Universidad Nacional Autónoma de México

Facultad de Estudios Superiores Acatlán

Apuntes
Estadística

Autor:

Jorge Miguel Alvarado Reyes

20 de septiembre de 2023

Índice

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. $Media(cx) = c \cdot Media(x)$
2. $Media(x + c) = Media(x) + c$

Ejemplo 4:
Mostrar que $Media(x + c) = Media(x) + c$.
Demostración:

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

Ejemplo 5:
Mostrar que $Media(cx) = c \cdot \text{Media}(x)$.
La **media ponderada** se define como:

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

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:

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

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 $f_1, ..., f_n$ 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 $f_1, ..., f_n$ las frecuencias de la varible x y $c_1, ..., c_n$ 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

$$Mediana(x) = L_i t(\frac{\frac{n}{2} - 3\sum f_i}{f_{mediana}}) \cdot c$$

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	Codificacion 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 linea

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

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

3. 21 Agosto 2023

3.1. Medidas de posición (1.6.4)

3.1.1. Cuantiles

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

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

Casos particulares

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

Ejemplo 10
Calcula los <u>deciles</u> de la variable $x = \{0{,}1\}$
Supongamos:
Si c=0

- $\frac{\#\{x_i|x_i \leq 0\}}{n} \geq p$
  $\frac{\#\{0\}}{2} \geq p$
  $\frac{1}{2} \geq p$
  $p = \{0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5\}$
- $\frac{\#\{x_i|x_i \geq 0\}}{n} \geq 1 - p$
  $\frac{\#\{0{,}1\}}{2} \geq 1 - p$
  $1 \geq 1 - p$
  $p = \{0.1,\ 0.2,\ 0.3,\ \ldots,\ 0.8,\ 0.9\}$

Si c=1

- $\frac{\#\{x_i|x_i \leq 1\}}{n} \geq p$
  $\frac{\#\{0{,}1\}}{2} \geq p$
  $1 \geq p$
  $p = \{0.5,\ 0.6,\ 0.7,\ 0.8,\ 0.9\}$??? para todos no?
- $\frac{\#\{x_i|x_i \geq 1\}}{n} \geq 1 - p$
  $\frac{\#\{1\}}{2} \geq p$
  $\frac{1}{2} \geq 1 - p$
  $p = \{0.5,\ 0.6,\ 0.7,\ 0.8,\ 0.9\}$

- 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

- $$\frac{\#\{x_i|x_i \leq 2\}}{n} \geq p$$
  $$\frac{\#\{0,1,2\}}{4} \geq p$$
  $$\frac{3}{4} \geq p$$
  $$p = \{0.75\}$$

- $\frac{\#\{x_i | x_i \geq 2\}}{n} \geq 1 - p$

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

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

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

Si c=3

- $\frac{\#\{x_i | x_i \leq 3\}}{n} \geq p$

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

  $1 \geq p$

  $p = \{0.25,\ 0.5,\ 0.75\}$

- $\frac{\#\{x_i | x_i \geq 3\}}{n} \geq 1 - p$

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

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

  $p = \{0.75\}$

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

3.1.2. Varianza

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

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

$$De(x) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_1 - \bar{x})^2}$$

Propiedades

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

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

3.1.4. Desviacion media

Sean $x_1, \ldots, x_n$ 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 $x_1, \ldots, x_n$ 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 $var(x + c) = var(x)$

3.2. Medidas de forma (1.6.5)

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

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

Mientras que el k-esimo momento central de x es

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

3.2.1. Asimetria

El coeficiente de asimetria mide la asimetria de los datos respecto a la media. Sean $x_1, \ldots, x_n$ observaciones de la variable x. El coeficiente de asimetria de x es

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

Donde s es la desviacion estandar de x.

4. 23 agosto 2023

4.1. Medidas de forma (1.6.5) (continuacion)

4.1.1. Coeficiente de asimetria

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

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:

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

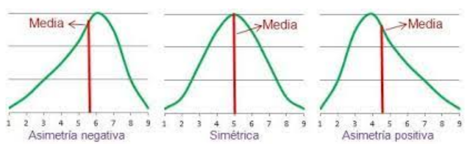


Figura 1: negativa = $sk(x) < 0$, Simetrica = $sk(x) = 0$, positiva = $sk(x) > 0$

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.

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

4.1.2. Curtosis

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

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

4.1.3. Coeficiente de Fisher

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

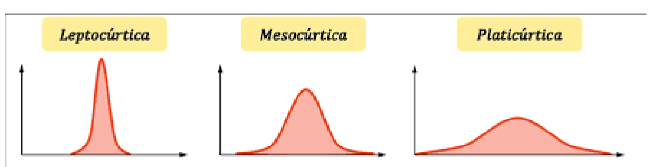
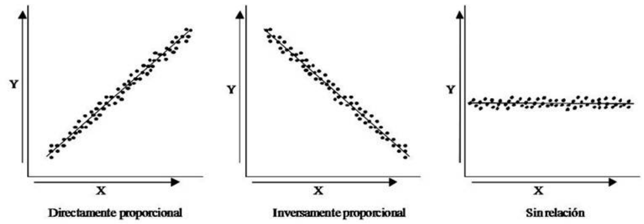


Figura 2:

4.2. Medidas de asociacion (1.6.6)



4.2.1. Covarianza

Sean $(x_1, y_1) \ldots, (x_n, y_n)$ 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

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

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

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

$$r_x y = \frac{s_{xy}}{s_x s_y}$$

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:

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

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

Ejemplo 16

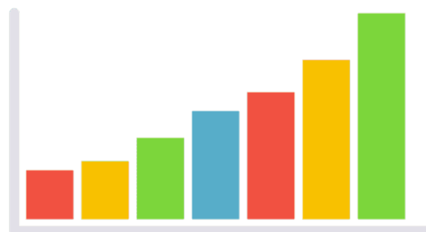
Supongamos que se tiene una variable cualitativa ordinal con valores ordenados de menor a mayor, A,B,C,D,E,F con las siguientes frecuencias,

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

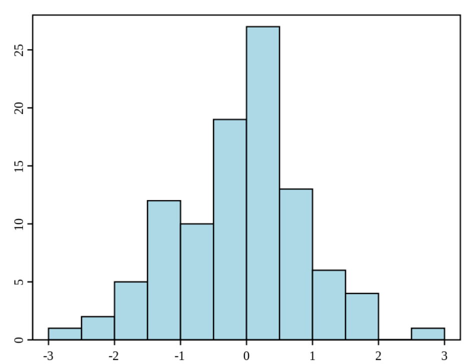
5.1. Graficas de barras

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.



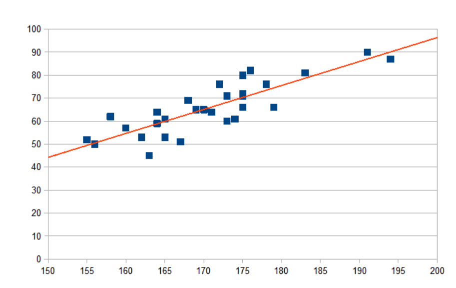
5.2. Histograma

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



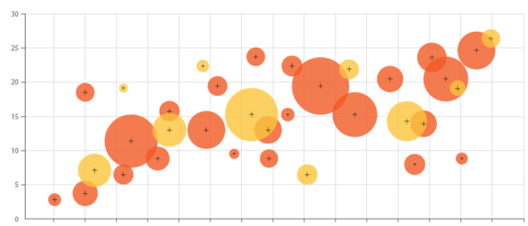
5.3. Gráficos de dispersión

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

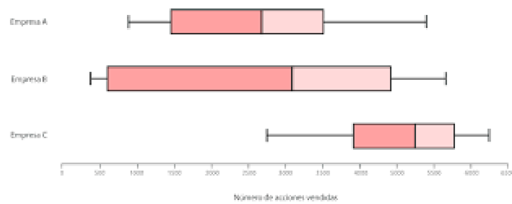
R es un programa útil para el análisis y visualización de datos. Es abierto y gratuito.
Es un lenguaje interpretado y tipado.
Fue creado por Ross lhaka y Robert Gentle-man.
Instalación.
Descargar R: https://cran.r-project.org/
Descargar R studio: https://posit.co/download/rstudio-desktop/

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 espacion. 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 despúes de una coma.
- Ayuda, se puede usar el comando help(mean) o ?mean
- Tipos de datos, enteros, númericos 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 (+)
- revista (-)
- 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 ¡- list(1,"Manzana")

6.2.6. Matrices

- $matriz_1 < -matrix(1:10, nrow = 2, ncol = 5)$

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 (Ω, F, P). Una variable aleatoria es una transformación X del espacio de resultados Ω 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 $x_0, x_1, \ldots$ y probabilidades respectivas $\mathbb{P}(X = x_0), \mathbb{P}(X = x_1), \ldots$ la función masa de probabilidad de x denotada por $f(x) : \mathbb{R} \to [0, \infty)$ se define como sigue:

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

Y cumple con las siguentes 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 $F(x) \to [0, 1]$ 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 $f(x) : \mathbb{R} \to [0, \infty)$ 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:

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

Definición

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

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

Definición

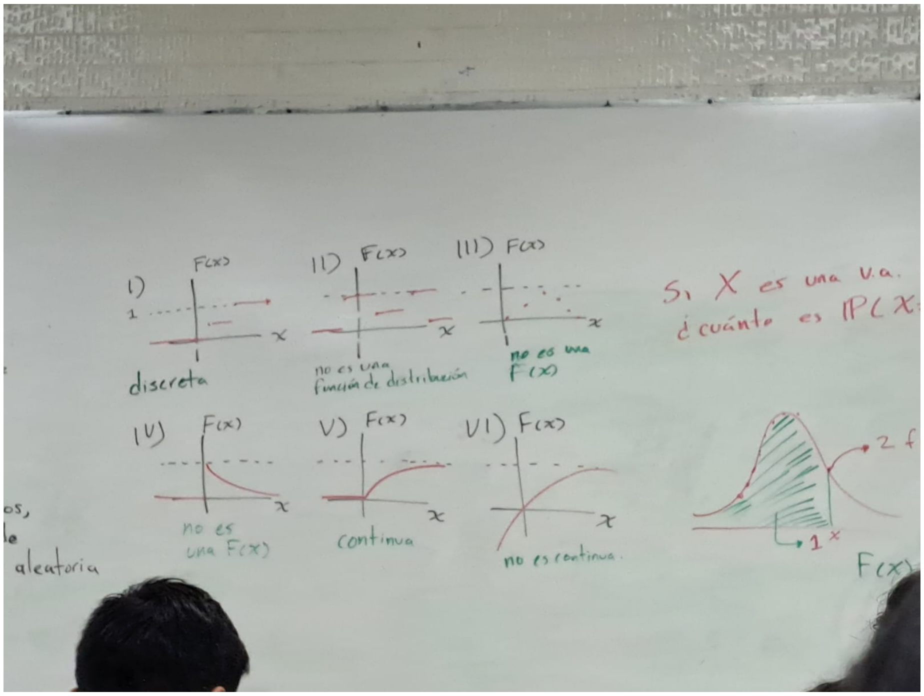
La variable aleatoria X se llama continua si su correspondiente función de distribución es continua y creciente

Definición

La variable aleatoria X se llama discrtea 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 φ una función de X, entonces la variable aleatoria $Y = \varphi(x)$ con funciónde distribución $F_y(y) = \mathbb{P}(Y \leq y)$, $f(y)$ se puede calcular al integrar la región para la cual $Y \leq y$. Ademas, 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 $Y = x_1 - X_2$

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

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 $\varphi(a, b) \to \mathbb{R}$ 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

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

Ejemplo 4
Sea $x \backsim Unif(0, 1)$ 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.

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

Definición

La varianza de la variable aleatoria X se define como:

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

Definición

Sea X una variable aleatoria continua con media μ, 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:

$$M_x(t) = E(e^t x)$$

Ejemplo 6

Insertar las fotos del ejercicio 6 y 7

Ejemplo 8

$$-M_x(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n)$$

$$-\frac{d^n[M_x(t)]}{dt^n}|_{t=0} = E(x^n)$$

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 $X \backsim N(\mu, G^2)$, entonces $Y = \frac{X-\mu}{G} \backsim N(0,1)$
Si $N \backsim (\mu G^2)$, entonces $M_x(t) = e(\mu t + \frac{G^2 t^2}{2})$
Si $Y \backsim N(0,1)$
$M_y(t) = e^{\frac{t^2}{2}}$
De manera que

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

11. 18 de Septiembre del 2023

Ejemplo 11
Sea $X \backsim Exp(1/100)$. Encuentra $f_{x(1)}(x)$ para una muestra de tamaño 2.

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

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

11.1. Distribuciones derivadas de la normal

11.1.1. Normal estándar

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

- La media, la mediana y la moda son iguales a μ
- 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 $Z_1, \cdots, Z_k$ variables aleatorias con distribución normal estándar

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

Entonces, $X \backsim X_k^2$ 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 $X_1, \cdots, X_n$ 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 $g(X_1, \cdots, X_n)$ en donde $X_1, \cdots, X_n$ es una muestra aleatoria y $g : \mathbb{R}^n \to \mathbb{R}$
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 párametro es una caracterización numérica de la población de manera que describe parcialmente la función de la probabilidad.

población	Muestra
N	n
μ	$\bar{x}$
σ^2	s^2

Proposición

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

$$y = a_1X_1 + \cdots + a_nX_n$$

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

$$M_y(t) = M_{x1}(a_1t) \cdots M_{xn}(a_nt)$$

Ejemplo 1

Sea $X_1, \cdots, X_n$ una muestra aleatoria con distribución normal donde $E(x_i) = \mu_i$ y $Var(X_i) = \sigma_i^2$, $i = 1, \cdots, n$. Si $Y = a_1X_1 + \cdots + a_nX_n$, muestra que Y se distribuye como una normal.

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

entonces

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

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