

Clase 6 Elementos de Estadística Descriptiva

La estadística permite realizar inferencias y sacar conclusiones a partir de datos.

- Planificación: recolección de datos y diseño del estudio a realizar
- Descripción: exploración de los datos, su estructura.
- Inferencia: son los métodos que permiten hacer predicciones

- Población: total de elementos que se consideran en el análisis para realizar el estudio.
- Muestra: cualquier subconjunto de una población de donde se recolectarán los datos.
- Parámetros: medidas que se calculan sobre la población.
- Estadísticos: medidas que se calculan sobre la muestra.

Tabla de frecuencias

La tabla de frecuencias de n observaciones de la variable X :

- X : variable, x_1, x_2, \dots, x_n , observaciones de la variable
- f frecuencia absoluta con la que aparece cada x_i
- f_r frecuencia relativa con la que aparece cada x_i
- F frecuencia acumulada, (la suma final es igual a n)
- F_r frecuencia acumulada relativa, (la suma final es igual a 1)

- n_i : número de ocurrencia de cada observación x_i , $i = 1, \dots, n$
- N : número total de casos

- Las frecuencias absolutas: $f_i = f(x_i) = n_i$, $i = 1, \dots, n$
- $\sum_{i=1}^n f_i = N$
- Las frecuencias relativas: $f_{r_i} = f_r(x_i) = \frac{n_i}{N}$, $i = 1, \dots, n$
- $\sum_{i=1}^n f_{r_i} = 1$

- Las frecuencias absolutas acumuladas:

$$F_k = \sum_{i=1}^k f_i, \quad k = 1, \dots, n$$
- $\sum_{i=1}^n F_i = N$
- Las frecuencias relativas acumuladas:

$$F_{r_k} = \sum_{i=1}^k f_{r_i}, \quad k = 1, \dots, n = 1$$
- $\sum_{i=1}^n F_{r_i} = 1$

Ejemplo

Se arroja un dado 20 veces, calcular:

- las frecuencias absolutas de aparición de las observaciones
- las frecuencias relativas
- las frecuencias acumuladas absoluta y relativa

```
> (dado = sample(1:6,20,replace=T))  
[1] 3 3 2 6 4 6 5 6 1 1 4 6 5 1 6 6 5 1 5 6
```

```
#frecuencias absolutas  
> (f_dado = table(dado))
```

```
dado
```

```
1 2 3 4 5 6
```

```
4 1 2 2 4 7
```

```
#frecuencias relativas
```

```
> (f_r_dado = prop.table(f_dado))
```

```
dado
```

```
      1      2      3      4      5      6  
0.20 0.05 0.10 0.10 0.20 0.35
```

```
# frecuencias acumuladas absolutas y relativas
```

```
> cumsum(f_dado)
```

```
1  2  3  4  5  6
```

```
4  5  7  9 13 20
```

```
> cumsum(f_r_dado)
```

```
      1      2      3      4      5      6  
0.20  0.25 0.35  0.45  0.65  1.00
```

Intervalos de frecuencias

- Generar los intervalos
- Consideremos el rango de los valores que toma la variable x
- $\text{rango} = \max(x_i) - \min(x_i), i = 1, \dots, n$
- Dividimos el rango en n intervalos
- $[a_1, a_2), [a_2, a_3), \dots, [a_{n-1}, a_n]$

Ejemplo

Generamos un vector de 120 elementos, con valores entre 37 y 85

```
> u=sample(c(37:85),120,replace=TRUE)
```

Calculamos frecuencias absolutas con la funcion `table()`:

```
> utabla=table(u)
```

Calculamos frecuencias relativas usando la funcion `prop.table()` que se aplica a `table()`

```
> utabla_frec=prop.table(utabla)
```

Tabla de frecuencias usando intervalos.

```
# Armamos intervalos (120 intervalos)

> intervalos_utabla=cut(u,breaks=seq(37,85,
  length=nclass.Sturges(u)),
  include.lowest=TRUE)

> intervalos_utabla
# es un objeto tipo factor, de 120 intervalos
  con los levels correspondientes, cada
  intervalo es de longitud 6.9.

# (64.4,71.3] (71.3,78.1] ... (43.9,50.7]
Levels: [37,43.9] (43.9,50.7] (50.7,57.6]
  (57.6,64.4] (64.4,71.3] (71.3,78.1] (78.1,85]
```

Generamos la tabla de frecuencias absolutas en intervalos

```
> utabla_intervalos = table(intervalos_utabla)
```

```
> utabla_intervalos
```

```
# [37,43.9] (43.9,50.7] (50.7,57.6] (57.6,64.4]
#          16          16          18          17
# (64.4,71.3] (71.3,78.1]   (78.1,85]
#          20          17          16
```

Generamos la tabla de frecuencias relativas en intervalos

```
> utabla_frec_intervalos = prop.table(
  utabla_intervalos)
> utabla_frec_intervalos
```

# [37,43.9]	(43.9,50.7]	(50.7,57.6]	(57.6,64.4]
# 0.133	0.133	0.150	0.141
# (64.4,71.3]	(71.3,78.1]	(78.1,85]	
# 0.166	0.141	0.133	

Definiciones

Sean x_1, x_2, \dots, x_n , observaciones de la variable X

- media: $\mu = \frac{1}{n} \sum_{i=1}^n x_i$
- varianza: $\sigma = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$
- desviación standard: $s = \sqrt{\sigma}$

Ejercicio

A partir de un experimento de 1000 tiradas de una moneda, armar una tabla con la siguiente descripción:
media, mediana, varianza, desviación standard, mínimo, máximo y rango

Ejercicio

Se realizó una encuesta a un conjunto de n personas acerca de la construcción de edificios de más de 4 pisos en un barrio residencial de casas bajas.

Las n respuestas se midieron en una escala de 1 a 9, donde:

- 1 representa un total desacuerdo con la medida,
- 9 un acuerdo total.

El resultado de la encuesta fue:

1 5 6 8 6 5 1 5 1 6 5 7 5 5 4 5 8 5 4 2 1 6 4 1 4 8 4 3 4 3 3 1 4 5
6 5 8 5 4 9 4 3 5 3 4 9 4 2 6 3 4 2 4 1 3 6 3 1 2 4 4 6 2 4 7 4 2 4
6 4 4 6 7 5 8 5 7 6 5 6 5 7 5 6 4 5 4 1 6 5 1 5 5 5 4 6 2 5 1 6 5 4
4 3 5 5 9 4 3 6 5 7 3 2 9 4 7 4 2 1 8 2 7 4 5 5 2 5 5 1 5 8 5 6 7 6
6 7 7 5 2 5 6 5 8 1 3 6 2 5

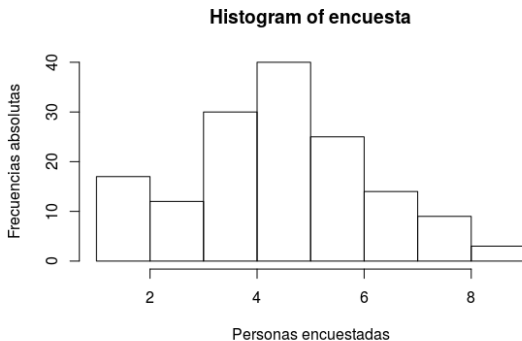
- ¿Cuántas personas participaron de la encuesta ?
- ¿Cuál fue la respuesta más frecuente?
- ¿Cuántas personas tienen, como máximo, una actitud de 2 puntos?
- ¿Cuál es el promedio de las respuestas?
- ¿Cuántas personas tienen, como mínimo, 8 puntos?
- ¿Qué desviación tienen los datos con respecto al promedio?
- Según las respuestas de los items anteriores ¿qué conclusiones se pueden hacer?

Histogramas

- La distribución de las frecuencias se construye de tal manera que se puede clasificar los datos en clases.
- Las clases o intervalos de clases son mutuamente excluyentes
- Construcción de la tabla de frecuencias
 - Dividir el rango total de los datos en clases o intervalos, no necesariamente de la misma longitud
 - En cada intervalo se cuenta el número de observaciones que caen en dicho intervalo y se determina la frecuencia de cada clase

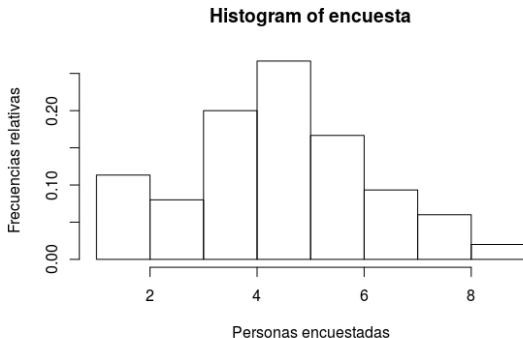
Consideremos los datos de la encuesta que están en el archivo "encuesta.txt"

```
encuesta=scan(file="encuesta.txt")  
% Read 150 items  
Frecuencias absolutas  
hist(encuesta,freq=TRUE,ylab="Frecuencias  
absolutas", xlab="Personas encuestadas")
```



Frecuencias relativas

```
hist_encuesta=hist(encuesta,freq=FALSE,ylab="Frecuencias relativas", xlab="Personas encuestadas")
```



La información del histograma usando la función histinfo

```
> histinfo=hist_encuesta
> histinfo
$breaks
[1] 1 2 3 4 5 6 7 8 9
$counts
[1] 17 12 30 40 25 14 9 3
$density
[1] 0.11333333 0.08000000 0.20000000 0.26666667
    0.16666667 0.09333333 0.06000000 0.02000000
$mids
[1] 1.5 2.5 3.5 4.5 5.5 6.5 7.5 8.5
$xname
[1] "encuesta"
$equidist
[1] TRUE
attr(,"class")
[1] "histogram"
```