\mainmatter

(PART) Estadística descriptiva

Prerrequisitos

wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww

Variables

Es una **característica** de personas cosas u objetos que son propensos a ser medidas o cualificadas

Variables cualitativas

Denotan cualidades de objetos personas o animales tales como características inherentes que *no son medibles por números*, tenemos dos casos de esta variable.

Nominales

Son caracteristicas que simplemente nominan y están propensos a ser jerarquizados u ordenados tales como: El estado civil (soltero, casado, divorciado, viudo), Religión (católica, evangélico, judío, etc).

Ordinales

Son caracteristicas que que si están propensos a ser jerarquizados tales como: Nivel de instrucción (inicial, primaria, secundaria, superior).

Variables cuantitativas

Son aquellas variables que están propensos a ser medidas mediante números ya sean números enteros o reales.

Discretas

Aquellas que solo son medidos mediante numeros enteros por ejemplo: Número de hijos, número de habitaciones.

Continuas

Aquellas que solo son medidos mediante numeros reales es decir este incluye a los numeros racionales e irracionales. Estatura, volumen, peso.

Asignación

1. Reconosca **5** variables **cualitativas** de una persona, admosfera, una pintura
    - Persona
        - Color (moreno, blanco, trigueo)
        - Religión (catolico, evangelico, pentecostal, etc)
        - wwwwwwww
        - wwww
        - wwwwww

    - Admosfera
        - www
        - wwwww
        - wwwwwww
        - wwww
        - wwwwww
    - Pintura
        - www
        - wwwww
        - wwwwwww
        - wwww
        - wwwwww
2. Reconosca **5** variables **cuantitativas** de una video, tela, un celular.
    - Video
        - Duracion (x segundos)
        - Numero video en youtube a la semana (n cantidades)
        - wwwwwww
        - wwww
        - wwwwww
    - Tela
        - www
        - wwwww
        - wwwwwww
        - wwww
        - wwwwww
    - Celular
        - www
        - wwwww
        - wwwwwww
        - wwww
        - wwwwww

Organización de datos en tablas de frecuencias

Distribución de frecuencias

La tabulación es un proceso en el cual los datos son ordenados en grupos llamados *clases* para un análisis más eficaz de estos, los datos podrían estar clasificados mediante una variable cualitativa o cuantitativa en el caso de las variables cualitativas Y_i, se considera la siguiente Tabla @ref(tab:ww)

Y_i	f_i	F_i	F_i^*	h_i	H_i	H_i^*	$h_i\%$	$H_i\%$	$H_i^*\%$
Y_1	f_1	F_1	F_1^*	$\frac{f_1}{n}$	$\frac{F_1}{n}$	$\frac{F_1^*}{n}$	h_1	H_1	H_1^*
Y_2	f_2	F_2	F_2^*	$\frac{f_2}{n}$	$\frac{F_2}{n}$	$\frac{F_2^*}{n}$	h_2	H_2	H_1^*
Y_3	f_3	F_3	F_3^*	$\frac{f_3}{n}$	$\frac{F_3}{n}$	$\frac{F_3^*}{n}$	h_3	H_3	H_1^*
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Y_r	f_r	F_r	F_r^*	$\frac{f_r}{n}$	$\frac{F_r}{n}$	$\frac{F_r^*}{n}$	h_r	H_r	H_1^*

: (#tab:ww) Caption

En el caso de variables cuantitativas ademas si los datos son muy variados, que para se clasificados adecuadamente, necesitan generarse particiones de longitudes semejantes entonces se utiliza el siguiente proceso; el **número de las particiones** r se consideran de acuerdo a **tres criterios**

1. Criterio del investigador r no puede ser más de 20 ni menos de 5
2. $r = \sqrt{n}$ donde n es el número de datos
3. La regla de Starges que consiste en considerar la fórmula $r = 3.322 \cdot \log_{10} n$ Una vez establecido el número de particiones se procede a generar los límites laterales de cada una de las particiones, sea L la longitud de todo el conjunto es decir $L = x_{\max} - x_{\min}$ entonces la longitud de las particiones o amplitud interválica se obtiene con $l = \frac{L}{r}$

```
library(openxlsx)
opts <- options(knitr.kable.NA = "",ggrepel.max.overlaps = Inf)
new <-read.xlsx(xlsxFile="levene.xlsx", sheet=2, cols=c(1:11), rows=c(1:5), colNames=T)
knitr::kable(new, escape = FALSE, digits = 2, booktabs=T, caption='Datos cuantitativos (intervalos)', linesep = "", longtable=T, align = "c")
```

Tenga en cuenta que n es el número de datos, es decir $n = f_1 + f_2 + \ldots + f_r = \sum_{i=1}^{r}$ donde f_i es número de datos en la partición X_i, una de las r particiones del conjunto total de datos.

1. Las **frecuencias absolutas**\index{frecuencias absolutas} f_i indican el número de datos con la característica X_i.
2. Las **frecuencias absolutas acumuladas menor que**\index{frecuencias absolutas acumuladas menor que} F_i obedecen a la fórmula

$$F_m = f_1 + f_2 + \ldots + f_m = \sum_{i=1}^{m} f_i$$

3. Las **frecuencias absolutas acumuladas mayor que** F_i^* obedecen a la fórmula

$$\begin{aligned} F_m^* &= f_m + f_{m+1} + \ldots + f_r \\ &= \sum_{i=m}^{r} f_i \\ &= n - \sum_{i=1}^{m-1} f_i \\ &= n - (f_1 + f_2 + \ldots + f_{m-1}) \end{aligned}$$

4. Las **frecuencias absolutas relativas**\index{frecuencias absolutas relativas} obedecen a la fórmula

$$h_m = \frac{f_m}{n}$$

5. Las **frecuencias absolutas relativas menor que**\index{frecuencias absolutas relativas menor que} obedecen a la fórmula

$$H_m = \frac{f_m}{n}$$

6. Las **frecuencias absolutas relativas mayor que** obedecen a la fórmula

$$H_m^* = \frac{F_m}{n}$$

7. Las **frecuencias absolutas relativas porcentuales** obedecen a la fórmula $h_i\% = 100 \cdot h_i$
8. Las **frecuencias absolutas relativas menor que porcentuales** obedecen a la fórmula $H_i\% = 100 \cdot H_i$
9. Las **frecuencias absolutas relativas mayor que porcentuales** obedecen a la fórmula $H_i^*\% = 100 \cdot H_i^*$
10. Y_i marca de clase o punto medio de la clase i

Ejemplo sin intervalos

Sean Los 16 tipos de personalidad en un grupo social encuestado.

1. ESTJ (Extraverted Sensing Thinking Judging)
Personas a las que les gusta tener el control sobre lo que ocurre a su alrededor, siempre buscan la manera de que todo funcione como debe y, si es necesar

2. ESTP ((Extraverted Sensing Thinking Perceiving)
Las personas que pertenecen a esta categoría son espontáneas, alegres y activas, pero al igual que lo que ocurre con los ESTJ, tienden a ejercer dominio s

3. ESFJ (Extraverted Sensing Feeling Judging)
Se trata de personas muy volcadas en la atención de las necesidad de los demás, especialmente si forman parte de su círculo cercano: familia y amistades.

4. ESFP (Extraverted Sensing Feeling Perceiving)
Se trata de personas alegres y espontáneas que disfrutan entreteniéndose y entreteniendo a los demás. La diversión es uno de los pilares más importantes d

5. ISTJ (Introverted Sensing Thinking Perceiving)
Un tipo de personalidad definido por su fuerte sentido de la moralidad y del deber. Les gusta planear e implementar sistemas de reglas que permitan que eq

6. ISTP (Introverted Sensing Thinking Perceiving)
Se trata de personas reservadas, orientadas a la acción y a las soluciones prácticas ante problemas del día a día. También son definidas por su tendencia

7. ISFJ (Introverted Sensing Feeling Judging)
Son personas definidas principalmente por sus ganas de proteger y ayudar a los demás y, en definitiva, de resultar confiables para los otros.Se esfuerzan

8. ISFP (Introverted Sensing Feeling Perceiving)
Personas que viven totalmente en el aquí y el ahora, en constante búsqueda de la novedad y de las situaciones sensorialmente estimulantes. Son reservadas,

9. ENTJ (Extraverted Intuitive Thinking Judging)
Este es uno de los 16 tipos de personalidad más relacionados con el liderazgo y la asertividad. Las personas descritas por esta categoría son comunicativa

10. ENTP (Extraverted Intuitive Thinking Perceiving)
Personas especialmente movidas por la curiosidad y por los retos que para ser resueltos requieren afrontar preguntas intelectualmente estimulantes. Su agi

11. ENFJ (Extraverted Intuitive Feeling Judging)
Personas que aprenden constantemente acerca de todos los ámbitos del conocimiento (o una buena parte de ellas) y ayudan a aprender a las demás, guiándolas

12. ENFP (Extraverted Intuitive Feeling Perceiving)
Uno de los 16 tipos de personalidad con mayor propensión al pensamiento creativo, las artes y a la sociabilidad. Son alegres, disfrutan de la interacción

13. INTJ (Introverted Intuitive Thinking Judging)
Un tipo de personalidad orientado hacia la resolución de problemas específicos a partir del razonamiento analítico. Las descritas por esta categoría son p

Es muy frecuente que lleguen a ser expertas en un ámbito de conocimiento muy específico, ya que les gusta tener el suficiente conocimiento sobre algo como

14. INTP (Introverted Intuitive Thinking Perceiving)
Uno de los 16 tipos de personalidad más definido por la propensión a la reflexión. A estas personas les gustan las teorías con capacidad para explicar tod

15. INFJ (Introverted Intuitive Feeling Judging)
Personas muy sensibles, reservadas y movidas por unos ideales muy definidos y que, además, sienten la necesidad de hacer que los demás también se benefici

16. INFP (Introverted Intuitive Feeling Perceiving)
Menos moralistas que los INFJ, los INFP también se preocupan mucho por ayudar a los demás desde su posición de personas reservadas. Muestran una sensibili

```
Se tiene 16 características con los siguientes datos

| Personalidad | Cantidad |
|:------------:|:--------:|
|     ESTJ     |    1     |
|     ESTJ     |    2     |
|     ESTP     |    3     |
|     ESFJ     |    4     |
|     ESFP     |    5     |
|     ISTJ     |    6     |
|     ISTP     |    7     |
|     ISFJ     |    8     |
|     ISFP     |    9     |
|     ENTJ     |    10    |
|     ENTP     |    6     |
|     ENFJ     |    5     |
|     ENFP     |    3     |
|     INTJ     |    2     |
|     INTP     |    1     |
|     INFJ     |    1     |
|     INFP     |    1     |
```

```
library(openxlsx)
opts <- options(knitr.kable.NA = "",ggrepel.max.overlaps = Inf)
new <-read.xlsx(xlsxFile="levene.xlsx", sheet=1, cols=c(5:14), rows=c(1:19), colNames=T)
knitr::kable(new, escape = FALSE, digits = 2, booktabs=T, caption='Datos cualitativos', linesep = "", longtable=T, align = "c")
```

Example intervalos

```
Edades de cierta comunidad


25
35 38
45 47 48
51 52 53 55
60 62 63 66 67
70 71 72 75 77 78
81 88 89
90 99

|   Clase    | $f_i$ |
|:----------:|:-----:|
| $[20-30>$  |   1   |
| $[30-40>$  |   2   |
| $[40-50>$  |   3   |
| $[50-60>$  |   4   |
| $[60-70>$  |   5   |
| $[70-80>$  |   6   |
| $[80-90>$  |   3   |
| $[90-100]$ |   2   |
```

Tabulando

```
const number = 50//parseInt(prompt('Enter the number of terms: '));
let n1 = 0, n2 = 1, nextTerm;
console.log('Fibonacci Series:');
for (let i = 1; i <= number; i++) {
    console.log(n1);
    nextTerm = n1 + n2;
    n1 = n2;
    n2 = nextTerm;
}
```

```
import matplotlib.pyplot as plt
plt.plot([1,2,3, 4])
plt.show() # show figure
```

```
library(openxlsx)
opts <- options(knitr.kable.NA = "",ggrepel.max.overlaps = Inf)
new <-read.xlsx(xlsxFile="levene.xlsx", sheet=1, cols=c(5:14), rows=c(30:41), colNames=T)
knitr::kable(new, escape = FALSE, digits = 2, booktabs=T, caption='Datos cuantitativos (intervalos)', linesep = "", longtable=T, align = "c")
```

Gráficos estadísticos

Histograma de frecuencias

```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <-c("Automatic", "Manual")
```

Some text

```
hist(mtcars$mpg[mtcars$am=="Automatic"], breaks=12, main="mpg for automatic vehicles", xlab="mpg", xlim=c(10, 35))
```

Circulares

```
par(mar = c(0, 1, 0, 1))
pie(
  c(280, 60, 20),
  c('Sky', 'Sunny side of pyramid', 'Shady side of pyramid'),
  col = c('#0292D8', '#F7EA39', '#C4B632'),
  init.angle = -50, border = NA
)
```

```
plot(cars)
lines(lowess(cars))
```

```
library(ggplot2)

# Create data
data <- data.frame(
  name=c("A","B","C","D","E") ,
  value=c(3,12,5,18,45)
  )

# Barplot
ggplot(data, aes(x=name, y=value)) +
  geom_bar(stat = "identity", width=0.5) +
  coord_flip()
```

Histograma de frecuencias

```
hist(mtcars$mpg[mtcars$am=="Automatic"], breaks=12, main="mpg for automatic vehicles", xlab="mpg", xlim=c(10, 35))
```

Medidas de tendencia central

Son aquellas medidas que buscan un dato representivo central de un conjunto de datos tales como la media, la moda y la mediana.

La media

A veces llamada *promedio aritmético*, es la medida de tendencia central que pondera los datos.

Media de datos no agrupados

Los datos no están agrupados cuando no están ordenados sobre una tabla de distribución de frecuencias. Sean los n datos $x_1, x_2, \ldots, x_n$ entonces la media o promedio aritmético se define como

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

Media de datos agrupados

Considérese la siguiente tabla de distribucion de frecuencias entonces el promedio es

$$\overline{x} = \frac{y_1 f_1 + y_2 f_2 + \cdots + y_n f_n}{n} = \frac{1}{n}\sum_{i=1}^{n} y_i f_i$$

Clase	Y_i	f_i	F_i	$\ldots$	$H_i^*\%$
$[y_1, y_2)$	y_1	f_1	$\ldots$	$\ldots$	$H_1^*\%$
$[y_2, y_3)$	y_2	f_2	$\ldots$	$\ldots$	$H_1^*\%$
$[y_3, y_4)$	y_3	f_3	$\ldots$	$\ldots$	$H_1^*\%$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\ldots$	$\vdots$
$[y_{r-1}, y_r]$	y_r	f_r	$\ldots$	$\ldots$	$H_1^*\%$

: (#tab:www) Captionww

Ejemplo

Clase	Y_i	f_i	$Y_i * f_i$
$[10, 15)$	12.5	1	12.5
$[15, 20)$	17.5	2	35
$[20, 25)$	22.5	5	112.5
$[25, 30)$	27.5	3	82.5
$[30, 35]$	32.5	2	65
$\sum$		13	307.5

$$\overline{x}=\frac{12.5+35++112.5+82.5+65}{13}=\frac{307.5}{13}=23.65$$

```
Si el promedio de $n$ datos es $\overline{x}$ entonces el promedio del conjunto inicial más un dato adicional $x_{n+1}$ es $$\overline{x}'=\frac{n\overlin
```

```
En efecto sea el promedio
\begin{align*}
\overline{x}'&=\frac{x_1+x_2+\cdots+x_{n+1}}{n+1}\\
&=\frac{n\frac{x_1+x_2+\cdots x_n}{n}+x_{n+1}}{n+1}\\
&=\frac{n\overline{x}+x_{n+1}}{n+1}
\end{align*}
```

La moda (Mo)

- La moda es el valor que tiene mayor frecuencia absoluta.
- Se representa por Mo
- Si en un grupo hay dos o varias puntuaciones con la misma frecuencia y esa frecuencia es la máxima, entonces la distribución es bimodal es decir, tiene varias modas.
- Cuando todas las puntuaciones de un grupo tienen la misma frecuencia, no hay moda.
- Se puede hallar la moda para variables cualitativas y cuantitativas.
- Cuando todas las puntuaciones de un grupo tienen la misma frecuencia, no hay moda.
- Si dos puntuaciones adyacentes tienen la frecuencia máxima, la moda es el promedio de las dos puntuaciones adyacentes.
- Si dos puntuaciones adyacentes tienen la frecuencia máxima, la moda es el promedio de las dos puntuaciones adyacentes.Ejemplos de ejercicios de moda

Moda de datos no tabulados

En este caso es dato que más repite en un conjunto de datos dados.

La moda es el dato que más se repite por ejemplo sea el conjunto de datos x_1, x_2, x_2, x_2, x_3 entonces la moda es $\mathrm{Mo}=x_2$

Halle la moda de los siguinetes datos 3, 5,3,6,7,3,4,5,5 ya que hay hay precencia de dotas que se repiten dos veces en tonces este conjunto de datos recibe el nombre de datos bimodal Mo=3 y Mo=5

Moda de datos tabulados

La moda es el dato que más se repite por ejemplo sea el conjunto de datos tabulados entonces la moda es

$$M_o=L_i+\frac{f_i-f_{i-1}}{(f_i-f_{i-1})+(f_i-f_{i+1})}\cdot a_i$$

- L_i es el límite inferior de la clase modal
- f_i es la frecuencia absoluta de la clase modal
- f_{i-1} es la frecuencia absoluta inmediatamente inferior a la clase modal
- f_{i+1} es la frecuencia absoluta inmediatamente posterior a la clase modal
- a_i es la amplitud de la clase

Clase	Y_i	f_i	F_i	$\ldots$	$H_i^*\%$
$[y_1,y_2)$	y_1	f_1	$\ldots$	$\ldots$	$H_1^*\%$
$[y_2,y_3)$	y_2	f_2	$\ldots$	$\ldots$	$H_1^*\%$
$[y_3,y_4)$	y_3	f_3	$\ldots$	$\ldots$	$H_1^*\%$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\ldots$	$\vdots$

Clase	Y_i	f_i	F_i	$\ldots$	$H_i^*\%$
$[y_{r-1}, y_r]$	y_r	f_r	$\ldots$	$\ldots$	$H_1^*\%$

: (#tab:wwwww) wwwww

Ejemplo

Clase	f_i
$[10, 15)$	2
$[15, 20)$	5
$[20, 25)$	10
$[25, 30)$	3
$[30, 35]$	1

Primeramente la mayor frecuencia absoluta es 10 y corresponde $f_3 = 10$ por tanto $i = 3$. $L_i = 20$

$$M_o = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot a_i =$$

$$= 20 + \frac{10 - 5}{(10 - 5) + (10 - 3)} \cdot 5 = 20 + \frac{5}{12} * 5 = 22.08$$

Más información

La mediana (Me)

Mediana de datos no tabulados

Obtener la mediana consiste en ordenar los datos de menor a mayor y considerar dos casos: El primero si el número de datos es impar entonces el dato $x_{\frac{n+1}{2}}$ del conjunto ordenado será la mediana es decir

$$\mathrm{Me} = x_{\frac{n+1}{2}}$$

de otro lado si el número de datos es par entonces la mediana es la semisuma de los dos datos intermedios es decir

$$\mathrm{Me} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

```
Sean los conjuntos de datos 5, 6, 8, 2, 1, 5, 6, 7, 10, 0, 14 y 20, 25, 6, 5, 19, 5 obtener la mediana de estos conjuntos de datos.
```

```
Al ordenarlos  se obtiene el siguiente arreglo 0, 1, 2, 5, 5, 6, 6, 7, 8, 10, 14 y considerando que $x_1=0$, $x_2=1$, $\ldots$, $x_{11}=14$ en este caso e
```

Mediana de datos tabulados

La mediana se encuentra en el intervalo donde la frecuencia acumulada llega hasta la mitad de la suma de las frecuencias absolutas.

Es decir tenemos que buscar el intervalo en el que se encuentre.

$$M_e = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a_i$$

L_i es el límite inferior de la clase donde se encuentra la mediana

$\frac{N}{2}$ es la semisuma de las frecuencias absolutas

f_i es la frecuencia absoluta de la clase mediana

F_{i-1} es la frecuencia acumulada anterior a la clase mediana

a_i es la amplitud de la clase

La mediana es independiente de las amplitudes de los intervalos

Clase	Y_i	f_i	F_i	$\dots$	$H_i^*\%$
$[y_1, y_2)$	y_1	f_1	$\dots$	$\dots$	$H_1^*\%$
$[y_2, y_3)$	y_2	f_2	$\dots$	$\dots$	$H_1^*\%$
$[y_3, y_4)$	y_3	f_3	$\dots$	$\dots$	$H_1^*\%$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\dots$	$\vdots$
$[y_{r-1}, y_r]$	y_r	f_r	$\dots$	$\dots$	$H_1^*\%$

: (#tab:mediana) Mediana

Más información

Ejemplo

Clase	f_i	F_i
$[10, 15)$	1	1
$[15, 20)$	2	3
$[20, 25)$	5	8
$[25, 30)$	3	11
$[30, 35]$	1	12
$\sum$	12	

$$\frac{N}{2} = 12/2 = 6$$

ubicando en las frecuencias absolutas acumuladas que corresponde al intervalo $[20, 25)$

$$M_e = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a_i$$

$$= 20 + \frac{\frac{12}{2} - 3}{5} \cdot 5 = 23$$

por lo tanto la mediana de este conjunto de datos tabulados (agrupados) es $Me = 23$

Asignación

Halle le media, la moda y la mediana de los siguientes datos tabulados

Clase	Y_i	f_i	F_i
$[100, 150)$		1	1
$[150, 200)$		2	3
$[200, 250)$		5	
$[250, 300)$		7	
$[300, 350]$		10	
$[350, 400]$		6	
$[400, 450]$		5	
$[450, 500]$		2	
$[500, 550]$		1	

$$\overline{x} = \frac{y_1 f_1 + y_2 f_2 + \cdots + y_n f_n}{n} = \frac{1}{n}\sum_{i=1}^{n} y_i f_i$$

$$M_o = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot a_i$$

$$M_e = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a_i$$

Medidas de dispersión

Rango

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

Sean los datos 2, 5, 6, 1, 7, 5, 8, 6

$x_{max} = 8$
$x_{min} = 1$

por lo tanto $R = x_{max} - x_{min} = 8 - 1 = 7$

Varianza

Datos no tabulados

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}$$

Sean los datos 2, 5, 6, 1, 7, 5, 8, 6

$$\overline{x} = 40/8 = 5$$

$$
\begin{aligned}
s^2 &= \frac{\sum (x_i-\overline{x})^2}{n-1}\\
&= \frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+(x_3-\overline{x})^2+(x_4-\overline{x})^2+(x_5-\overline{x})^2+(x_6-\overline{x})^2+(x_7-\overline{x})^2+(x_8-\overline{x})^2}{8-1}\\
&= \frac{(2-5)^2+(5-5)^2+(6-5)^2+(1-5)^2+(7-5)^2+(5-5)^2+(8-5)^2+(6-5)^2}{8-1}\\
&= \frac{9+0+1+16+4+0+9+1}{7}\\
&= \frac{40}{7}=5.71
\end{aligned}
$$

Datos tabulados

$$s^2=\frac{\sum f_i\,(Y_i-\overline{x})^2}{n-1}$$

$\overline{x}=\frac{\sum Y_i * f_i}{n}=747.5/31=24.11$

Clase	Y_i	f_i	$Y_i * f_i$
$[5,10)$	7.5	1	7.5
$[10,15)$	12.5	2	25
$[15,20)$	17.5	5	87.5
$[20,25)$	22.5	7	157.5
$[25,30]$	27.5	10	275
$[30,35]$	32.5	6	195
$\sum$		31	

Por lo tanto la varianza para dotos agrupados es

$$
\begin{aligned}
s^2 &= \frac{\sum f_i\,(Y_i-\overline{x})^2}{n-1}\\
&= \frac{f_1\,(Y_1-\overline{x})^2+f_2\,(Y_2-\overline{x})^2+f_3\,(Y_3-\overline{x})^2+f_4\,(Y_4-\overline{x})^2+f_5\,(Y_5-\overline{x})^2+f_6\,(Y_6-\overline{x})^2}{31-1}\\
&= \frac{1\,(7.5-24.11)^2+2\,(12.5-24.11)^2+5\,(17.5-24.11)^2+7\,(22.5-24.11)^2+10\,(27.5-24.11)^2+6\,(32.5-24.11)^2}{31-1}\\
&= \frac{1*275.89+2*134.79+5*43.69+7*2.59+10*11.49+6*7.39}{31-1}\\
&= \frac{275.89+269.58+218.45+18.13+114.9+44.34}{30}\\
&= \frac{941.29}{30}=31.38
\end{aligned}
$$

Por lo tanto

$$s^2=31.38$$

Desviación típica

$$s=\sqrt{s^2}$$

La desviacion típica o estandar del siguiente conjunto de datos tabulados

Clase	Y_i	f_i	$Y_i * f_i$
$[5, 10)$	7.5	1	7.5
$[10, 15)$	12.5	2	25
$[15, 20)$	17.5	5	87.5
$[20, 25)$	22.5	7	157.5
$[25, 30]$	27.5	10	275
$[30, 35]$	32.5	6	195
$\sum$		31	

es

$$s = \sqrt{s^2} = \sqrt{31.38} = 5.60$$

Desviación media absoluta

Datos no tabulados

$$DM = \frac{1}{n} \sum |x_i - \overline{x}|$$

Sean los datos 2, 5, 6, 1, 7, 5, 8, 6

$$\overline{x} = 40/8 = 5$$

Entonces

$$\begin{aligned} DM &= \frac{\sum |x_i - \overline{x}|}{n} \\ &= ...Resolver \end{aligned}$$

Datos tabulados

$$DM = \frac{1}{n} \sum f_i \, |Y_i - \overline{x}|$$

y_i marca de clase o punto medio de la clase i

$\overline{x} = \frac{\sum Y_i * f_i}{n} = 747.5/31 = 24.11$

Clase	Y_i	f_i	$Y_i * f_i$
$[5, 10)$	7.5	1	7.5
$[10, 15)$	12.5	2	25
$[15, 20)$	17.5	5	87.5
$[20, 25)$	22.5	7	157.5
$[25, 30]$	27.5	10	275
$[30, 35]$	32.5	6	195
$\sum$		31	

Por lo tanto la desviación media absoluta es

$$\begin{aligned} DM &= \frac{\sum f_i \left|Y_i - \overline{x}\right|}{n} \\ &= \frac{f_1 \left|Y_1 - \overline{x}\right| + f_2 \left|Y_2 - \overline{x}\right| + f_3 \left|Y_3 - \overline{x}\right| + f_4 \left|Y_4 - \overline{x}\right| + f_5 \left|Y_5 - \overline{x}\right| + f_6 \left|Y_6 - \overline{x}\right|}{31} \\ &= \frac{1\left|7.5 - 24.11\right| + 2\left|12.5 - 24.11\right| + 5\left|17.5 - 24.11\right| + 7\left|22.5 - 24.11\right| + 10\left|27.5 - 24.11\right| + 6\left|32.5 - 24.11\right|}{31} \\ &= \frac{1 * 16.61 + 2 * 11.61 + 5 * 6.61 + 7 * 1.61 + 10 * 3.39 + 6 * 8.39}{31} \\ &= \frac{277.33}{31} = 8.94 \end{aligned}$$

por lo tanto

$$DM = 8.94$$

Desviación mediana absoluta

Datos no tabulados

$$DMe = \frac{1}{n} \sum \left|Y_i - Me\right|$$

Sean los datos 2, 5, 6, 1, 7, 5, 8, 6 (Ejercicio)

Datos tabulados

$$DMe = \frac{1}{n} \sum f_i \left|Y_i - Me\right|$$

$Me = ?$ (Ejercicio)

Clase	Y_i	f_i	$Y_i * f_i$
$[5, 10)$	7.5	1	7.5
$[10, 15)$	12.5	2	25
$[15, 20)$	17.5	5	87.5
$[20, 25)$	22.5	7	157.5
$[25, 30]$	27.5	10	275
$[30, 35]$	32.5	6	195
$\sum$		31	

Por lo tanto la desviacion de la mediana absoluta es

(Ejercicio)

$$\begin{aligned} s^2 &= \frac{\sum f_i \left|Y_i - Me\right|}{n} \\ &= \frac{f_1 \left(Y_1 - Me\right)^2 + f_2 \left(Y_2 - Me\right)^2 + f_3 \left(Y_3 - Me\right)^2 + f_4 \left(Y_4 - Me\right)^2 + f_5 \left(Y_5 - Me\right)^2 + f_6 \left(Y_6 - Me\right)^2}{31} \\ &= complete \end{aligned}$$

Coeficiente de variacion

$$Cv = \frac{s}{\overline{x}} \cdot 100$$

Si $Cv > 25\%$ se dice que los datos estan muy dispersos
Si $Cv < 25\%$ se dice que los datos estan muy juntos

Para el conjunto de datos

Clase	Y_i	f_i	$Y_i * f_i$
$[5, 10)$	7.5	1	7.5
$[10, 15)$	12.5	2	25
$[15, 20)$	17.5	5	87.5
$[20, 25)$	22.5	7	157.5
$[25, 30]$	27.5	10	275
$[30, 35]$	32.5	6	195
$\sum$		31	

$$Cv = \frac{5.60}{24.11} \cdot 100 = 0.23 \cdot 100 = 23\%$$

Asignación

Halle el rango, la varianza, la desviación típica, desviación media, desviación mediana absoluta y el coeficiente de variación. Grafique el hstograma y ubique estos estadigrafos

Clase	Y_i	f_i	F_i
$[50, 100)$		8	1
$[100, 150)$		20	3
$[150, 200)$		50	
$[200, 250)$		70	
$[250, 300]$		100	
$[300, 350]$		60	
$\sum$		20	

Medidas de posicion (cuantiles)

- Los deciles, que dividen a la distribución en diez partes;
- Los percentiles, que dividen a la distribución en cien partes.

Cuartiles

Los cuartiles, que dividen a la distribución en cuatro partes (corresponden a los cuantiles 0,25; 0,50 y 0,75); Q_1, Q_2, Q_3

Datos no agrupados

Sean los datos 1, 2, 5, 1 ,5, 6, 7, 8, 9, 3, 4, 5, 2, 6, 2, 5, 6, 7 Ordenar de menor a mayor (creciente)

Si

$$Q_k = \frac{k(n+1)}{4}$$

es entero entonces el cuartil es el dato de la posicion $Q_k = x_{\frac{k(n+1)}{4}}$ en caso contrario se interpola los datos extremos donde se encuentra el valor Q_k

- Ejemplo
  1, 2, 5, 1 ,5, 6, 7, 8, 9, 3, 4, 5, 2, 6, 2, 5, 6, 7 ordenados de menor a mayor
  1, 1, 2, 2, 2, 3, 4, 5, 5, 5, 5, 6, 6, 6, 7, 7, 8, 9

$Q_1 = \frac{1(18+1)}{4} = 4.75$ interpolando $Q_1 = 2 + (2-2) \cdot 0.75 = 2$

$Q_2 = \frac{2(18+1)}{4} = 9.5$ interpolando $Q_2 = 5 + (5-5) \cdot 0.5 = 5$

$Q_3 = \frac{3(18+1)}{4} = 14.25$ interpolando $Q_3 = 6 + (7-6) \cdot 0.25 = 6.25$

Datos agrupados

$$Q_k = L_i + A\left(\frac{\frac{kn}{4} - F_{i-1}}{F_i - F_{i-1}}\right); k = 1, 2, 3$$

- L_i limite inferior del intervalo que contiene al decil
- F_{i-1} frecuencia acumulada en la clase anterior al decil
- F_i frecuencia acumulada en la clase al decil
- A amplitud intervalica
- n numero de datos
- k indice del cuartil correspondiente

$$\begin{aligned} Q_1 &= L_i + A\left(\frac{\frac{1n}{4} - F_{i-1}}{F_i - F_{i-1}}\right) \\ &= L_i + A\left(\frac{\frac{1*39}{4} - F_{i-1}}{F_i - F_{i-1}}\right) \\ &= 20 + 5 \cdot \left(\frac{9.75 - 8}{15 - 8}\right) \\ &= 21.5 \end{aligned}$$

$$\begin{aligned} Q_2 &= L_i + A\left(\frac{\frac{2n}{4} - F_{i-1}}{F_i - F_{i-1}}\right) \\ &= L_i + A\left(\frac{\frac{2*39}{4} - F_{i-1}}{F_i - F_{i-1}}\right) \\ &= 25 + 5 \cdot \left(\frac{19.5 - 15}{25 - 15}\right) \\ &= 27.5 \end{aligned}$$

$$\begin{aligned} Q_3 &= L_i + A\left(\frac{\frac{3n}{4} - F_{i-1}}{F_i - F_{i-1}}\right) \\ &= L_i + A\left(\frac{\frac{2*39}{4} - F_{i-1}}{F_i - F_{i-1}}\right) \\ &= 30 + 5 \cdot \left(\frac{29.25 - 25}{31 - 25}\right) \\ &= 33.542 \end{aligned}$$

Clase	Y_i	f_i	F_i
$[5, 10)$	7.5	1	1
$[10, 15)$	12.5	2	3
$[15, 20)$	17.5	5	8
$[20, 25)$	22.5	7	15
$[25, 30]$	27.5	10	25
$[30, 35]$	32.5	6	31
$[35, 40]$	37.5	5	36
$[40, 45]$	42.5	3	39
$\sum$		39	

Quintiles

wwwwwwwww

Deciles

Los deciles, que dividen a la distribución en diez partes es decir $D_1, D_2, \ldots, D_9$

Datos no agrupados

Sean los datos 1, 2, 5, 1 ,5, 6, 7, 8, 9, 3, 4, 5, 2, 6, 2, 5, 6, 7 Ordenar de menor a mayor (creciente)

Si

$$D_k = \frac{k(n+1)}{10}$$

es entero entonces el decil es el dato de la posicion $D_k = x_{\frac{k(n+1)}{10}}; k = 1, 2, 3, \ldots, 9$
Si

$$D_k = \frac{k(n+1)}{10}$$

no es entero entonces el decil es la interpolacion lineal de de los dos valores entre las cuales se encuentra $D_k = \frac{k(n+1)}{10}$

- Ejemplo
  1, 2, 5, 1 ,5, 6, 7, 8, 9, 3, 4, 5, 2, 6, 2, 5, 6, 7 ordenados de menor a mayor
  1, 1, 2, 2, 2, 3, 4, 5, 5, 5, 5, 6, 6, 6, 7, 7, 8, 9
  entonces

$$D_9 = \frac{9(18+1)}{10} = 17.1$$

interpolando el decil 9 es $D_9 = 8 + (9 - 8) \cdot 0.1 = 8.1$

Datos agrupados

$$D_k = L_i + A\left(\frac{\frac{kn}{10} - F_{i-1}}{F_i - F_{i-1}}\right)$$

Clase	Y_i	f_i	F_i
$[5, 10)$	7.5	1	1
$[10, 15)$	12.5	2	3
$[15, 20)$	17.5	5	8
$[20, 25)$	22.5	7	15
$[25, 30]$	27.5	10	25
$[30, 35]$	32.5	6	31
$[35, 40]$	37.5	5	36
$[40, 45]$	42.5	3	39
$\sum$		39	

$$D_9 = L_i + A\left(\frac{\frac{9\cdot 39}{10} - F_{i-1}}{F_i - F_{i-1}}\right)$$

Entonces $\frac{9\cdot 39}{10} = 35.1$

$$D_9 = 35 + 5\left(\frac{35.1 - 31}{36 - 31}\right) = 39.1$$

Percentiles

Los percentiles, que dividen a la distribución en diez partes es decir $P_1, P_2, \ldots, P_{99}$

Datos no agrupados

Si

$$P_k = \frac{k(n+1)}{100}; k = 1, 2, \ldots, 99$$

es entero entonces el cuartil es el dato de la posicion $P_k = x_{\frac{k(n+1)}{100}}$

Si

$$P_k = \frac{k(n+1)}{100}$$

no es entero entonces el cuartil es la interpolacion lineal de de los dos valores entre las cuales se encuentra $Q_k = \frac{k(n+1)}{100}$

- Ejemplo
  1, 2, 5, 1 ,5, 6, 7, 8, 9, 3, 4, 5, 2, 6, 2, 5, 6, 7
  Al ordenar de manera creciente
  1, 2, 5, 1 ,5, 6, 7, 8, 9, 3, 4, 5, 2, 6, 2, 5, 6, 7
  y

$$P_k = \frac{k(18+1)}{100}$$

Datos agrupados

$$P_k = L_i + A\left(\frac{\frac{kn}{100} - F_{i-1}}{F_i - F_{i-1}}\right)$$

Clase	Y_i	f_i	F_i
$[5, 10)$	7.5	1	1
$[10, 15)$	12.5	2	3
$[15, 20)$	17.5	5	
$[20, 25)$	22.5	7	
$[25, 30]$		10	
$[30, 35]$		6	
$[35, 40]$		5	
$[40, 45]$		3	
$\sum$		2	

Medidas de asimetria

wwwwww

wwwwww

wwwwww

Medidas de curtosis

leptocurtica

mesocurtica

palticurtica

Medidas de asimetría