

$$P_{r_1, r_2, \dots}^n = \frac{n!}{r_1! \cdot r_2! \cdot \dots}$$

##### 10.1.3. Permutacion Circular

Si tenemos  $n$  elementos, los podemos ordenar de :

$$PC_n = \frac{n!}{n} = (n - 1)!$$

## 10.2. Ejercicios

- a.) ¿De cuantas maneras se pueden ordenar las letras de la palabra LEMUR para formar otra palabra (con o sin sentido)

$$P_5 = 5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

- b.) ¿Y para la palabra ARROYO? ¿Y CARRERA?

**ARROYO:**

**CARRERA**

$$P_{2,2}^6 = \frac{6!}{2! \cdot 2!} = \frac{720}{4} = 180$$

$$P_{2,3}^7 = \frac{7!}{2! \cdot 3!} = \frac{7 \cdot 6 \cdot 5 \cdot 4}{2} = 420$$

- c.) ¿De cuantas maneras se pueden sentar en una mesa de 4 sillas, 4 personas?

$$PC_4 = \frac{4!}{4} = \frac{4 \cdot 3!}{4} = 3! = 6$$