



# Universidad Nacional Autónoma de México

Facultad de Estudios Superiores Acatlán

## Apuntes Estadística

**Autor:**

Jorge Miguel Alvarado Reyes

29 de enero de 2024

# Índice

<b>1. 16 de agosto 2023</b>	<b>2</b>
1.1. Medidas de Tendencia Central . . . . .	2
1.1.1. Media . . . . .	2
1.1.2. Mediana . . . . .	2
1.1.3. Moda . . . . .	3
1.2. Media para una serie de frecuencias . . . . .	3
1.3. Media para datos agrupados . . . . .	3
<b>2. 18 Agosto 2023</b>	<b>5</b>
2.1. Breve introduccion a latex . . . . .	5
2.1.1. Principales clases de documentos . . . . .	5
2.1.2. Paquetes . . . . .	5
2.1.3. Estructura de un documento . . . . .	5
2.1.4. LaTeX en linea . . . . .	6
2.1.5. Partes de un documento . . . . .	6
2.1.6. Tamaños de fuente . . . . .	6
2.1.7. Listas numeradas y viñetas . . . . .	6
2.1.8. Alineacion de texto . . . . .	6
2.1.9. Composicion de ecuaciones . . . . .	6
2.1.10. Alinear expresion con algun elemento . . . . .	7
2.1.11. Tablas . . . . .	7
2.1.12. Como insertar una imagen . . . . .	7
2.2. Clase . . . . .	8
<b>3. 21 Agosto 2023</b>	<b>8</b>
3.1. Medidas de posición (1.6.4) . . . . .	8
3.1.1. Cuantiles . . . . .	8
3.1.2. Varianza . . . . .	11
3.1.3. Desviacion estandar . . . . .	12
3.1.4. Desviacion media . . . . .	12
3.1.5. Rango . . . . .	13
3.2. Medidas de forma (1.6.5) . . . . .	13
3.2.1. Asimetria . . . . .	13
<b>4. 23 agosto 2023</b>	<b>14</b>
4.1. Medidas de forma (1.6.5) (continuacion) . . . . .	14
4.1.1. Coeficiente de asimetria . . . . .	14
4.1.2. Curtosis . . . . .	15
4.1.3. Coeficiente de Fisher . . . . .	15
4.2. Medidas de asociacion (1.6.6) . . . . .	16
4.2.1. Covarianza . . . . .	16
4.2.2. Coeficiente de correlacion de Pearson . . . . .	17
4.3. Presentacion grafica y tabular de los datos (1.6.7) . . . . .	17

<b>5. 25 agosto 2023</b>	<b>18</b>
5.1. Graficas de barras . . . . .	18
5.2. Histograma . . . . .	18
5.3. Gráficos de dispersión . . . . .	19
5.4. Gráfica de burbujas . . . . .	19
5.5. Serie temporal . . . . .	19
5.6. Introducción a R . . . . .	20
5.7. Reglas de sintaxis . . . . .	20
<b>6. 28 agosto 2023</b>	<b>21</b>
6.1. Diagrama de caja y brazos . . . . .	21
6.2. Continuación de R . . . . .	21
6.2.1. Operadores aritméticos . . . . .	21
6.2.2. Operadores de asignacion . . . . .	21
6.2.3. Operadores de comparacion e identidad . . . . .	21
6.2.4. Funciones para cadenas de texto . . . . .	22
6.2.5. Listas . . . . .	22
6.2.6. Matrices . . . . .	22
6.2.7. Convertir tipo de datos . . . . .	22
6.3. Funciones con dataFrames . . . . .	22
<b>7. 30 Agosto 2023</b>	<b>23</b>
7.1. Funciones . . . . .	23
7.2. Unidad 2 . . . . .	23
<b>8. 1 Septiembre 2023</b>	<b>25</b>
8.1. Métodos de las funciones de distribución . . . . .	25
8.2. Teorema de cambio de variable . . . . .	27
<b>9. 08 de Septiembre del 2023</b>	<b>29</b>
9.1. Método de la función generadora de momentos . . . . .	29
<b>10.13 de Septiembre del 2023</b>	<b>30</b>
<b>11.18 de Septiembre del 2023</b>	<b>30</b>
11.1. Distribuciones derivadas de la normal . . . . .	31
11.1.1. Normal estándar . . . . .	31
11.1.2. Ji cuadrada . . . . .	31
<b>12.Unidad 3</b>	<b>31</b>
12.1. Introducción . . . . .	31
<b>13.22 de Septiembre del 2023</b>	<b>32</b>
13.1. Distribución de muestreo de $\bar{X}$ . . . . .	32
13.2. Teorema central del limite . . . . .	33

<b>14.25 de Septiembre del 2023</b>	<b>34</b>
14.1. La distribución de muestreo de $s^2$ . . . . .	34
14.2. Aproximación normal . . . . .	35
<b>15.27 de Septiembre del 2023</b>	<b>35</b>
15.1. Estimacion puntual y por intervalos . . . . .	35
15.2. Propiedades de los estimadores . . . . .	35
<b>16.2 de Octubre del 2023</b>	<b>36</b>
16.1. Métodos para obtener estimadores puntuales . . . . .	36
16.1.1. Método de momentos . . . . .	36
<b>17.4 de Octubre del 2023</b>	<b>37</b>
<b>18.11 de Octubre del 2023</b>	<b>41</b>
<b>19.13 de Octubre del 2023</b>	<b>43</b>
<b>20.16 de octubre del 2023</b>	<b>44</b>
<b>21.20 de octubre del 2023</b>	<b>50</b>
<b>22.23 de Octubre del 2023</b>	<b>51</b>
<b>23.6 de Noviembre del 2023</b>	<b>52</b>
<b>24.10 de Noviembre del 2023</b>	<b>54</b>
24.1. Muestreo aleatorio estratificado . . . . .	54
<b>25.hh</b>	<b>56</b>
<b>26.ss</b>	<b>57</b>
<b>27.Reapso 22 de noviembre</b>	<b>63</b>
<b>28.coursera</b>	<b>63</b>

## 1. 16 de agosto 2023

Las medidas de tendencia central son valores de un conjunto de datos que se encuentran en el centro de los datos ordenados.

### 1.1. Medidas de Tendencia Central

#### 1.1.1. Media

Existen dos tipos de media: la aritmética y la ponderada.

La **media aritmética** se calcula sumando todos los valores y dividiendo por la cantidad de valores:

$$\text{Media}(x) = \frac{\sum_{i=1}^n x_i}{n} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

**Propiedades:**

1.  $\text{Media}(cx) = c \cdot \text{Media}(x)$
2.  $\text{Media}(x + c) = \text{Media}(x) + c$

**Ejemplo 4:**

Mostrar que  $\text{Media}(x + c) = \text{Media}(x) + c$ .

**Demostración:**

$$\begin{aligned}\text{Media}(x + c) &= \frac{x_1 + c + \dots + x_n + c}{n} \\ &= \frac{x_1 + \dots + x_n + n \cdot c}{n} \\ &= \frac{x_1 + \dots + x_n}{n} + c \\ &= \text{Media}(x) + c\end{aligned}$$

**Ejemplo 5:**

Mostrar que  $\text{Media}(cx) = c \cdot \text{Media}(x)$ .

La **media ponderada** se define como:

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

#### 1.1.2. Mediana

La mediana es el valor central que divide a un conjunto de datos ordenados en dos partes iguales. Si  $n$  es par, se calcula como:

$$\text{Mediana}(x) = \frac{x(\frac{n}{2}) + x(\frac{n}{2} + 1)}{2}$$

### 1.1.3. Moda

Es el valor que mas se repite en un conjunto de observaciones.

**Ejemplo 6:**

1.  $[1, 2, 3, 4, 5]$  Aqui no existe moda
2.  $[3, 4, 4, 5, 5, 6]$  Moda = 4.5
3.  $[3, 3, 4, 5, 6, 6]$  Moda = 3 y 6
4.  $[2, 7, 7, 7, 9]$  Moda = 7

### 1.2. Media para una serie de frecuencias

Si  $f_1, \dots, f_n$  son las frecuencias de la variable  $x$ . Entonces.

$$Mediana(x) = \frac{\sum_{i=1}^n f_i \cdot x_i}{\sum_{i=1}^n f_i}$$

**Ejemplo 7:**

Calcula la media para los siguientes valores

$x$	$f_i$
2	4
5	1
6	3
8	4

$$Mediana(x) = \frac{(2 \cdot 4) + (5 \cdot 1) + (6 \cdot 3) + (8 \cdot 4)}{4 + 1 + 3 + 4} = \frac{8 + 5 + 18 + 32}{12} = \frac{63}{12}$$

### 1.3. Media para datos agrupados

Sean  $f_1, \dots, f_n$  las frecuencias de la variable  $x$  y  $c_1, \dots, c_n$  las marcas de clase, entonces: (Marca de clase es un representante)

$$Media(x) = \frac{\sum_{i=1}^n f_i \cdot c_i}{\sum_{i=1}^n f_i}$$

**Ejemplo 8:**

Calcula la edad promedio para el siguiente conjunto de datos

Adulto	25
Adulto de la tercera edad	10

Adulto, edad =  $[20, 65]$ ,  $c_1 = 43$   
Adulto tercera edad, edad =  $[65, 100]$ ,  $c_1 = 83$

25 veces 43 y 10 veces 83

$$Mediana(x) = L_i t \left( \frac{\frac{n}{2} - \sum f_i}{f_{mediana}} \right) \cdot c$$

$L_i$  = limite inferior de la clase que contiene la mediana

$n$  = frecuencia total

$\sum f_i$  = suma de las frecuencias menores a la mediana

$f_{mediana}$  = Frecuencia de la clase que contiene la mediana

$c$  = longitud del intervalo que contiene la mediana

## 2. 18 Agosto 2023

Sitio del curso: <https://piazza.com/unam.mx/other/ei20241>  
codigo de acceso: 150621

### 2.1. Breve introduccion a latex

LaTeX es una herramienta para crear documentos de una gran calidad tipográfica, en donde los usuarios se ocupan en mayor medida del contenido del texto en lugar del formato.

#### 2.1.1. Principales clases de documentos

Clase	Proposito
article	Articulos de revista
report	Textos largos como tesis o reportes
book	Libros o documentos con una estructura similar
lette	cartas

#### 2.1.2. Paquetes

Nombre	Función
amsmath, amssymb, amsfont	Permiten el uso de símbolos matemáticos.
babel	Escribir en diversos idiomas.
inputec	Codificación de entradas.

#### 2.1.3. Estructura de un documento

```
\documentclass[11pt, a4paper]{report}
\usepackage[utf8]{inputec}
\usepackage[spanish]{babel}
\usepackage{amsmath}
\usepackage{amssymb}
\usepackage{amsfont}

\title{Titulo}
\author{Nombre}
\date{\today}
\begin{document}
\maketitle
...
\end{document}
```



#### 2.1.4. LaTeX en línea

Crear cuenta en <https://es.overleaf.com>

New project → Blank project → Escribir nombre del documento → Create  
Menu → spell check spanish

#### 2.1.5. Partes de un documento

```
\section*{title}
\subsection*{title}
\subsubsection*{title}

\part*{title}
\chapter*{title}
```

#### 2.1.6. Tamaños de fuente

```
\huge
\Huge
\LARGE
\Large
\large
\normalsize
\small
\tiny
```

#### 2.1.7. Listas numeradas y viñetas

```
\begin{itemize}[a]
  \item
  \item
\end{itemize}

\begin{enumerate}
  \item
  \item
\end{itemize}
```

#### 2.1.8. Alineacion de texto

```
\begin{center}
...
\end{center}
```

#### 2.1.9. Composicion de ecuaciones

```

$$x^2+2x+3=0$$

```

#### 2.1.10. Alinear expresion con algun elemento

```
\begin{align*}
c^2 &= a^2 + b^2 \\
&= 2^2 + 3^2 \\
&= 13
\end{align*}
```

#### 2.1.11. Tablas

```
\begin{table}[h]
\centering
\begin{tabular}{c | c c}
a & b & c \\
a & b & c \\
a & b & c
\end{tabular}
\end{table}
```

```
\begin{table}[h]
\centering
\begin{tabular}{| c c c |}
\hline
a & b & c \\
\hline
a & b & c \\
\hline
a & b & c \\
\hline
\end{tabular}
\end{table}
```

#### 2.1.12. Como insertar una imagen

```
\usepackage{graphicx}
\includegraphics[width = , height = ]{archivo.jpg,png,etc.}
```

## 2.2. Clase

### Ejemplo 9:

Encuentran la mediana para las siguientes observaciones

Intervalo	Frecuencia	Frecuencia acumulada
(118.5,126.5]	3	3
(126.5,135.5]	5	8
(135.5,144.5]	9	17
(144.5,153.5]	12	29
(153.5,162.5]	5	34
(162.5,171.5]	4	38
(171.5, 180.5]	2	40

$L_i$  = limite inferior de la clase que contiene la mediana

$n$  = frecuencia total

$\sum f_i$  = suma de las frecuencias menores a la mediana

$f_{mediana}$  = Frecuencia de la clase que contiene la mediana

$c$  = longitud del intervalo que contiene la mediana

$$L_i = 144.5$$

$$n = 40$$

$$\sum f_i = 17$$

$$f_{mediana} = 12$$

$$c = 153.5 - 144.5 = 9$$

$$\begin{aligned} Mediana(x) &= \frac{x_{20} + x_{21}}{2} = 146,75 \\ &= L_i + \left( \frac{\frac{n}{2} - \sum f_i}{f_{mediana}} \right) \cdot c \end{aligned}$$

## 3. 21 Agosto 2023

### 3.1. Medidas de posición (1.6.4)

#### 3.1.1. Cuantiles

Sean  $x_1, \dots, x_n$  observaciones de una variable aleatoria  $x$  y  $p \in (0, 1)$ . Un cuantil al 100 $p$ % es el numero  $c$  que cumple con las siguientes condiciones.

$$\begin{aligned} \blacksquare \quad \frac{\#\{x_i | x_i \leq c\}}{n} &\geq p \\ \blacksquare \quad \frac{\#\{x_i | x_i \leq c\}}{n} &\geq 1 - p \end{aligned}$$

Casos particulares

- Deciles: si  $p = \{0,1, \dots, 0,9\}$
- Cuartiles: si  $p = \{0,25, 0,50, 0,75\}$
- Percentiles: si  $p = \{0,01, 0,02, \dots, 0,99\}$

### Ejemplo 10

Calcula los deciles de la variable  $x = \{0,1\}$

Supongamos:

Si  $c=0$

$$\blacksquare \frac{\#\{x_i | x_i \leq 0\}}{n} \geq p$$

$$\frac{\#\{0\}}{2} \geq p$$

$$\frac{1}{2} \geq p$$

$$p = \{0.1, 0.2, 0.3, 0.4, 0.5\}$$

$$\blacksquare \frac{\#\{x_i | x_i \geq 0\}}{n} \geq 1 - p$$

$$\frac{\#\{0,1\}}{2} \geq 1 - p$$

$$1 \geq 1 - p$$

$$p = \{0.1, 0.2, 0.3, \dots, 0.8, 0.9\}$$

Si  $c=1$

$$\blacksquare \frac{\#\{x_i | x_i \leq 1\}}{n} \geq p$$

$$\frac{\#\{0,1\}}{2} \geq p$$

$$1 \geq p$$

$p = \{0.5, 0.6, 0.7, 0.8, 0.9\}$  ??? para todos no?

$$\blacksquare \frac{\#\{x_i | x_i \geq 1\}}{n} \geq 1 - p$$

$$\frac{\#\{1\}}{2} \geq p$$

$$\frac{1}{2} \geq 1 - p$$

$$p = \{0.5, 0.6, 0.7, 0.8, 0.9\}$$

- $q(0.1) = 0$
- $q(0.2) = 0$
- $q(0.3) = 0$
- $q(0.4) = 0$
- $q(0.5) = 0.5$

- $q(0.6) = 1$

- $q(0.7) = 1$

- $q(0.8) = 1$

- $q(0.9) = 1$

### Ejemplo 11

Calcula los cuartiles cd la variable  $x = \{0, 1, 2, 3\}$ ,  $n = 4$

Si  $c=0$

- $\frac{\#\{x_i | x_i \leq 0\}}{n} \geq p$

$$\frac{\#\{0\}}{4} \geq p$$

$$\frac{1}{4} \geq p$$

$$p = \{0.25\}$$

- $\frac{\#\{x_i | x_i \geq 0\}}{n} \geq 1 - p$

$$\frac{\#\{0,1,2,3\}}{4} \geq 1 - p$$

$$1 \geq 1 - p$$

$$p = \{0.25, 0.50, 0.75\}$$

Si  $c=1$

- $\frac{\#\{x_i | x_i \leq 1\}}{n} \geq p$

$$\frac{\#\{0,1\}}{4} \geq p$$

$$\frac{1}{2} \geq p$$

$$p = \{0.25, 0.5\}$$

- $\frac{\#\{x_i | x_i \geq 1\}}{n} \geq 1 - p$

$$\frac{\#\{1,2,3\}}{4} \geq 1 - p$$

$$\frac{3}{4} \geq 1 - p$$

$$p = \{0.75\}$$

Si  $c=2$

- $\frac{\#\{x_i | x_i \leq 2\}}{n} \geq p$

$$\frac{\#\{0,1,2\}}{4} \geq p$$

$$\frac{3}{4} \geq p$$

$$p = \{0.75\}$$

$$\begin{aligned}
& \blacksquare \frac{\#\{x_i | x_i \geq 2\}}{n} \geq 1 - p \\
& \frac{\#\{2,3\}}{4} \geq 1 - p \\
& \frac{1}{2} \geq 1 - p \\
& p = \{0.50, 0.75\}
\end{aligned}$$

Si  $c=3$

$$\begin{aligned}
& \blacksquare \frac{\#\{x_i | x_i \leq 3\}}{n} \geq p \\
& \frac{\#\{0,1,2,3\}}{4} \geq p \\
& 1 \geq p \\
& p = \{0.25, 0.5, 0.75\} \\
& \blacksquare \frac{\#\{x_i | x_i \geq 3\}}{n} \geq 1 - p \\
& \frac{\#\{3\}}{4} \geq 1 - p \\
& \frac{1}{4} \geq 1 - p \\
& p = \{0.75\}
\end{aligned}$$

- $q(0.25) = 0.5$
- $q(0.50) = 1.5$
- $q(0.75) = 2.5$

### 3.1.2. Varianza

Sean  $x_1, \dots, x_n$  observaciones de la variable aleatoria  $x$ . La varianza de  $x$  es

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

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

Propiedades de la varianza

$$var(x + c) = var(x)$$

$$var(ax) = a^2 var(x)$$

**Ejemplo 12**

Muestra que  $var(ax) = a^2 var(x)$

Demostracion:

$$\begin{aligned}
 var(ax) &= \frac{1}{n} \sum_{i=1}^n (ax_i - a\bar{x})^2 \\
 &= \frac{1}{n} \sum_{i=1}^n (a(x_i - \bar{x}))^2 \\
 &= a^2 \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \\
 &= a^2 var(x)
 \end{aligned}$$

**3.1.3. Desviacion estandar**

Sean  $x_1, \dots, x_n$  observaciones de la variable x. la desviacion estandar de x es

$$De(x) = \sqrt{var(x)}$$

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

Propiedades

$$De(x + c) = De(x)$$

$$De(ax) = |a|De(x)$$

**3.1.4. Desviacion media**

Sean  $x_1, \dots, x_n$  observaciones de la variable x. la desviacion media de x es

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

Propiedades

$$Dm(x + c) = Dm(x)$$

$$Dm(ax) = |a|Dm(x)$$

### 3.1.5. Rango

Sean  $x_1, \dots, x_n$  observaciones de la variable  $x$ . El rango de  $x$  es

$$R(x) = x_n - x_1$$

Propiedades

$$R(x + c) = R(x)$$

$$R(ax) = |a|R(x)$$

**Ejemplo 13** (ejercicio)

Muestra que  $\text{var}(x + c) = \text{var}(x)$

### 3.2. Medidas de forma (1.6.5)

Sean  $x_1, \dots, x_n$  observaciones de la variable  $x$ . El  $k$ -ésimo momento de  $x$  es

$$m'_k(x) = \frac{1}{n} \sum_{i=1}^n x_i^k$$

Mientras que el  $k$ -ésimo momento central de  $x$  es

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

#### 3.2.1. Asimetría

El coeficiente de asimetría mide la asimetría de los datos respecto a la media.

Sean  $x_1, \dots, x_n$  observaciones de la variable  $x$ . El coeficiente de asimetría de  $x$  es

$$sk(x) = \frac{1}{s^3} \left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \right]$$

Donde  $s$  es la desviación estándar de  $x$ .



## 4. 23 agosto 2023

### 4.1. Medidas de forma (1.6.5) (continuacion)

#### 4.1.1. Coeficiente de asimetria

$$sk(x) = \frac{1}{s^3} \left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \right]$$

Donde s es la desviacion estandar de x.

$$\begin{aligned} sk(x) &= \frac{1}{s^3} [m_3] \\ &= \frac{m_3}{(m_2)^{\frac{3}{2}}} \end{aligned}$$

Donde:

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

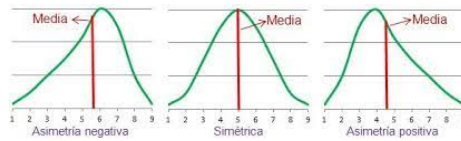


Figura 1: negativa =  $sk(x) < 0$ , Simetrica =  $sk(x) = 0$ , positiva =  $sk(x) > 0$

#### Ejemplo 14

Muestra que el coeficiente de asimetria es adimensional e invariante a cambios de escala y origen.

$$sk(ax + c) = sk(x)$$

Demostracion:

Debido a que  $m_3$  y  $(m_2)^{\frac{3}{2}}$  tienen la misma unidad de medida,  $sk(x)$  es adimensional.

$$\begin{aligned}
sk(ax + c) &= \frac{1}{(s(ax + c))^3} \left[ \frac{1}{n} \sum_{i=1}^n ((ax + c) - (a\bar{x} + c))^3 \right] \\
Media(ax + c) &= aMedia(x) + c \\
s(ax + c) &= |a|s(x) \\
&= \frac{1}{[|a|s(x)]^3} \left[ \frac{a^3}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \right] \\
&= \frac{a^3}{|a|^3} \frac{1}{s^3} \left[ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})^3 \right] \\
&= \frac{a^3}{|a|^3} sk(x)
\end{aligned}$$

#### 4.1.2. Curtosis

Sean  $x_1, \dots, x_n$  observaciones de la variable  $x$ . La curtosis de  $x$  se define como.

$$\begin{aligned}
k(x) &= \frac{1}{s^4} \left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 \right] \\
&= \frac{m_4}{(m_2)^2}
\end{aligned}$$

#### 4.1.3. Coeficiente de Fisher

$$k_3(x) = k(x) - 3$$

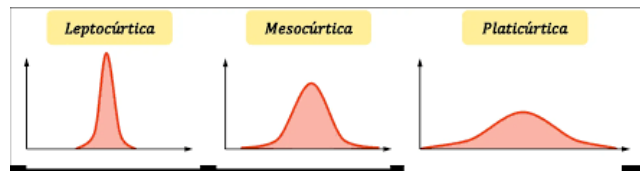
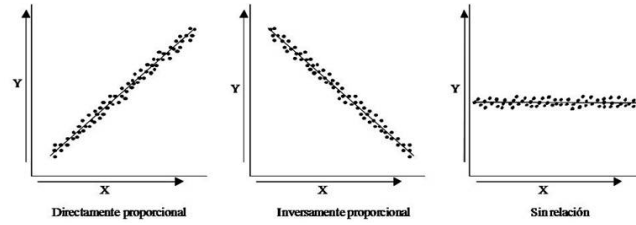


Figura 2:

## 4.2. Medidas de asociacion (1.6.6)



### 4.2.1. Covarianza

Sean  $(x_1, y_1) \dots, (x_n, y_n)$  observaciones de las variables  $(x, y)$ . La covarianza entre las variables  $(x, y)$  se define como

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

#### Ejemplo 15

Muestra que

$$s_{xy} = \frac{1}{n} \left[ \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right]$$

$$\begin{aligned} s_{xy} &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{1}{n} \sum_{i=1}^n (x_i y_i - x_i \bar{y} - y_i \bar{x} + \bar{x} \bar{y}) \\ &= \frac{1}{n} \left[ \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \bar{y} - \sum_{i=1}^n y_i \bar{x} + \sum_{i=1}^n \bar{x} \bar{y} \right] \\ &= \frac{1}{n} \left[ \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} - n \bar{x} \bar{y} + n \bar{x} \bar{y} \right] \\ &= \frac{1}{n} \left[ \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right] \end{aligned}$$

Nota:

$$n \bar{x} \bar{y} = n \frac{\sum x_i}{n} \bar{y} = \sum x_i \bar{y}$$

#### 4.2.2. Coeficiente de correlacion de Pearson

Sean  $(x_1, y_1) \dots, (x_n, y_n)$  observaciones de las variables  $(x, y)$ . El coeficiente de correlacion de Pearson entre las variables  $(x, y)$  se define como

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

Donde  $s_x$  es la desviacion estandar de  $x$  y  $s_y$  es la desviacion estandar de  $y$

Alguna consideraciones del coeficiente de correlacion son las siguientes:

$$\begin{aligned} r_{xy} &\in [-1, 1] \\ \text{si } x \perp y, r_{xy} &= 0 \\ \text{si } r_{xy} < 0, &\text{ relacion inversa entre las variables} \end{aligned}$$

#### 4.3. Presentacion grafica y tabular de los datos (1.6.7)

##### Ejemplo 16

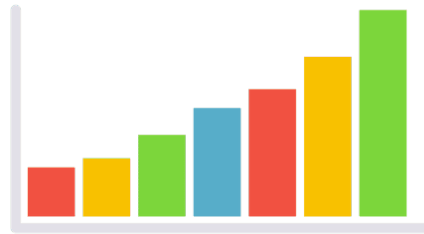
Supongamos que se tiene una variable cualitativa ordinal con valores ordenados de menor a mayor, A,B,C,D,E,F con las siguientes frecuencias,

Valor	F absoluta	F absoluta acumulada	F relativa	F relativa acumulada
A	2	2	2/28	2/28
B	8	10	8/28	10/28
C	6	16	6/28	16/28
D	4	20	4/28	20/28
E	3	23	3/28	23/28
F	5	28	5/28	28/28

## 5. 25 agosto 2023

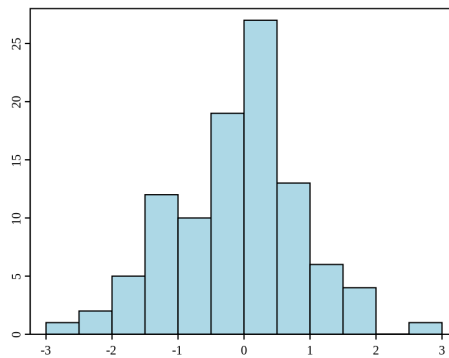
### 5.1. Graficas de barras

A cada clase de una variable se le asocia una barra de la altura la frecuencia de las observaciones. Se utiliza para cualquier tipo de variables.



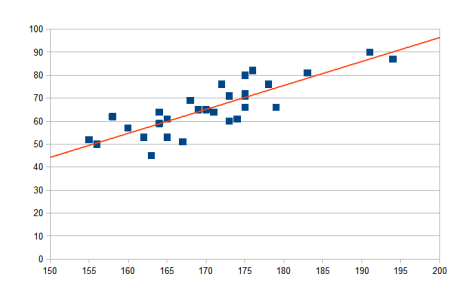
### 5.2. Histograma

Es una grafica donde los valores de la variable tiene un orden



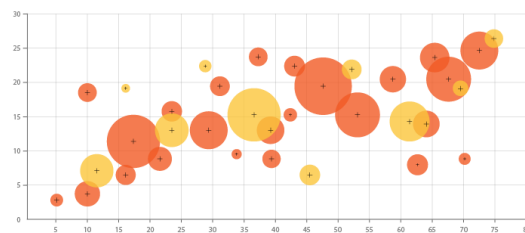
### 5.3. Gráficos de dispersión

Muestra la relacion entre dos variables



### 5.4. Gráfica de burbujas

Muestra la relación para tres variables



### 5.5. Serie temporal

Por medio de una linea se recorren diferentes valores o frecuencias a lo largo del tiempo.



## 5.6. Introducción a R

R es un programa útil para el análisis y visualización de datos. Es abierto y gratuito.

Es un lenguaje interpretado y tipado.

Fue creado por Ross Ihaka y Robert Gentleman.

Instalación.

Descargar R: <https://cran.r-project.org/>

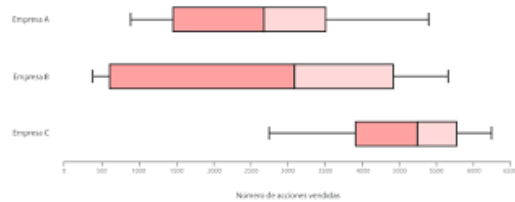
Descargar R studio: <https://posit.co/download/rstudio-desktop/>

## 5.7. Reglas de sintaxis

- R distingue entre mayúsculas y minúsculas
- Los nombres de las variables pueden contener letras, números y puntos. Sin embargo, deben comenzar con una letra y no pueden contener espacio.  
Ejemplo:  
Uso correcto  
*monto\_total* < -1200  
*monto\_mensual* < -200  
Uso incorrecto  
*montoTotal* < -1200  
*MontoMensual* < -200
- Usar espacios alrededor de todos los operadores binarios (=, +, -, < -, etc) y un espacio después de una coma.
- Ayuda, se puede usar el comando `help(mean)` o `?mean`
- Tipos de datos, enteros, numéricos y complejos.  
Ejemplo:  
*entero* < -1L  
*numerico* < -1  
*complejo* < -3 + 4i  
`print(entero, str(entero))`
- Cadena de texto  
Ejemplo:  
*mensaje* < -"Holamundo"  
`print(mensaje, str(mensaje))`
- Factores  
Ejemplo:  
*colores* < -factor(levels = c("azul", "verde"))  
`print(colores)`

## 6. 28 agosto 2023

### 6.1. Diagrama de caja y brazos



### 6.2. Continuación de R

#### 6.2.1. Operadores aritméticos

- Suma (+)
- resta (-)
- Multiplicación (\*)
- División (/)
- División entera (% / % )
- Módulo (% %)

#### 6.2.2. Operadores de asignacion

- $valor_1 < -5$
- $valor_2 = 6$
- $7 - > valor_3$

#### 6.2.3. Operadores de comparacion e identidad

- Menor
- Mayor
- Menor o igual
- Mayor o igual
- Igual
- Distinto



#### 6.2.4. Funciones para cadenas de texto

- `paste()` Concatena varias cadenas en una sola cadena
- `rep()` Repite un objeto n cantidad de veces
- `grepl()` Busca un patron en una cadena de texto y devuelve un vector logico

#### 6.2.5. Listas

- `lista j-` `list(1,"Manzana")`

#### 6.2.6. Matrices

- `matriz1 <- -matrix(1 : 10, nrow = 2, ncol = 5)`

#### 6.2.7. Convertir tipo de datos

- `as.integer`
- `as.numeric`
- `as.complex`
- `as.factor`
- `as.logical`

#### 6.3. Funciones con dataFrames

- `str()`
- `head()`
- `tail()`
- `summary()`

## 7. 30 Agosto 2023

### 7.1. Funciones

```
nombre_funcion <- function(argumentos){  
  operaciones  
  return(resultado)  
}
```

### 7.2. Unidad 2

Métodos para la obtención de funciones de variables aleatorias

#### Definición

Consideremos un fenómeno aleatorio junto con un espacio de probabilidad  $(\Omega, F, P)$ . Una variable aleatoria es una transformación  $X$  del espacio de resultados  $\Omega$  al conjunto de los reales, tal que:

$$\{w \in \Omega : X(w) \leq x\} \in F$$

#### Definición

Una variable aleatoria es discreta cuando el conjunto de valores que toma es un conjunto discreto

#### Definición

Una variable aleatoria en continua cuando el conjunto de valores que toma está en un intervalo  $(a, b) \in \mathbb{R}$

#### Definición

Sea  $X$  una variable aleatoria discreta con valores  $x_0, x_1, \dots$  y probabilidades respectivas  $\mathbb{P}(X = x_0), \mathbb{P}(X = x_1), \dots$  la función masa de probabilidad de  $x$  denotada por  $f(x) : \mathbb{R} \rightarrow [0, \infty)$  se define como sigue:

$$f(x) = \begin{cases} \mathbb{P}(X = x), & \text{si } x = x_1, x_2, \dots, \\ 0, & \text{c.o.c.} \end{cases}$$

Y cumple con las siguientes Propiedades

- a)  $f(x) \geq 0 \forall x \in \mathbb{R}$
- b)  $\sum_{x \in X} f(x) = 1$

#### Definición

Sea  $X$  una variable aleatoria discreta y  $f(x)$  la función masa de probabilidad de  $X$ , la función de distribución de  $X$ , denotada por  $F(x) \rightarrow [0, 1]$  se define como:

$$F(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} (f(x_i))$$

### Definición

Sea  $X$  una variable aleatoria continua, decimos que la función integrable  $f(x) : \mathbb{R} \rightarrow [0, \infty)$  es la función de densidad de  $X$  si para cualquier intervalo  $(a, b) \in \mathbb{R}$  se cumple que:

$$\mathbb{P}(X \in (a, b)) = \int_a^b f(x) dx$$

y cumple:

$$\begin{array}{l} \text{a) } f(x) \geq 0 \\ \text{b) } \int_{-\infty}^{\infty} f(x) = 1 \end{array}$$

### Definición

Sea  $X$  una variable aleatoria continua y  $f(t)$  la función de densidad de  $x$ , la función de distribución de  $X$ , denotada por  $F(x) = \mathbb{R} \rightarrow [0, 1]$ , se define como

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f(t) dt$$

### Definición

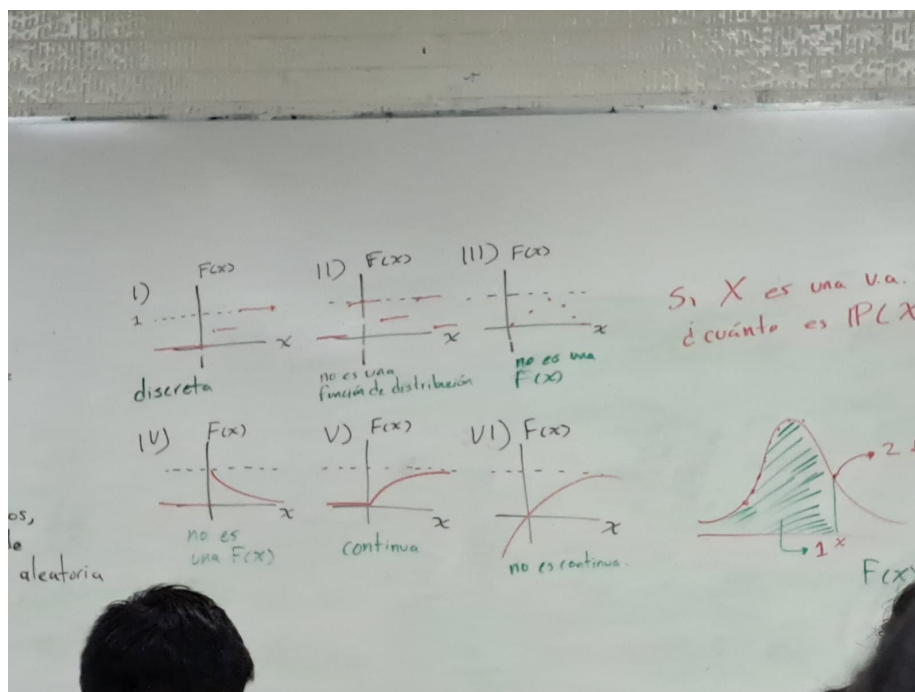
La variable aleatoria  $X$  se llama continua si su correspondiente función de distribución es continua y creciente

### Definición

La variable aleatoria  $X$  se llama discreta si su correspondiente función de distribución es constante por pedazos

### Ejemplo

Para cada uno de los siguientes incisos, identifica si la gráfica de la función de distribución corresponde a una variable aleatoria discreta o continua.



## 8. 1 Septiembre 2023

### 8.1. Métodos de las funciones de distribución

Sea  $X$  una variable aleatoria con función de densidad  $f(x)$  y  $\varphi$  una función de  $X$ , entonces la variable aleatoria  $Y = \varphi(x)$  con función de distribución  $F_Y(y) = \mathbb{P}(Y \leq y)$ ,  $f(y)$  se puede calcular al integrar la región para la cual  $Y \leq y$ . Además, la función de densidad de la variable  $y$  se puede calcular al derivar  $F_Y(y)$ .

#### Ejemplo 2

Encuentra la función de densidad de la variable  $Y$ , si  $Y = 3X - 1$  donde.

$$f(x) = \begin{cases} 2x & \text{si } 0 \leq x \leq 1 \\ 0, & \text{c.o.c.} \end{cases}$$

$$\begin{aligned}
F_y(y) &= \mathbb{P}(Y \leq y) \\
&= \mathbb{P}(3x - 1 \leq y) \\
&= \mathbb{P}(3x \leq y + 1) \\
&= \mathbb{P}\left(x \leq \frac{y+1}{3}\right) \\
&= \int_0^{\frac{y+1}{3}} f(x) dx
\end{aligned}$$

- Si  $y = -1$ ,  $\frac{y+1}{3} = 0$
- Si  $y < -1$  entonces  $F_y(y) = 0$
- Si  $y = 2$  entonces  $F_y(y) = 1$

$$\begin{aligned}
F_y(y) &= \int_0^{\frac{y+1}{3}} 2x dx \\
&= \frac{(y+1)^2}{9}
\end{aligned}$$

$$F_y(y) = \begin{cases} 0 & \text{si } y < -1 \\ \frac{(y+1)^2}{9}, & -1 \leq y \leq 2 \\ 1 & \text{si } y > 2 \end{cases}$$

Si derivamos  $\frac{d(F_y(y))}{dy}$

$$f_y(y) = \begin{cases} \frac{2}{9}(y+1)^2 & \text{si } -1 \leq y \leq 2 \\ 0 & \text{c.o.c} \end{cases}$$

### Ejemplo 3

La función de densidad conjunta de  $X_1$  y  $X_2$  está dada por

$$f(x_1, x_2) = \begin{cases} 3x_1 & \text{si } -0 \leq x_2 \leq x_1 \leq 1 \\ 0 & \text{c.o.c} \end{cases}$$

Encuentra la función de densidad  $Y = x_1 - X_2$

$$\begin{aligned}
F_y(y) &= \mathbb{P}(Y \leq y) \\
&= \mathbb{P}(X_1 - X_2 \leq y) \\
&= 1 - \mathbb{P}(X_1 - X_2 > y) \\
&= 1 - \int_y^1 \int_0^{X_1-y} f(x_1, x_2) dx_2 dx_1 \\
&= 1 - \int_y^1 \int_0^{X_1-y} 3x_1 dx_2 dx_1 \\
&= 1 - \int_y^1 [3x_1 x_2]_0^{x_1-y} dx_1 \\
&= 1 - \int_y^1 [3x_1(x_1 - y)] dx_1 \\
&= 1 - \int_y^1 [3x_1^2 - 3x_1 y] dx_1 \\
&= 1 - [x_1^3 - \frac{3}{2}x_1^2 y]_y^1 \\
&= 1 - [1 - \frac{3}{2}y - (y^3 - \frac{3}{2}y^3)] \\
&= [1 - \frac{3}{2}y + \frac{1}{2}y^3] \\
&= \frac{3}{2}y - \frac{1}{2}y^3 \\
&= \frac{1}{2}(3y - y^3)
\end{aligned}$$

## 8.2. Teorema de cambio de variable

Sea  $X$  una variable aleatoria continua con valores en el intervalo  $(a, b) \in \mathbb{R}$  y con función de densidad  $f_x(x)$ . Sea  $\varphi(a, b) \rightarrow \mathbb{R}$  una función continua, estrictamente creciente o decreciente y con inversa. Entonces la variable aleatoria  $Y = \varphi$  toma valores en el intervalo  $\varphi(a, b)$  y tiene función de densidad

$$f(x_1, x_2) = \begin{cases} f_x(\varphi^{-1}(y)) \left| \frac{d(\varphi^{-1}(y))}{dy} \right| & \text{si } y \in \varphi(a, b) \\ 0 & \text{c.o.c} \end{cases}$$

### Ejemplo 4

Sea  $x \sim Unif(0, 1)$  y  $\varphi(x) = \frac{1}{\lambda} \ln(x) = y$  obten  $f_y(y)$

$$f_x(x) = \begin{cases} 1 & \text{si } 0 \leq x \leq 1 \\ 0 & \text{c.o.c} \end{cases}$$

- Si  $x = 1$ ,  $\varphi(1) = -\frac{1}{\lambda} \ln(1) = 0$

- Si  $x = 0$ ,  $\varphi(0) = -\frac{0}{\lambda} \ln(0) = \infty$

Para obtener  $\varphi^{-1}(y)$ , se tiene que:

$$\begin{aligned} -\frac{1}{\lambda} \ln(x) &= y \\ \ln(x) &= -\lambda y \\ x &= e^{-\lambda y} \\ &= \mathbb{P}\left(x \leq \frac{y+1}{3}\right) \\ &= \int_0^{\frac{y+1}{3}} f(x) dx \end{aligned}$$

## 9. 08 de Septiembre del 2023

### 9.1. Método de la función generadora de momentos

#### Definición

El valor esperado o esperanza de la variable aleatoria de  $X$ , denotada por  $E(x)$  es.

$$E(x) = \begin{cases} \int_{-\infty}^{\infty} xf(x)dx & \text{si } X \text{ es una v.a.c} \\ \sum_x xf(x) & \text{si } X \text{ es una v.a.d} \end{cases}$$

#### Definición

La varianza de la variable aleatoria  $X$  se define como:

$$\begin{aligned} \text{var}(x) &= E[x - E(x)]^2 \\ &= E[x^2 - 2xE(x) + E^2(x)] \\ &= E(x^2) - 2E(x)E(x) + E^2(x) \\ &= E(x^2) - 2E^2(x) + E^2(x) \\ &= E(x^2) - E^2(x) \end{aligned}$$

#### Definición

Sea  $X$  una variable aleatoria continua con media  $\mu$ , el k-ésimo momento central de  $X$  es

$$m_k = \int_{-\infty}^{\infty} (x - \mu)^k f(x) dx$$

#### Definición

Sea  $X$  una variable aleatoria, la función generadora de momentos es:

$$M_x(t) = E(e^{tx})$$

#### Ejemplo 6

\*Insertar las fotos del ejercicio 6 y 7\*

#### Ejemplo 8

$$\begin{aligned} -M_x(t) &= \sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n) \\ -\frac{d^n[M_x(t)]}{dt^n} \Big|_{t=0} &= E(x^n) \end{aligned}$$

Si  $X$  y  $Y$  son independientes. Entonces  $M_{x+y}(t) = M_x(t) \cdot M_y(t)$

Si  $M_x(t) = M_y(t)$  para todos los valores de  $t$ , entonces  $X$  y  $Y$  tienen la misma función de distribución



pregunta de examen: si ya conoces la función de densidad y como se identificara la moda?

## 10. 13 de Septiembre del 2023

### Ejemplo 8

Muestra que si  $X \sim N(\mu, G^2)$ , entonces  $Y = \frac{X-\mu}{G} \sim N(0, 1)$

Si  $N \sim (\mu G^2)$ , entonces  $M_x(t) = e(\mu t + \frac{G^2 t^2}{2})$

Si  $Y \sim N(0, 1)$

$$M_y(t) = e^{\frac{t^2}{2}}$$

De manera que

$$\begin{aligned} M_y(t) &= E(e^{ty}) \\ &= E(e^{t \frac{x-\mu}{G}}) \\ &= E(e^{\frac{tx}{G}}) E(e^{-\frac{t\mu}{G}}) \\ &= e^{-\frac{t\mu}{G}} E(e^{\frac{t}{G}x}) \\ &= e^{-\frac{t\mu}{G}} M_x\left(\frac{t}{G}\right) \\ &= e^{-\frac{t\mu}{G}} \left[ e^{\mu \frac{t}{G} + \frac{G^2}{2} \frac{t^2}{G^2}} \right] \\ &= e^{-\frac{t\mu}{G} + \frac{t\mu}{G} + \frac{t^2}{2}} = e^{\frac{t^2}{2}} \end{aligned}$$

## 11. 18 de Septiembre del 2023

### Ejemplo 11

Sea  $X \sim \text{Exp}(1/100)$ . Encuentra  $f_{x(1)}(x)$  para una muestra de tamaño 2.

$$\begin{aligned} f_{x(1)}(x) &= n f_{(x)} [1 - F_{(x)}]^{n-1} \\ \text{Si } X &\sim \text{Exp}(1/100) \\ f(x) &= 100 \exp^{-100x} \\ F(x) &= 1 - \exp^{-100x} \\ n &= 2 \\ f_{x(1)}(x) &= 2(100 \exp^{-100x})(1 - 1 + \exp^{-100x})^{2-1} \\ &= 200 \exp^{-200x} \end{aligned}$$

por lo tanto:  $f_{x(1)}(x) \sim \text{Exp}(\beta = 1/200)$

## 11.1. Distribuciones derivadas de la normal

### 11.1.1. Normal estándar

Si  $X \sim N(0, 1)$  se dice que  $x$  que se distribuye como una normal estándar.

#### Propiedades

- La media, la mediana y la moda son iguales a  $\mu$
- El 68 % de las observaciones caen en el intervalo  $[\mu - \sigma, \mu + \sigma]$
- El 95 % de las observaciones caen en el intervalo  $[\mu - 2\sigma, \mu + 2\sigma]$
- El 99,7 % de las observaciones caen en el intervalo  $[\mu - 3\sigma, \mu + 3\sigma]$
- $\mathbb{P}(-1 \leq x \leq 1) = 68 \%$
- $\mathbb{P}(x \leq 0) = 50 \%$
- $\mathbb{P}(x = 0) = 0 \%$
- $\mathbb{P}(0 \leq x \leq 1) = 34 \%$

### 11.1.2. Ji cuadrada

Sean  $Z_1, \dots, Z_k$  variables aleatorias con distribución normal estándar

$$x = z_1^2 + \dots + z_k^2$$

Entonces,  $X \sim X_k^2$  donde  
 $E(x) = k$  y  $Var(x) = 2k$

## 12. Unidad 3

### Distribuciones muestrales

### 12.1. Introducción

#### Definición

Una muestra aleatoria es una colección de variable aleatoria  $X_1, \dots, X_n$  que cumplen con la condición de ser independientes y de tener cada una de ellas la misma distribución.

#### Definición

Un estadístico es una variable aleatoria de la forma  $g(X_1, \dots, X_n)$  en donde  $X_1, \dots, X_n$  es una muestra aleatoria y  $g : \mathbb{R}^n \rightarrow \mathbb{R}$

Por ejemplo, la media muestral es un estadístico al igual que la varianza muestral

#### Definición

Dado que los estadísticos son funciones de variables aleatorias entonces los estadísticos tienen funciones de probabilidad que se denominan Distribuciones muestrales.

### Definición

Un parámetro es una caracterización numérica de la población de manera que describe parcialmente la función de la probabilidad.

población	Muestra
N	n
$\mu$	$\bar{x}$
$\sigma^2$	$s^2$

### Proposición

Sea  $X_1, \dots, X_n$  una muestra aleatoria con función generadora de momentos  $M_{x1}(t), \dots, M_{xn}(t)$  y

$$y = a_1 X_1 + \dots + a_n X_n$$

En donde  $a_1, \dots, a_n$  son constantes, entonces.

$$M_y(t) = M_{x1}(a_1 t) \dots M_{xn}(a_n t)$$

### Ejemplo 1

Sea  $X_1, \dots, X_n$  una muestra aleatoria con distribución normal donde  $E(x_i) = \mu_i$  y  $Var(X_i) = \sigma_i^2$ ,  $i = 1, \dots, n$ . Si  $Y = a_1 X_1 + \dots + a_n X_n$ , muestra que  $Y$  se distribuye como una normal.

Si  $X_i \sim N(\mu_i, \sigma_i^2)$

entonces

$$M_{xi}(t) = e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}}$$

$$M_{xi}(t) = e^{\mu_i a_i t + \frac{a_i^2 \sigma_i^2 t^2}{2}}$$

## 13. 22 de Septiembre del 2023

### 13.1. Distribución de muestreo de $\bar{X}$

Sea  $X_1, \dots, x_n$  una muestra aleatoria tal que  $E(x_i) = \mu$  y  $Var(x_i) = \sigma^2$  para  $i = 1, \dots, n$ . Entonces

$$E(\bar{X}) = \mu$$

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

### Demostración

$$\begin{aligned}
E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^{x_i}\right) \\
&= E\left(\frac{1}{n} [x_1, \dots, c_n]\right) \\
&= \frac{1}{n} (E(x_1) + \dots + E(x_n)) \\
&= \frac{1}{n} (\mu + \dots + \mu) \\
&= \frac{n\mu}{n} = \mu
\end{aligned}$$

$$\begin{aligned}
Var(\bar{X}) &= Var\left(\frac{1}{n} \sum_{i=1}^{x_i}\right) \\
&= Var\left(\frac{1}{n} [x_1, \dots, c_n]\right) \\
&= \frac{1}{n^2} (Var(x_1) + \dots + Var(x_n)) \\
&= \frac{1}{n^2} (\sigma^2 + \dots + \sigma^2) \\
&= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}
\end{aligned}$$

### Ejemplo 3

Sea  $X_1, \dots, x_n$  v.a.i.i.d con distribución normal donde  $E(x_1) = 3,25$ ,  $E(x_2) = 3,3$ ,  $\sqrt{Var(x_1)} = 0,005$  y  $\sqrt{Var(x_2)} = 0,006$   
 Sea  $y = x_2 - x_1$ . Calcula  $E(y)$  y  $Var(y)$

$$E(y) = E(x_2 - x_1) = E(x_2) - E(x_1) = 3,3 - 3,25 = 0,05$$

$$Var(y) = Var(x_2 - x_1) = Var(X_2) + Var(x_1) = (0,000061)^2 + (0,005)^2 = 0,000061$$

$$\text{Si } z \sim x_4^2$$

$$\mathbb{P}(z > 4) = \mathbb{P}\left(\frac{z - E(z)}{\sqrt{Var(z)}} > \frac{4 - E(z)}{\sqrt{Var(z)}}\right) = \mathbb{P}\left(\frac{z - 4}{\sqrt{8}} > \frac{4 - 4}{\sqrt{8}}\right) = \mathbb{P}(x > 0)$$

### 13.2. Teorema central del limite

Sea  $X_1, \dots, x_n$  una sucesión de v.a.i.i.d. tales que para cada natural  $n$ ,  $E(x_n) = \mu$  y  $Var(x_n) = \sigma^2$ . Entonces

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

## 14. 25 de Septiembre del 2023

### Ejemplo 4

Se ha seleccionado una muestra al azar de 50 mujeres de una población de mayores de 18 años. se desconoce la estatura media de la población pero en la muestra observada se obtuvo que  $\bar{x} = 1,60\text{m}$ . Si se sabe por otros estudios que la desviación estandar poblacional es de  $0,33\text{m}$

Determina la probabilidad de que la media muestral no difiera en más de  $0,01\text{m}$  de la media poblacional.

$$\bar{x}, E(\bar{x}) = \mu$$

$$(\bar{x}) = \mu \qquad \qquad \qquad = \bar{x} = 1,60$$

$$\begin{aligned} \mathbb{P}(|\bar{x} - \mu| < 0,01) &= \mathbb{P}(-0,01 < \bar{x} - \mu < 0,01) \\ &= \mathbb{P}\left(\frac{-0,01}{\frac{0,33}{\sqrt{50}}} < \frac{\bar{x} - \mu}{\frac{0,33}{\sqrt{50}}} < \frac{0,01}{\frac{0,33}{\sqrt{50}}}\right) \\ &= \mathbb{P}\left(\frac{-0,01}{\frac{0,033}{\sqrt{50}}} < z < \frac{0,01}{\frac{0,033}{\sqrt{50}}}\right) \\ &= \mathbb{P}(-2,14 < z < 2,14) \\ &= \mathbb{P}(z < 2,14) - \mathbb{P}(z < -2,14) \end{aligned}$$

### 14.1. La distribución de muestreo de $s^2$

Sean  $(x_1, y_1) \dots, (x_n, y_n)$  v.a.i.i con distribución normal con media  $\mu$  conocida y varianza  $\sigma^2$ . Entonces

$$\frac{nS_n^2}{\sigma^2} \sim \chi_n^2$$

Donde

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

### Ejemplo 5

Sean  $(x_1, y_1) \dots, (x_n, y_n)$  v.a.i.i con distribución  $N(\mu = 10, \sigma^2 = 0,01)$ . Calcula la probabilidad de que la varianza muestral sea mayor que  $0,014$

$$\begin{aligned}
\mathbb{P}(S_n^2 > 0,014) &= \mathbb{P}\left(\frac{nS_n^2}{\sigma^2} > \frac{0,014n}{\sigma^2}\right) \\
&= \mathbb{P}\left(X_n^2 > \frac{0,014}{0,01}n\right) \\
&\quad \text{Sin} = 25 \\
\mathbb{P}(X_{25}^2 > \frac{0,014(25)}{0,01}) &= \mathbb{P}(X_{25}^2 > 35) \\
&= 1 - \mathbb{P}(X_n^2 < 35)
\end{aligned}$$

En r 1 - pchisq(35,25) = 0.088203

## 14.2. Aproximación normal

$$\mathbb{P}(X_n^2 > 35)$$

## 15. 27 de Septiembre del 2023

### 15.1. Estimacion puntual y por intervalos

estimador *neq* Estimacion

### 15.2. Propiedades de los estimadores

**Definición** Un estimador  $\theta$  del parametro  $\theta$  se dice que es insesgado si  $E(\hat{\theta}) = \theta$

#### Ejemplo 2

Define el espacio parametral de las siguientes Distribuciones

$$Exp(\lambda), H = (0, \infty)$$

$$N(\mu, \sigma^2), H = (-\infty, \infty) \times (0, \infty)$$

**Definición** Un estadístico  $\theta_n$  basado en una muestra de tamaño  $n$  es un estimador asintotico insesgado para el parametro  $\theta$  si

$$\lim_{n \rightarrow \infty} E(\theta_n) = \theta$$

$$\mathbb{P}(z \leq 0,41) = 0,6591$$

**Definición.** Sea  $\hat{\theta}_n$  un estimador para el parametro  $\theta$ , basado en una muestra de tamaño  $n$ ,  $\hat{\theta}_n$  es consistente si para cualquier  $\epsilon > 0$

## 16. 2 de Octubre del 2023

### 16.1. Métodos para obtener estimadores puntuales

#### 16.1.1. Método de momentos

Consiste en igualar los momentos poblacionales con los correspondientes momentos muestrales y resolver esta ecuación o sistema de ecuaciones para el parámetro o vector de parámetros desconocidos.

$$E(x) \simeq \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

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

Si  $X$  es una v.a continua

$$E(x^j) = \int_{-\infty}^{\infty} x^j f(x) dx$$

Por ejemplo, si  $X \sim N(\mu, \sigma^2)$

$$E(x) = \mu$$

#### Ejemplo 4

Sea  $X$  una variable aleatoria con función de densidad

$$f(x) = \begin{cases} \theta x^{\theta-1} & \text{si } 0 \leq x \leq 1 \\ 0 & \text{eoc} \end{cases}$$

Utiliza el método de momentos para estimar  $\theta$

$$\begin{aligned} E(x) &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_{-1}^1 x \theta x^{\theta-1} \\ &= \int_{-1}^1 \theta x^{\theta} dx \\ &= \frac{\theta}{\theta+1} x^{\theta+1} \Big|_{-1}^1 \\ &= \frac{\theta}{\theta+1} \end{aligned}$$

Por lo tanto

$$\frac{\hat{\theta}}{\hat{\theta} + 1} = \bar{x}$$

$$\hat{\theta} = \frac{\bar{x}}{1 - \bar{x}}$$

### Ejemplo 5

Sea  $X_1, \dots, X_n$  una muestra aleatoria con distribución  $N(\mu, \sigma^2)$  usa el metodo de momentos para estimar  $\mu$  y  $\sigma^2$

$$E(x) = \frac{1}{n} \sum x_i$$

$$E(x^2) = \frac{1}{n} \sum x_i^2$$

$$\mu = \bar{x}$$

Por otro lado

$$\begin{aligned} E(x^2) &= \text{var}(x) + E(x)^2 \\ &= \sigma^2 + \mu^2 \\ &= \int_{-\infty}^{\infty} x^2 f(x) dx \end{aligned}$$

$$\begin{aligned} \hat{\sigma}^2 + \hat{\mu}^2 &= \frac{1}{n} \sum x_i^2 \\ \hat{\sigma}^2 + \bar{x}^2 &= \frac{1}{n} \sum x_i^2 \\ \hat{\sigma}^2 &= \frac{1}{n} \sum x_i^2 - \bar{x}^2 \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

Insertar imagen del 2 de Octubre

Insertar ejemplo 6

## 17. 4 de Octubre del 2023

### Ejemplo 7

$$L(\beta|x) = \prod_{i=1}^n f(x_i|\beta)$$



$$= \frac{1}{\beta^n} e^{-\sum x_i/\beta}$$

$$\begin{aligned}\log[L(\beta|x)] &= \log\left[\frac{1}{\beta^n} e^{-\sum x_i/\beta}\right] \\ &= \log\left[\frac{1}{\beta^n}\right] + \log[e^{-\sum x_i/\beta}] \\ &= \log[\beta^{-n}] - \frac{\sum x_i}{\beta} \\ &= -n \log(\beta) - \sum x_i \beta^{-1}\end{aligned}$$

### Ejemplo 8

Sea  $X_1, \dots, X_n$  una m.a que se distribuye como una  $N(\mu, \sigma^2)$ . Utiliza el método de máxima verisimilitud para encontrar  $\hat{\mu}$ ,  $\hat{\sigma}^2$

$$\begin{aligned}L(\mu, \sigma^2|x) &= \prod_{i=1}^n f(\mu, \sigma^2|x_i) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\ &= \prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum (x_i-\mu)^2}{2\sigma^2}}\end{aligned}$$

$$\begin{aligned}\log[L(\mu, \sigma^2|x)] &= \log \left[ (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum (x_i-\mu)^2}{2\sigma^2}} \right] \\ &= \log \left[ (2\pi\sigma^2)^{-\frac{n}{2}} \right] + \log \left[ e^{-\frac{\sum (x_i-\mu)^2}{2\sigma^2}} \right] \\ &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{\sum (x_i - \mu)^2}{2\sigma^2}.\end{aligned}$$

$$\frac{d \log[L(\mu, \sigma^2|x)]}{d\mu}$$

Para resolver la derivada con respecto a  $\mu$  de la log-verosimilitud de una distribución normal, primero recordemos la expresión para  $\log[L(\mu, \sigma^2|x)]$ :

$$\log[L(\mu, \sigma^2|x)] = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{\sum (x_i - \mu)^2}{2\sigma^2}$$

Ahora, vamos a derivar con respecto a  $\mu$ :

$$\frac{d \log[L(\mu, \sigma^2|x)]}{d\mu} = \frac{d}{d\mu} \left( -\frac{n}{2} \log(2\pi\sigma^2) - \frac{\sum (x_i - \mu)^2}{2\sigma^2} \right)$$

El primer término,  $-\frac{n}{2} \log(2\pi\sigma^2)$ , no depende de  $\mu$ , por lo que su derivada es 0.

El segundo término, centrémonos en derivar  $-\frac{\sum (x_i - \mu)^2}{2\sigma^2}$  con respecto a  $\mu$ :

$$\frac{d}{d\mu} \left( -\frac{\sum (x_i - \mu)^2}{2\sigma^2} \right) = -\frac{1}{2\sigma^2} \sum 2(x_i - \mu)(-1) = \frac{\sum (x_i - \mu)}{\sigma^2}$$

Por lo tanto, la derivada de la log-verosimilitud con respecto a  $\mu$  es:

$$\frac{d \log[L(\mu, \sigma^2|x)]}{d\mu} = \frac{\sum (x_i - \mu)}{\sigma^2}$$

Esta derivada nos da la tasa de cambio de la log-verosimilitud con respecto a  $\mu$ . Para encontrar el valor de  $\mu$  que maximiza la log-verosimilitud (y, por lo tanto, la verosimilitud), podemos resolver para  $\mu$  cuando esta derivada es igual a cero. Este proceso nos dará el estimador de máxima verosimilitud (MLE) para  $\mu$  en una distribución normal.

$$\begin{cases} \frac{\sum (x_i - \hat{\mu})}{\sigma^2} = 0 \\ -\frac{n}{2\sigma^2} + \frac{\sum (x_i - \mu)^2}{2(\sigma^2)^2} = 0 \end{cases}$$

Vamos a resolver el sistema de ecuaciones que nos has dado paso a paso.  
Dadas las ecuaciones:

1)

$$\frac{\sum (x_i - \hat{\mu})}{\sigma^2} = 0$$

2)

$$-\frac{n}{2\sigma^2} + \frac{\sum (x_i - \mu)^2}{2(\sigma^2)^2} = 0$$

\*\*Resolviendo la primera ecuación:\*\* De la ecuación (1), tenemos:

$$\sum (x_i - \hat{\mu}) = 0$$

Sumando todos los  $x_i$  y reorganizando los términos, tenemos:

$$\sum x_i = n\hat{\mu}$$

De aquí, podemos deducir que:

$$\hat{\mu} = \frac{\sum x_i}{n}$$

Esto es simplemente el promedio de los valores de  $x_i$ .

**\*\*Resolviendo la segunda ecuación:\*\*** La ecuación (2) es:

$$-\frac{n}{2\sigma^2} + \frac{\sum (x_i - \mu)^2}{2(\sigma^2)^2} = 0$$

Multiplicando todo por  $2(\sigma^2)^2$ :

$$-n\sigma^2 + \sum (x_i - \mu)^2 = 0$$

Reorganizando, obtenemos:

$$\sum (x_i - \mu)^2 = n\sigma^2$$

Dividimos ambos lados por  $n$ :

$$\frac{\sum (x_i - \mu)^2}{n} = \sigma^2$$

Aquí,  $\sigma^2$  es simplemente la varianza de los valores de  $x_i$ .

Por lo tanto, las soluciones para el sistema de ecuaciones son:

$$\hat{\mu} = \frac{\sum x_i}{n}$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

### **Teorema Cota inferior de Cramér-Rao**

Sea  $X_1, \dots, X_n \sim f(x|\theta)$ ,  $\theta \in \Theta$  Sea  $T$  un estimador de  $\theta$  y

$$Var(T) \geq \frac{1}{nE\left[\frac{d \log(f(x|\theta))}{d\theta}\right]^2}$$

Si  $T$  es insesgado para  $\theta$  y  $Var(T) = CICR((\theta))$ .  $T$  es UMVUE (Estimador, insesgado de mínima varianza)

### **Ejemplo 9**

Calcula el  $CICR(\theta)$  para el modelo Bernoulli

Recordemos que una variable aleatoria  $X$  sigue una distribución Bernoulli con parámetro  $\theta$  si su función de masa de probabilidad (fmp) es:

$$P(X = k|\theta) = \theta^k(1 - \theta)^{1-k}$$

para  $k \in \{0, 1\}$  y  $0 < \theta < 1$ .

Para calcular la Cota Inferior de Cramér-Rao, primero necesitamos encontrar la información de Fisher  $I(\theta)$ . La información de Fisher para una muestra es:

$$I(\theta) = E \left[ \left( \frac{\partial \ln f(X; \theta)}{\partial \theta} \right)^2 \right]$$

Calculemos primero la derivada del logaritmo de la fmp respecto de  $\theta$ .

Dada:

$$\ln f(X|\theta) = \ln(\theta^x) + \ln((1-\theta)^{1-x})$$

$$\ln f(X|\theta) = x \ln(\theta) + (1-x) \ln(1-\theta)$$

Derivando con respecto a  $\theta$  obtenemos:

$$\frac{\partial \ln f(X|\theta)}{\partial \theta} = \frac{x}{\theta} - \frac{1-x}{1-\theta}$$

Elevando al cuadrado esta derivada:

$$\left( \frac{\partial \ln f(X|\theta)}{\partial \theta} \right)^2 = \left( \frac{x}{\theta} - \frac{1-x}{1-\theta} \right)^2$$

Pedirle a chatgpt concluir El profe concluyo  $CICR(\theta) = \frac{\theta(1-\theta)}{n}$

### **Ejemplo 10**

Muestar que  $\hat{\theta} = \bar{x}$  es UMVUE para una m.a con distribucion  $Bern(\theta)$

## **18. 11 de Octubre del 2023**

### **Ejemplo**

Muestra que la media muestral es UMVUE para el parametro  $\beta$  de una exponencial

Cuota inferior de Cramér-Rao

$$CICR(\theta) = \frac{1}{nE \left[ \frac{d[Ln(f(X|\theta))]}{d\theta} \right]^2}$$

Funcion de dednsidad exponencial

$$f(x|\beta) = \frac{1}{\beta} e^{-x/\beta}$$

Primero, calculemos el logaritmo de esta función:

$$\ln \left( \frac{1}{\beta} e^{-\frac{x}{\beta}} \right) = -\ln(\beta) - \frac{x}{\beta}$$

Ahora, calculamos su derivada con respecto de  $\beta$

$$\begin{aligned} E \left[ \frac{dLog(f(x|\beta))}{d\beta} \right]^2 &= E \left[ -\frac{1}{\beta} + \frac{x}{\beta^2} \right]^2 \\ &= E \left[ \frac{x}{\beta^2} - \frac{1}{\beta} \right]^2 \\ &= E \left[ \frac{x^2}{\beta^4} - \frac{2x}{\beta^3} + \frac{1}{\beta^2} \right] \end{aligned}$$

Sabemos que

$$Var(x) = \beta^2 \quad E(x) = \beta$$

Entonces al distribuir la esperanza tenemos:

Nota:  $\beta$  es una constante por lo tanto solo no fijamos en  $x$

$$\frac{E[x^2]}{\beta^4} - \frac{E[2x]}{\beta^3} + \frac{1}{\beta^2}$$

para calcular  $E[x^2]$  podemos observar la varianza y despejar el valor de  $E[x^2]$

$$\begin{aligned} Var(x) &= E[x^2] - E^2[x] = \beta^2 \\ &= E[x^2] - \beta^2 = \beta^2 \\ &= E[x^2] = 2\beta^2 \end{aligned}$$

por lo tanto

$$\frac{2\beta^2}{\beta^4} - \frac{2\beta}{\beta^3} + \frac{1}{\beta^2} = \frac{1}{\beta^2}$$

$$CIRC(\beta) = \frac{1}{n \frac{1}{\beta^2}} = \frac{\beta^2}{n}$$

$$\begin{aligned} E(\bar{x}) &= E\left(\frac{1}{n} \sum_{i=1}^n x_i\right) \\ &= \frac{1}{n} E(x_1 + \cdots + x_n) \\ &= \frac{1}{n} [E(x_1) + \cdots + E(x_n)] \\ &= \frac{1}{n} [n\beta] \\ &= \beta \end{aligned}$$

$$\begin{aligned} Var(\bar{x}) &= Var\left(\frac{1}{n} \sum_{i=1}^n x_i\right) \\ &= \frac{1}{n^2} Var(x_1 + \cdots + x_n) \\ &= \frac{1}{n^2} [Var(x_1) \cdots Var(x_n)] \\ &= \frac{1}{n^2} [n\beta^2] \\ &= \frac{\beta^2}{n} \end{aligned}$$

Como  $\bar{x}$  es insesgada y  $Var(\bar{x}) = \frac{\beta^2}{n} = CIRC(\beta)$ , entonces  $\bar{X}$  es UMVUE

## 19. 13 de Octubre del 2023

**Teorema de factorizacion** Un estadístico  $T = \tau(X_1, \dots, X_n)$  es suficiente para  $\theta$  si y solo si,  $L(\theta|X)$  se puede escribir como

$$g(T(X|\theta))h(X)$$

### Ejemplo 12

Muestra que  $T = \sum_{i=1}^n x_i$  es suficiente para  $f(X|\theta) = \theta^x(1-\theta)^{1-x}$

$$L(\theta|x) = \prod_{i=1}^n f(x_i|\theta)$$

## 20. 16 de octubre del 2023

### Ejemplo 13

Construye un intervalo de confianza de nivel  $1 - \alpha$  para el parámetro  $\theta$ , dada una muestra aleatoria de tamaño  $n$ , con función de distribución  $f(X|\theta) \sim \text{unif}(0, \theta)$ ,  $\theta \in \Theta$ ,

### Intervalos de confianza para la distribución Normal

Parámetro	Supuestos	Intervalo
$\mu$	$\sigma$ conocida	$\left( \bar{X} - q_{z(1-\frac{\alpha}{2})} \frac{\sigma}{\sqrt{n}}, \bar{X} + q_{z(1-\frac{\alpha}{2})} \frac{\sigma}{\sqrt{n}} \right)$ $z \sim N(0, 1)$
$\mu$	$\sigma$ desconocida	$\left( \bar{X} - q_{t(1-\frac{\alpha}{2})} \frac{s}{\sqrt{n}}, \bar{X} + q_{t(1-\frac{\alpha}{2})} \frac{s}{\sqrt{n}} \right)$ $t \sim t_{n-1}$
$\mu_1 - \mu_2$	$\sigma_1$ y $\sigma_2$ y conocidas	$\left( \bar{X}_1 - \bar{X}_2 \pm q_{z(1-\frac{\alpha}{2})} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$ $z \sim N(0, 1)$
$\mu_1 - \mu_2$	$\sigma_1 = \sigma_2$ y desconocidas	$\left( \bar{X}_1 - \bar{X}_2 \pm q_{t(1-\frac{\alpha}{2})} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$ $t \sim t_{n_1+n_2-2}$
$\mu_1 - \mu_2$	$\sigma_1 \neq \sigma_2$ y desconocidas	$\left( \bar{X}_1 - \bar{X}_2 \pm q_{t(1-\frac{\alpha}{2})} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$ $t \sim t_{\frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(n_1-1)^{-1}s_1^4/n_1^2 + (n_2-1)^{-1}s_2^4/n_2^2}}$



## Intervalos de confianza para la distribución Normal

Parámetro	Intervalo
$\sigma^2$	$\left( \frac{(n-1)s^2}{q_{\chi^2_{(1-\frac{\alpha}{2})}}}, \frac{(n-1)s^2}{q_{\chi^2_{(\frac{\alpha}{2})}}} \right)$ $\chi^2 \sim \chi^2_{n-1}$





## Intervalos de confianza para datos pareados

Parámetro	Intervalo
$\mu_d$	$\left( \bar{D} - q_{t(1-\frac{\alpha}{2})} \frac{s_d}{\sqrt{n}}, \bar{D} + q_{t(1-\frac{\alpha}{2})} \frac{s_d}{\sqrt{n}} \right)$ $t \sim t_{n-1}$



## Intervalos de confianza para la distribución Bernoulli

Parámetro	Intervalo
$p$	$\left( \bar{X} - q_{z(1-\frac{\alpha}{2})} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}, \bar{X} + q_{z(1-\frac{\alpha}{2})} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \right)$ $z \sim N(0, 1)$
$p_1 - p_2$	$\left( \bar{X}_1 - \bar{X}_2 \pm q_{z(1-\frac{\alpha}{2})} \sqrt{\frac{\bar{X}_1(1-\bar{X}_1)}{n_1} + \frac{\bar{X}_2(1-\bar{X}_2)}{n_2}} \right)$ $z \sim N(0, 1)$



#### Ejemplo 14

Se eligieron al azar 140 hombre y 150 mujeres entre 40 y 50 años de edad y se halló que 56 de los hombres y 27 de las mujeres usan lentes. Con estos datos, construye un intervalo de confianza de 95 % para la diferencia de las proporciones reales de hombres y mujeres que usan lentes en la población.

Primero, identificamos los datos:

$n_1$  = número de hombres = 140

$n_2$  = número de mujeres = 150

$x_1$  = número de hombres que usan lentes = 56

$x_2$  = número de mujeres que usan lentes = 27

De aquí, podemos encontrar las proporciones muestrales:

$$x_1 = \frac{x_1}{n_1} = \frac{56}{140} = 0,4$$

$$x_2 = \frac{x_2}{n_2} = \frac{27}{150} = 0,18$$

La diferencia de proporciones muestrales es:

$$x_1 - x_2 = 0,4 - 0,18 = 0,22$$

Para un intervalo de confianza del 95 %,  $z$  es aproximadamente 1.96 (dado que  $z$  sigue una distribución normal estándar).

$$q_z(1 - \frac{\alpha}{2}) = q_z(1 - \frac{1 - 0,95}{2})$$

$$q_z(1 - \frac{0,05}{2}) = q_z(1 - 0,025)$$

$$q_z(0,975) = 1,96$$

Esto se sabe por la tabla  $z$

$z$	0	0.01	0.02	0.03	0.04	0.05	0.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750

Usando la fórmula dada:

$$\bar{X}_1 - \bar{X}_2 \pm q_z \left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{\bar{X}_1(1 - \bar{X}_1)}{n_1} + \frac{\bar{X}_2(1 - \bar{X}_2)}{n_2}}$$

Calculamos el margen de error:

$$ME = 1,96 \sqrt{\frac{0,4(1 - 0,4)}{140} + \frac{0,18(1 - 0,18)}{150}}$$

$$ME = 1,96 \sqrt{\frac{0,24}{140} + \frac{0,1476}{150}}$$

$$ME \approx 1,96(0,0415)$$

$$ME \approx 1,96(0,0415)$$

$$ME \approx 0,08134$$

Ahora que tenemos el margen de error (ME), podemos calcular el intervalo de confianza para la diferencia de proporciones:

Límite inferior (LI):

$$LI = x_1 - x_2 - ME$$

$$LI = 0,22 - 0,08134$$

$$LI \approx 0,13866$$

Límite superior (LS):

$$LS = x_1 - x_2 + ME$$

$$LS = 0,22 + 0,08134$$

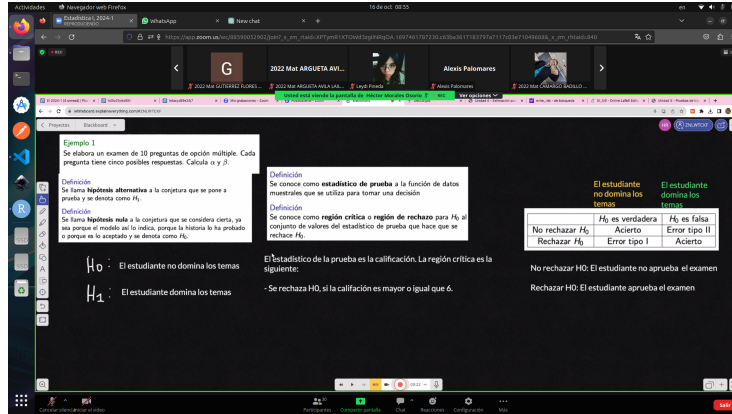
$$LS \approx 0,30134$$

Por lo tanto, el intervalo de confianza del 95 % para la diferencia de las proporciones reales de hombres y mujeres que usan lentes en la población es:

$$(0,13866, 0,30134)$$

Esto significa que, con un 95 % de confianza, la diferencia real entre las proporciones de hombres y mujeres que usan lentes se encuentra en el rango de 13.866 % a 30.134 %.

**Ejemplo 1** Se elabora un examen de 10 preguntas de opcion multiple. Cada pregunta tiene cinco posibles respuestas. Calcula  $\alpha$  y  $\beta$



## 21. 20 de octubre del 2023

### Ejemplo 3

Considera una muestra aleatoria de una normal. Usa el lema de Neyman-Pearson y encuentra la mejor región crítica para la prueba de hipótesis:

$$H_0 : \mu = \mu_0 \text{ contra } H_1 : \mu = \mu_1$$

### Lema de Neyman-Pearson:

Para la prueba de hipótesis:

$$H_0 : \theta = \theta_0 \text{ contra } H_1 : \theta = \theta_1$$

La mejor región crítica de tamaño  $\alpha$  es de la forma:

$$C = \left\{ (x_1, \dots, x_n) \mid \frac{L(\theta_0|x)}{L(\theta_1|x)} < k \right\}$$

con  $k$  una constante positiva.

Si  $X_i \sim N(\mu, \sigma^2)$

$$L(\mu_0, \sigma^2|x) = \prod_{i=1}^n f(x_i|M_0, \sigma^2)$$

$$L(\mu_0, \sigma^2|x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu_0)^2}{2\sigma^2}}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1-\mu_0)^2}{2\sigma^2}} \cdots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n-\mu_0)^2}{2\sigma^2}} \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \cdots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1-\mu_0)^2}{2\sigma^2}} \cdots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n-\mu_0)^2}{2\sigma^2}} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum (x_i-\mu_1)^2}{2\sigma^2}} \\
L(\mu_0, \sigma^2|x) &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^n (x_i-\mu_0)^2}{2\sigma^2}} \\
L(\mu_1, \sigma^2|x) &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^n (x_i-\mu_1)^2}{2\sigma^2}} \\
&= e^{-\frac{\sum (x_i-\mu_0)^2}{2\sigma^2}} + \frac{\sum (x_i-\mu_1)^2}{2\sigma^2}
\end{aligned}$$

$$\begin{aligned}
&e\left(\frac{\mu_0 - \mu_1}{\sigma}\right) \sum x_i < k_1 \\
&e(\mu_0 - \mu_1) \sum x_i < \ln(k_1) \\
&\frac{\mu_0 - \mu_1}{\sigma^2} \sum x_i < k_3 \\
&\text{Si } \mu_0 - \mu_1 > 0 \Rightarrow \mu_0 > \mu_1 \\
P(\text{error tipo 1}) &= P(\bar{X} \leq c | \mu = \mu_0) \\
&= P\left(\frac{n(\bar{X} - \mu_1)}{\sigma} \leq \frac{n(c - \mu_1)}{\sigma}\right) \\
&= P(Z_c < q_z(\alpha)) = \alpha \\
&\text{Si } \mu_0 - \mu_1 < 0 \Rightarrow \mu_0 < \mu_1 \\
P(\text{error tipo 1}) &= P(\bar{X} > c | \mu = \mu_1) \\
&= P(Z_c > q_z(1 - \alpha)) = \alpha
\end{aligned}$$

## 22. 23 de Octubre del 2023

### Ejemplo 6

Para evitar la contaminación ambiental producida por el consumo excesivo e innecesario de gasolina, se realiza un estudio para determinar si el motor del modelo nuevo es más rendidor que el del modelo viejo. Para ello, se eligieron 50 autos con este motor y en una prueba se obtuvo un promedio 18.9 km por litro. Además, se tiene la certeza que la desviación estándar es 4 km. Estos datos confirman que el nuevo motor tiene un rendimiento superior a 17.5 km por litro? Usa un nivel de significancia igual a 0.025.

$$\begin{cases} H_0 : m = 17,5 \\ H_1 : m > 17,5 \end{cases}$$

Prueba	Rc $\sigma^2$ conocida	Rc $\sigma^2$ desconocida
$H_1 : \mu > H_0$	$Z_c > Z(1 - \alpha)$	$tc > t(1 - \alpha)$
$H_1 : \mu < H_0$	$Z_c < Z(\alpha)$	$tc < t(\alpha)$
$H_1 : \mu \neq H_0$	$ Z_c  > Z(1 - \alpha/2)$	$ tc  > t(1 - \alpha/2)$

Cuando una variable aleatoria  $X$  se distribuye como una  $N(\mu, \sigma^2)$  y la hipótesis nula es  $H_0 : \mu = \mu_0$  y el lema de Newman-Pearson indica que las dos posibles realizaciones de prueba son:

$$Z_c = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \text{ cuando } \sigma \text{ es conocida, } Z_c \sim N(0, 1)$$

$$t_c = \frac{\sqrt{n}(\bar{X} - \mu_0)}{s} \text{ cuando } \sigma \text{ es desconocida, } t_c \sim t_{n-1}$$

#### Datos proporcionados

- $n = 50$  (número de autos)
- $\bar{X} = 18,9$  (promedio de kilometraje por litro)
- $\sigma = 4$  (desviación estándar)
- $\mu_0 = 17,5$  (valor hipotético para la media bajo  $H_0$ )
- $\alpha = 0,025$  (nivel de significancia)

#### Cálculo del estadístico de prueba

Ya que se conoce la desviación estándar ( $\sigma$ ), utilizaremos el estadístico  $Z$ . La fórmula para calcular el estadístico  $Z$  en este caso es:

$$Z_c = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}$$

Sustituimos los valores:

$$Z_c = \frac{\sqrt{50}(18,9 - 17,5)}{4} = 2,4748$$

## 23. 6 de Noviembre del 2023

#### Pendientes

- Cuarto parcial ( + cuestionario ) 17 de noviembre
- Tarea 4 en equipos 16 de noviembre
- Cuestionario ( Muestreo ) 22 de noviembre
- Práctica en SAS ( Muestreo ) 22 de noviembre

#### Solucion del examen

$$Y_n = \sum_N^{i=1} x_i^2$$

$$x_i \sim N(0, 1)$$

a)

$$Y_n \sim X_2^{(n)}$$

$$M_{y_n}(t) = (1 - 2t)^{\frac{-n}{2}}$$

$$\begin{aligned} E(Y_n) &= \frac{dM_{Y_n}(t)}{dt} \Big|_{t=0} \\ &= \frac{d(1 - 2t)^{\frac{-n}{2}}}{dt} \Big|_{t=0} \\ &= -\frac{n}{2}(1 - 2t)^{-\frac{n}{2}-1}(-2) \Big|_{t=0} \\ &= n(1 - 2t)^{-\frac{n}{2}-1} \Big|_{t=0} \\ &= n \end{aligned}$$

$$Var(Y_n) = E(y_n^2) - E^2(Y_n) = E(Y_n^2) - n^2$$

$$\begin{aligned} E(Y_n^2) &= \frac{dM_{Y_n}t}{d^2t} \Big|_{t=0} \\ &= \frac{dn(1 - 2t)^{-\frac{n}{2}-1}}{dt} \Big|_{t=0} \\ &= n\left(\frac{n}{2} + 1\right)(1 - 2t)^{-\frac{n}{2}-2} \Big|_{t=0} \\ &= n^2 + 2n \end{aligned}$$

$$\begin{aligned} Var(Y_n) &= E(Y_n^2) - E^2(Y_n) \\ &= n^2 + 2n - 2n^2 \\ &= 2n \end{aligned}$$

b)

$$\mathbb{P}(Y_{100} > 100) = \mathbb{P}\left(\frac{(Y_{100}) - E(Y_{100})}{\sqrt{Var(Y_{100})}} > \frac{100 - E(Y_{100})}{\sqrt{var(Y_{100})}}\right)P = (z > \frac{100 - 100}{\sqrt{200}})P = (z > 0) = 0,5$$

3)



$$f(x) = \begin{cases} 18/38 & \text{si } x = 1 \\ 20/38 & \text{si } x = -1 \end{cases}$$

$$Y_n = \sum_{i=1}^n x_i$$

Imagen del 6 de noviembre 08:44

$$\begin{aligned} E(Y_{100}) &= 100E(x_i) \\ &= 100(-2/38) \\ \text{var}(y_{100}) &= 100(1440/14444) \end{aligned}$$

## 24. 10 de Noviembre del 2023

**Ejemplo 1**

$$\text{Var}(\hat{\theta}) = |0,4|$$

**Ejemplo 3**

$$(69,3275,1306725)$$

**Ejemplo 4**

$$n = \frac{s^2 z_{1-\frac{\alpha}{2}}^2}{\epsilon^2 + \frac{s^2}{N} z_{1-\frac{\alpha}{2}}^2}$$

### 24.1. Muestreo aleatorio estratificado

Cuando la población de tamaño  $n$  es heterogénea, se divide en  $H$  subpoblaciones que se llamarían estratos.

En cada estrato, los elementos son homogéneos entre sí. En este caso, para cada estrato se utiliza  $N_h$  para representar el número de elementos de cada estrato, por lo tanto:

$$N = N_1 + \dots + N_H$$

Se extrae una muestra de cada estrato  $h$ , donde  $x_{ih}$  es el valor  $i$ -ésimo del estrato  $h$  de la variable de interés  $X$ .

**Notacion**

$$w_h = \frac{N_h}{N} \text{ proporción de elementos del estrato } h$$

$$s_h^2 = \text{varianza del estrato } h$$

$\bar{X}_h$  = promedio del estrato h

$$w_1 + \dots + w_h = \frac{N_1}{N} + \frac{N_2}{N} + \dots + \frac{N_H}{N}$$

1. Promedio

$$\bar{X}_s = \sum_{h=1}^H W_h \cdot \bar{X}_h$$

2. Proporción:

$$\bar{P}_s = \sum_{h=1}^H W_h \cdot P_h$$

3. Total:

$$T_s = \sum_{h=1}^H N_h \cdot \bar{X}_h$$

4. Para la varianza:

$$Var(\hat{\theta}) = \sum_{h=1}^H W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}$$

Mejor pegar foto del 10 de Noviembre

### **Ejemplo 5**

Para la siguiente muestra aleatoria estratificada:

Estrato	$N_h$	$W_h$	$\bar{X}_h$	$S_h^2$
A	98	0.249	240	6826.6
B	98	0.249	507.4	23749
C	98	0.249	861.7	42453.0
D	99	0.253	1716.1	152099.6

Calcula  $\bar{X}_s$  y  $Var(\bar{X}_s)$ , donde  $n_h = 10$  para cada estrato.

### **Ejemplo 5 - Resolución paso a paso**

$$\bar{X}_s = \sum_{h=1}^H W_h \bar{X}_h$$

Entonces

$$\bar{X}_s = (0,249 \times 240) + (0,249 \times 507,4) + (0,249 \times 861,7) + (0,253 \times 1716,1)$$

Calculando los productos y sumándolos, obtenemos:

$$\bar{X}_s = (59,76) + (126,3426) + (214,7623) + (434,1733) = 834,8392$$

Para la varianza de  $\bar{X}_s$ :

$$Var(\bar{X}_s) = \sum_{h=1}^H \left( \frac{W_h^2 S_h^2}{n_h} \right)$$

Sustituyendo los valores tenemos:

$$Var(\bar{X}_s) = \left( \frac{0,249^2 \times 6826,6}{10} \right) + \left( \frac{0,249^2 \times 23749}{10} \right) + \left( \frac{0,249^2 \times 42453,0}{10} \right) + \left( \frac{0,253^2 \times 152099,6}{10} \right)$$

Y calculamos cada término para obtener:

$$Var(\bar{X}_s) = (4,0840254) + (14,7430254) + (26,3530254) + (96,578876) = 142,6,358952$$

## 25. hh

El problema 1 trata sobre estadísticas e hipótesis de prueba con respecto a la diferencia en la estatura promedio entre mujeres y hombres. Vamos a resolverlo paso a paso:

Parte (a) Se nos pide construir un intervalo de confianza del 95

Para esto, utilizamos la fórmula del intervalo de confianza para la diferencia de dos medias, cuando las varianzas son conocidas y las muestras son grandes (mayor a 30), lo cual nos permite aplicar el teorema del límite central y usar la distribución Z.

La fórmula es:

$$IC = (\bar{X} - \bar{Y}) \pm Z_{\alpha/2} \cdot \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

Donde: -  $\bar{X} = 169$  cm y  $\bar{Y} = 175$  cm son las medias muestrales. -  $\sigma_X = \sigma_Y = 5$  son las desviaciones estándar poblacionales. -  $n_X = n_Y = 100$  son los tamaños de las muestras. -  $Z_{\alpha/2}$  es el valor crítico de Z para un intervalo de confianza del 95

Vamos a calcular el intervalo de confianza utilizando estas fórmulas.

Parte (b) Luego, debemos concluir si tenemos suficiente evidencia para decir que la estatura promedio de las mujeres es menor que la de los hombres. Para esto, evaluaremos si el intervalo de confianza incluye el 0 o no. Si no incluye el 0, entonces hay suficiente evidencia para afirmar que existe una diferencia significativa en las medias poblacionales.

Voy a calcular esto ahora.

Parece que hay un problema técnico con la herramienta de código en este momento, así que no puedo calcular el intervalo de confianza en este instante. Sin embargo, puedo mostrarte cómo realizar los cálculos manualmente.

Para calcular el intervalo de confianza para la diferencia de medias con un nivel de confianza del 95

$$IC = (\bar{X} - \bar{Y}) \pm Z_{\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n}}$$

donde: -  $\bar{X} = 169$  cm y  $\bar{Y} = 175$  cm son las medias muestrales. -  $\sigma = 5$  es la desviación estándar (igual para ambos grupos). -  $n = 100$  es el tamaño de la muestra (igual para ambos grupos). -  $Z_{\alpha/2}$  es el valor crítico de Z para un intervalo de confianza del 95

Primero calcula el error estándar de la diferencia de medias:

$$SE = \sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{n}} = \sqrt{2 \cdot \frac{25}{100}} = \sqrt{\frac{50}{100}} = \sqrt{0,5}$$

Luego, encuentra el margen de error:

$$ME = Z_{\alpha/2} \cdot SE = 1,96 \cdot \sqrt{0,5}$$

Finalmente, calcula el intervalo de confianza:

$$IC = (169 - 175) \pm ME$$

Esto te dará el intervalo de confianza inferior y superior. Si este intervalo no incluye el 0, entonces hay evidencia estadística suficiente para afirmar que la estatura promedio de las mujeres es significativamente menor que la de los hombres.

## 26. ss

Para resolver el problema 1 del ejercicio, necesitamos aplicar conocimientos de estadística para la estimación y prueba de hipótesis.

**\*\*1(a) Construcción de un intervalo de confianza del 95**

La fórmula para construir un intervalo de confianza para la diferencia de dos medias, cuando las varianzas son conocidas y las muestras son grandes, es la siguiente:

$$IC = (\bar{x} - \bar{y}) \pm Z_{\alpha/2} \cdot \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

Donde: -  $\bar{x}$  y  $\bar{y}$  son las medias muestrales. -  $\sigma_X^2$  y  $\sigma_Y^2$  son las varianzas poblacionales, ambas son 5 en este caso. -  $n_X$  y  $n_Y$  son los tamaños de las muestras, ambos 100 en este caso. -  $Z_{\alpha/2}$  es el valor crítico de la distribución normal estándar para un nivel de confianza del 95

**\*\*1(b) Prueba de hipótesis para la diferencia de medias\*\***

Para probar la hipótesis nula de que no hay diferencia en las medias poblacionales ( $H_0 : \mu_X - \mu_Y = 0$ ) contra la alternativa de que la media de las mujeres es menor que la de los hombres ( $H_1 : \mu_X - \mu_Y < 0$ ), se utiliza un valor Z de prueba:

$$Z = \frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}}$$

Luego se compara el valor Z de prueba con el valor crítico Z para un nivel de significancia dado (usualmente 0.05 para pruebas de dos colas).

Vamos a calcular estas dos partes del ejercicio utilizando Python.

Parece que hubo un problema técnico al intentar realizar los cálculos. Volveré a intentarlo para proporcionarte los resultados que necesitas. Un momento, por favor.

Lo siento, pero parece que hay un problema técnico persistente que me impide realizar los cálculos en este momento. Sin embargo, puedo guiarte a través de los pasos para que puedas calcularlo manualmente.

Para el \*\*1(a)\*\*, el intervalo de confianza del 95

$$IC = (\bar{x} - \bar{y}) \pm Z_{\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n_X} + \frac{\sigma^2}{n_Y}}$$

Donde: -  $\bar{x} = 169$  cm y  $\bar{y} = 175$  cm son las medias muestrales. -  $\sigma = 5$  cm es la desviación estándar conocida para ambas poblaciones. -  $n_X = n_Y = 100$  es el tamaño de cada muestra. -  $Z_{\alpha/2}$  es el valor crítico de la distribución normal estándar, que es aproximadamente 1.96 para un intervalo de confianza del 95

Sustituye los valores en la fórmula para encontrar el intervalo de confianza.

Para el \*\*1(b)\*\*, la prueba de hipótesis para la diferencia de medias se realiza comparando el valor Z calculado con el valor crítico Z para el nivel de significancia seleccionado. Si el valor Z calculado es menor que el valor crítico Z negativo, rechazamos la hipótesis nula.

El valor Z se calcula como:

$$Z = \frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma^2}{n_X} + \frac{\sigma^2}{n_Y}}}$$

En este caso, la hipótesis nula es que  $\mu_X - \mu_Y = 0$ , y dado que estamos buscando si la media de las mujeres es significativamente menor que la de los hombres, sería una prueba de una cola. El valor crítico para una prueba de una cola con un nivel de significancia de 0.05 es aproximadamente -1.645.

Calcula el valor Z usando las medias y la desviación estándar proporcionadas y compáralo con el valor crítico para determinar si rechazas o no la hipótesis nula.

Espero que estas instrucciones sean útiles para resolver el problema manualmente. Si tienes una calculadora estadística o acceso a un software como R o Python, puedes usar estos valores para encontrar los resultados necesarios.

It seems there was an internal error while trying to perform the calculation. Let me try that again.

I apologize for the inconvenience, but it seems there is a persistent technical issue preventing the calculation from being completed. However, I can guide you through the process of solving this problem.

To construct a 95

1. Calculate the standard error of the difference between the two means:

$$SE = \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

Where: -  $\sigma_X$  and  $\sigma_Y$  are the population standard deviations, and -  $n_X$  and  $n_Y$  are the sample sizes.

2. Find the z-score corresponding to a 95
3. The margin of error (ME) is then calculated as:

$$ME = z \times SE$$

4. The confidence interval is calculated as:

$$(\bar{X} - \bar{Y}) \pm ME$$

Where: -  $\bar{X}$  and  $\bar{Y}$  are the sample means.

5. With the given values ( $\bar{X} = 169$  cm,  $\bar{Y} = 175$  cm,  $\sigma_X = \sigma_Y = 5$ ,  $n_X = n_Y = 100$ ), you would plug these into the formulas to get your interval.

To conclude whether there is sufficient evidence to say that the average height of women is less than that of men, you would check if the confidence interval includes zero. If zero is not within the interval, it suggests that there is a significant difference between the two means.

You can perform these calculations using a scientific calculator or statistical software. If zero is not included in your interval, you can conclude that there is evidence to suggest that the average height of women is less than that of men at the 95

## Problema de Estadística

### 1(a) Intervalo de confianza del 95 % para la diferencia de medias

Dado que se conocen las varianzas y estas son iguales para hombres y mujeres, y que las muestras son grandes, se puede usar la distribución normal para el intervalo de confianza. La fórmula del intervalo de confianza para la diferencia de medias es:

$$IC = (\bar{x} - \bar{y}) \pm Z_{\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{n}}$$

Donde:

- $\bar{x} = 169$  cm y  $\bar{y} = 175$  cm son las medias muestrales.
- $\sigma = 5$  cm es la desviación estándar conocida para ambas poblaciones.
- $n = 100$  es el tamaño de las muestras.
- $Z_{\alpha/2}$  es el valor crítico para un intervalo de confianza del 95 %, que es aproximadamente 1.96.

Sustituimos los valores en la fórmula:

$$IC = (169 - 175) \pm 1,96 \cdot \sqrt{\frac{5^2}{100} + \frac{5^2}{100}}$$

Calculamos el margen de error y los límites del intervalo:

$$ME = 1,96 \cdot \sqrt{\frac{25}{100} + \frac{25}{100}}$$
$$IC = -6 \pm ME$$

### 1(b) Prueba de hipótesis para la diferencia de medias

Para la prueba de hipótesis utilizamos la fórmula del valor Z de prueba:

$$Z = \frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{n}}}$$

Bajo la hipótesis nula, asumimos que  $\mu_X - \mu_Y = 0$ . Entonces:

$$Z = \frac{(169 - 175)}{\sqrt{\frac{25}{100} + \frac{25}{100}}}$$

Calculamos el valor de Z y lo comparamos con el valor crítico  $Z_\alpha$  para un nivel de significancia de 0.05 en una prueba de una cola, que es aproximadamente -1.645.

$$Z = \frac{-6}{ME}$$

Si  $Z \leq -1,645$ , entonces rechazamos la hipótesis nula, indicando que hay suficiente evidencia para afirmar que la estatura promedio de las mujeres es significativamente menor que la de los hombres.



## Problema 2

Tenemos una muestra aleatoria observada de tamaño uno,  $X_1$ , de una distribución normal  $N(\mu, \sigma^2 = 1)$ . Queremos probar la hipótesis nula  $H_0 : \mu = 0$  contra la hipótesis alternativa  $H_1 : \mu = 2$ . La región de rechazo es  $X_1 \geq 1,5$ .

### (a) Probabilidad de cometer el error tipo I

El error tipo I se da cuando rechazamos la hipótesis nula siendo esta verdadera. En términos de nuestro problema, esto sucede si  $X_1$  es mayor o igual a 1.5 cuando  $\mu = 0$ . La probabilidad de cometer el error tipo I es el nivel de significancia  $\alpha$ , que se calcula como:

$$\alpha = P(X_1 \geq 1,5 | \mu = 0) = 1 - P(X_1 < 1,5 | \mu = 0)$$

Dado que  $X_1$  se distribuye normalmente, convertimos este valor a un puntaje  $Z$  y buscamos en la tabla de la distribución normal estándar o usamos una función de distribución acumulativa.

$$Z = \frac{1,5 - 0}{1} = 1,5$$
$$\alpha = 1 - \Phi(1,5)$$

Donde  $\Phi$  es la función de distribución acumulativa de una distribución normal estándar.

### (b) Probabilidad de cometer el error tipo II

El error tipo II ocurre cuando no rechazamos la hipótesis nula y esta es falsa. Para calcular esta probabilidad, necesitamos considerar que  $\mu = 2$  y calcular la probabilidad de que  $X_1$  sea menor que 1.5.

$$\beta = P(X_1 < 1,5 | \mu = 2) = \Phi\left(\frac{1,5 - 2}{1}\right)$$

### (c) Potencia de la prueba

La potencia de la prueba es la probabilidad de rechazar correctamente la hipótesis nula cuando la hipótesis alternativa es verdadera. Se calcula como  $1 - \beta$ .

$$\text{Potencia} = 1 - \beta = 1 - \Phi\left(\frac{1,5 - 2}{1}\right)$$

## 27. Reapso 22 de noviembre

Tipos de muestras

- Muestreo probabilistico
  - Se puede calcular  $\epsilon = |\hat{\theta} - \theta|$
  - Cada elemento de la poblacion tiene una probabilidad conocida de ser seleccionada
- Muestreo no probabilistico
  - No se puede calcular  $\epsilon$
  -
- Errores de muestreo
  - Se controlan con el diseno de muestreo
- Errores no de muestreo
  - Se controla con el diseno de la encuesta

## 28. coursera

1. **¿Cuál de las siguientes es una pregunta clave a la que intentan responder las retrospectivas de los sprints de diseño?**
  - ¿Cuáles son nuestros objetivos?
  - ¿Cuál es nuestra próxima tarea?
  - ¿Se realizó el sprint a la perfección?
  - ¿Qué salió bien o qué se puede mejorar? [✓]
2. **Completa el espacio en blanco: En una retrospectiva de sprint de diseño, el equipo trabaja en conjunto para revisar y criticar lo que sucedió en el sprint que -----.**
  - el equipo acaba de terminar [✓]
  - el equipo finalizará al final del proyecto
  - el equipo está completando actualmente
  - el equipo está a punto de empezar