Practica 1 Probabilidad

Robinson Aldair Cuayal

2023-03-03

Índice

Punto 1

Para dar solucion a este ejercicio en primer lugar se ingresara los datos

```
sexo = c('M','M','F','M','F','F','F','F','M','M','F','M','M','F','F','F',
         'M','F','F','M','F','M','F','M','M','F','F','F','F','M','F','F',
         'M','M','M','F','F','F','M','F')
diagnositico = c('EM','RM','FE','EM','RM','EM','FE','FE','RM','EM','FE',
                 'FE','EM','RM','RM','FE','RM','FE','RM','FE','EM','EM',
                 'FE','FE','FE','EM','RM','EM','RM','FE','FE','RM','FE',
                 'RM','FE','EM','EM','EM','RM','RM')
edad = c(29,35,34,36,25,20,31,89,42,41,47,41,87,56,50,28,35,23,39,42,72,
         52,31,35,42,29,61,18,64,51,30,35,40,76,59,71,62,65,51,18)
destino = c(2,7,7,7,7,7,7,1,7,7,7,2,7,1,7,7,7,7,3,7,6,7,7,7,7,2,7,3,7,7,
            7,7,6,3,7,6,7,3,7,7)
clinicaDB = data.frame(sexo, diagnositico,edad,destino)
clinicaDB
```

```
##    sexo diagnositico edad destino
## 1     M           EM   29       2
## 2     M           RM   35       7
## 3     F           FE   34       7
## 4     M           EM   36       7
## 5     F           RM   25       7
## 6     F           EM   20       7
## 7     F           FE   31       7
## 8     F           FE   89       1
## 9     M           RM   42       7
## 10    M           EM   41       7
## 11    F           FE   47       7
## 12    M           FE   41       2
## 13    M           EM   87       7
## 14    F           RM   56       1
## 15    F           RM   50       7
## 16    F           FE   28       7
## 17    M           RM   35       7
## 18    F           FE   23       7
## 19    F           RM   39       3
## 20    M           FE   42       7
## 21    F           EM   72       6
## 22    M           EM   52       7
## 23    F           FE   31       7
## 24    M           FE   35       7
## 25    M           FE   42       7
## 26    F           EM   29       2
## 27    F           RM   61       7
## 28    F           EM   18       3
## 29    F           RM   64       7
## 30    M           FE   51       7
## 31    F           FE   30       7
## 32    F           RM   35       7
## 33    M           FE   40       6
## 34    M           RM   76       3
```

```
## 35    M           FE   59       7
## 36    F           EM   71       6
## 37    F           EM   62       7
## 38    F           EM   65       3
## 39    M           RM   51       7
## 40    F           RM   18       7
```

El paso siguiente es examinar cuales son las variables cualitativas y cuantitativas, entre las que se encuentra la siguiente forma: cuantitativa -> edad y destino cualitativa -> sexo y diagnostico

Variables cuantitativas

Tabla de frecuencias variables cuantitativas

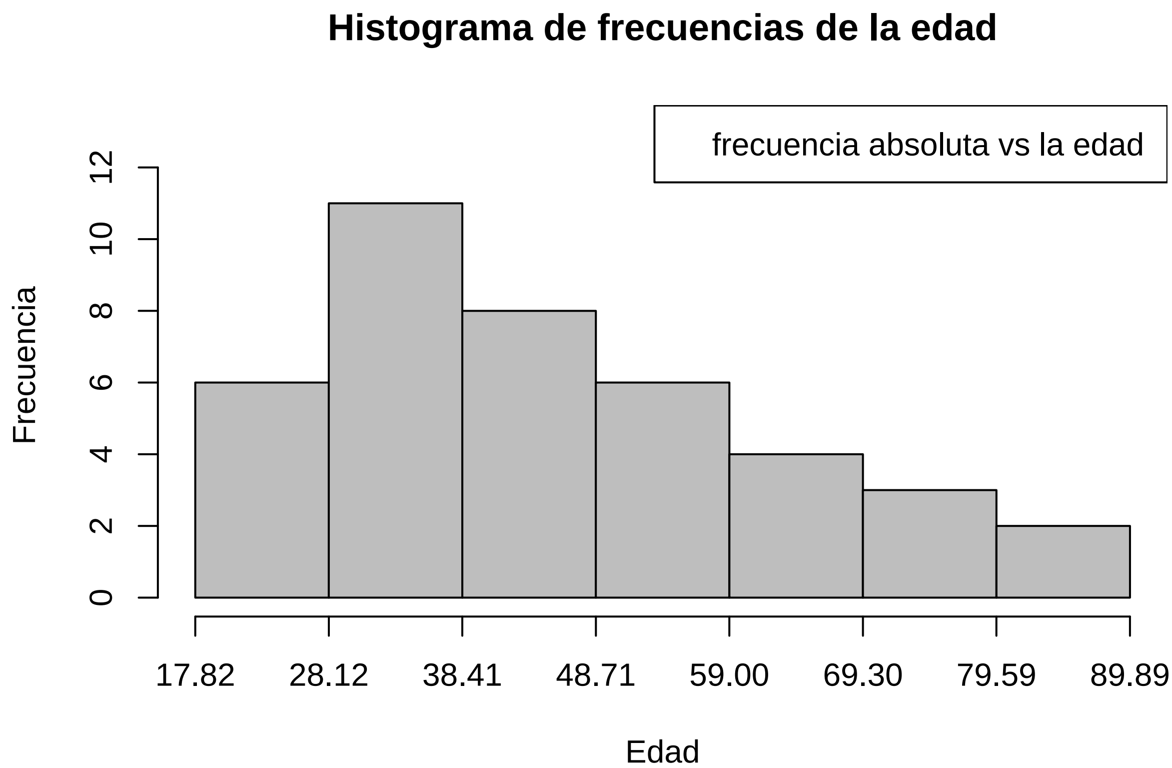
Edad

```
library(fdth)
frecEdad = fdt(clinicaDB$edad, breaks="Sturges")
frecEdad
```

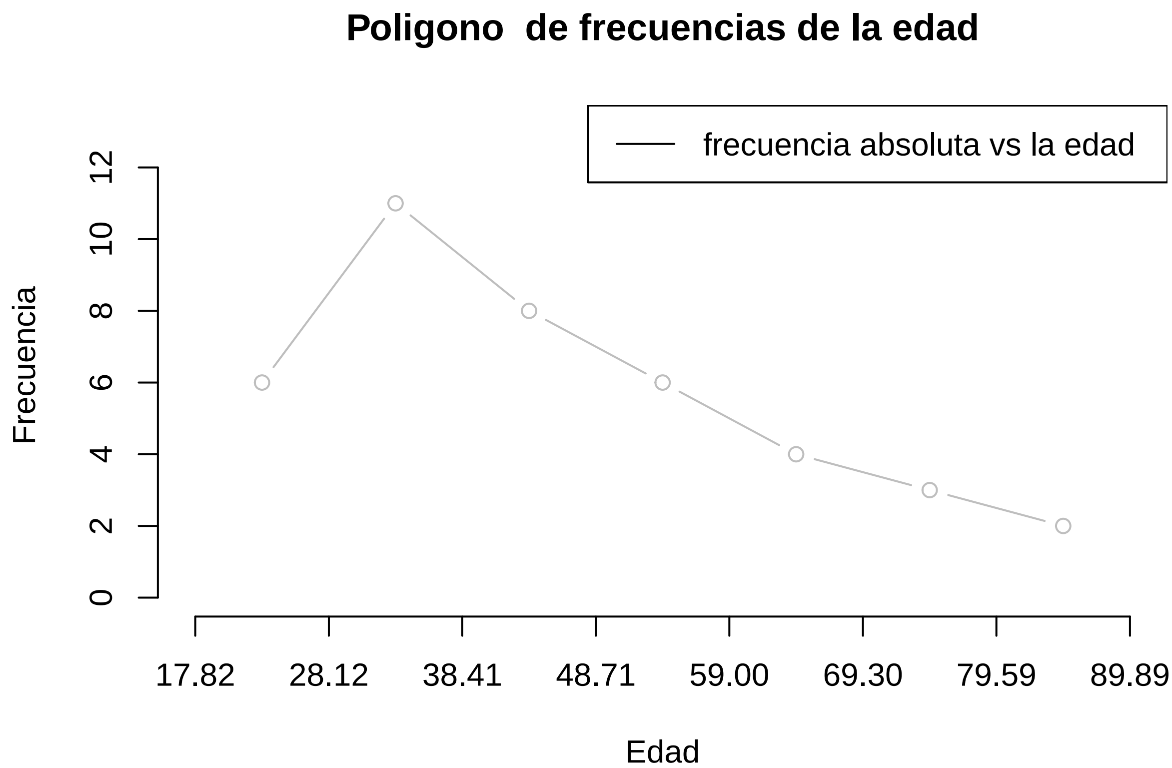
```
##     Class limits  f   rf rf(%) cf cf(%)
##   [17.82,28.116)  6 0.15  15.0  6  15.0
##  [28.116,38.411) 11 0.28  27.5 17  42.5
##  [38.411,48.707)  8 0.20  20.0 25  62.5
##  [48.707,59.003)  6 0.15  15.0 31  77.5
##  [59.003,69.299)  4 0.10  10.0 35  87.5
##  [69.299,79.594)  3 0.07   7.5 38  95.0
##   [79.594,89.89)  2 0.05   5.0 40 100.0
```

las graficas son las siguientes

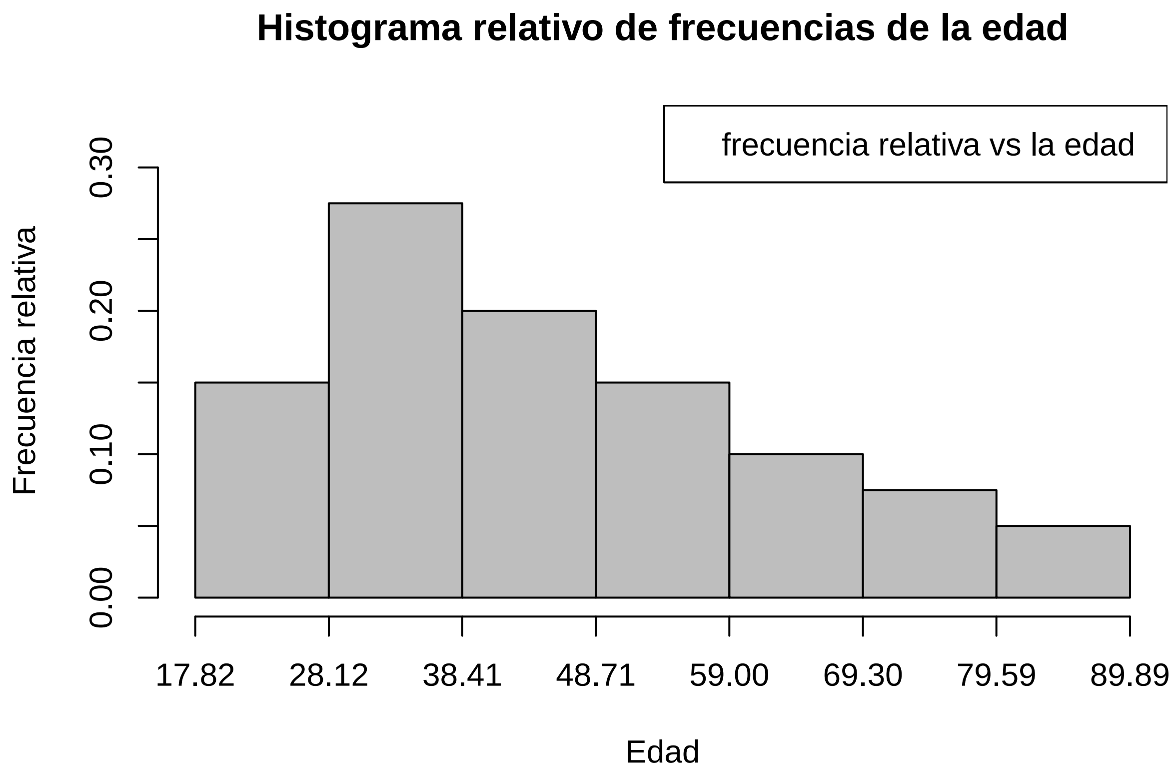
```
#plot(clinicaDB$edad,frecEdad$table$f, type = 'h')
plot(frecEdad, type = "fh", main = "Histograma de frecuencias de la edad",
     xlab = 'Edad', ylab = 'Frecuencia ')
  legend('topright',
       legend = c("frecuencia absoluta vs la edad "))
```




```r
plot(frecEdad, type = "fp", main = "Poligono  de frecuencias de la edad",
     xlab = 'Edad', ylab = 'Frecuencia ')
  legend('topright',
       legend = c("frecuencia absoluta vs la edad  "),
       col = c("black"),
       lty = 1)
```




```
plot(frecEdad, type = "rfh", main = "Histograma relativo de frecuencias de la edad",
     xlab = 'Edad', ylab = 'Frecuencia relativa')
  legend('topright',
       legend = c("frecuencia relativa vs la edad  "))
```



Para ver cuales es el valor maximo, minimo, media, mediana y el 3 cuartil.

```
summary(edad)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.00   31.00   41.00   44.80   56.75   89.00
```

```
# con los percentiles
quantile(edad, 0.25)
```

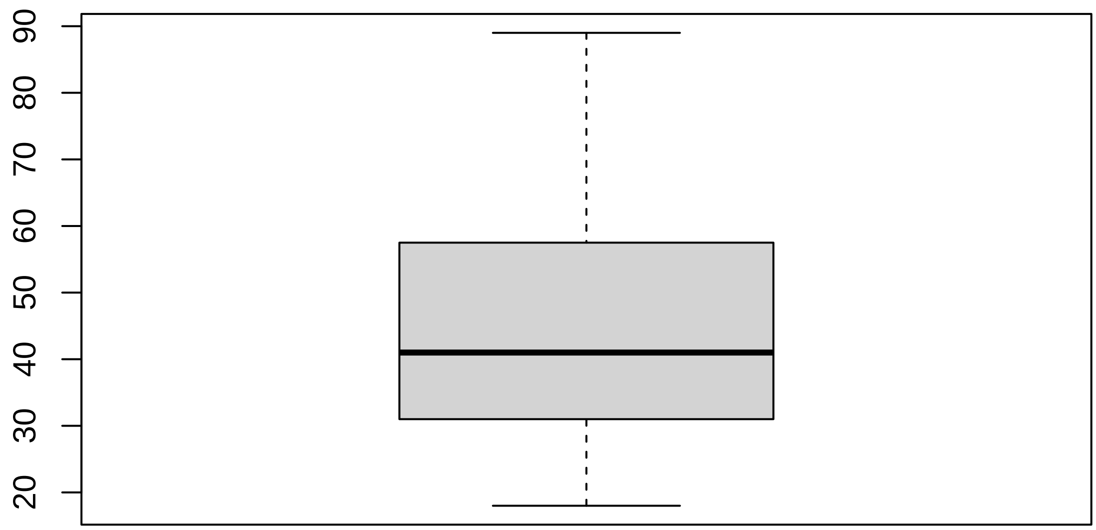
```
## 25%
##  31
```

```
quantile(edad, c(0.1,0.2,0.3,0.4))
```

```
##  10%  20%  30%  40%
## 24.8 29.8 34.7 35.6
```

Segun el diagrama de cajas se puede observar que se tiene una buena distribucion de datos, sin tener datos atipicos que se dispersen mucho de los demas,siendo confiable manejar la media para algunos procesos donde se la necesite.

```
#para generar el diagrama de cajas
boxplot(edad)
```



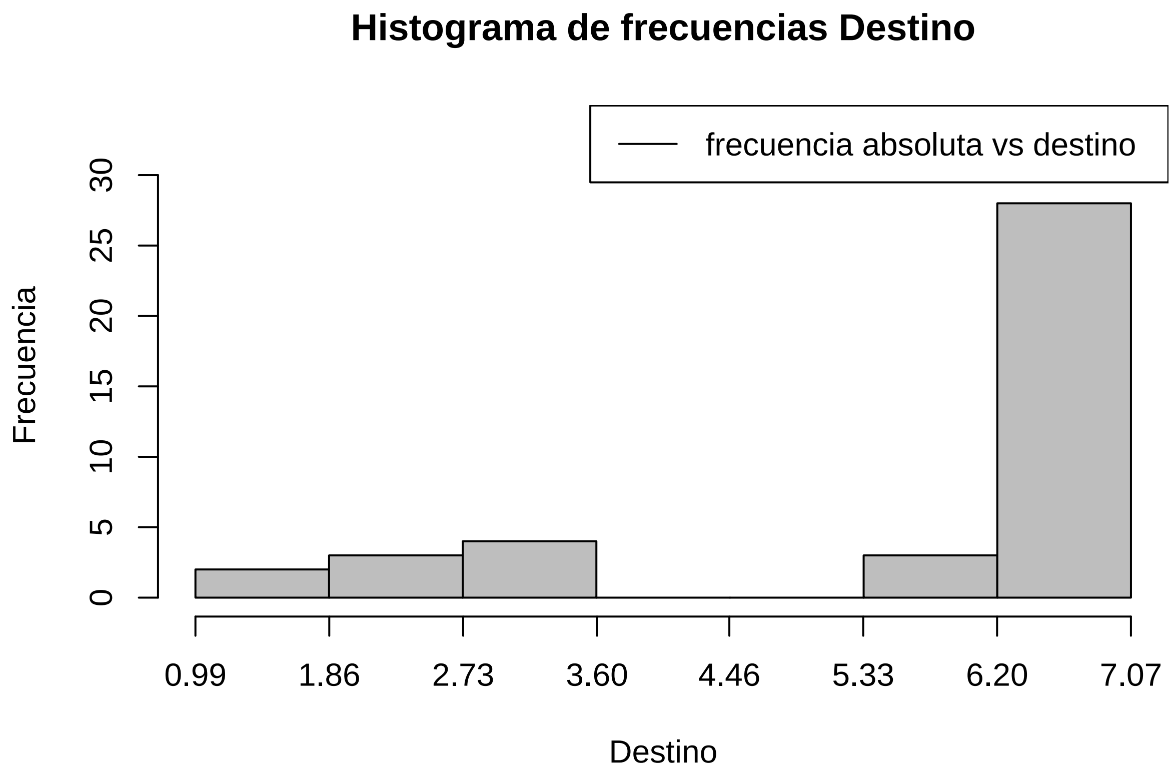
Destino

```r
frecDestino = fdt(clinicaDB$destino, breaks="Sturges")
frecDestino
```

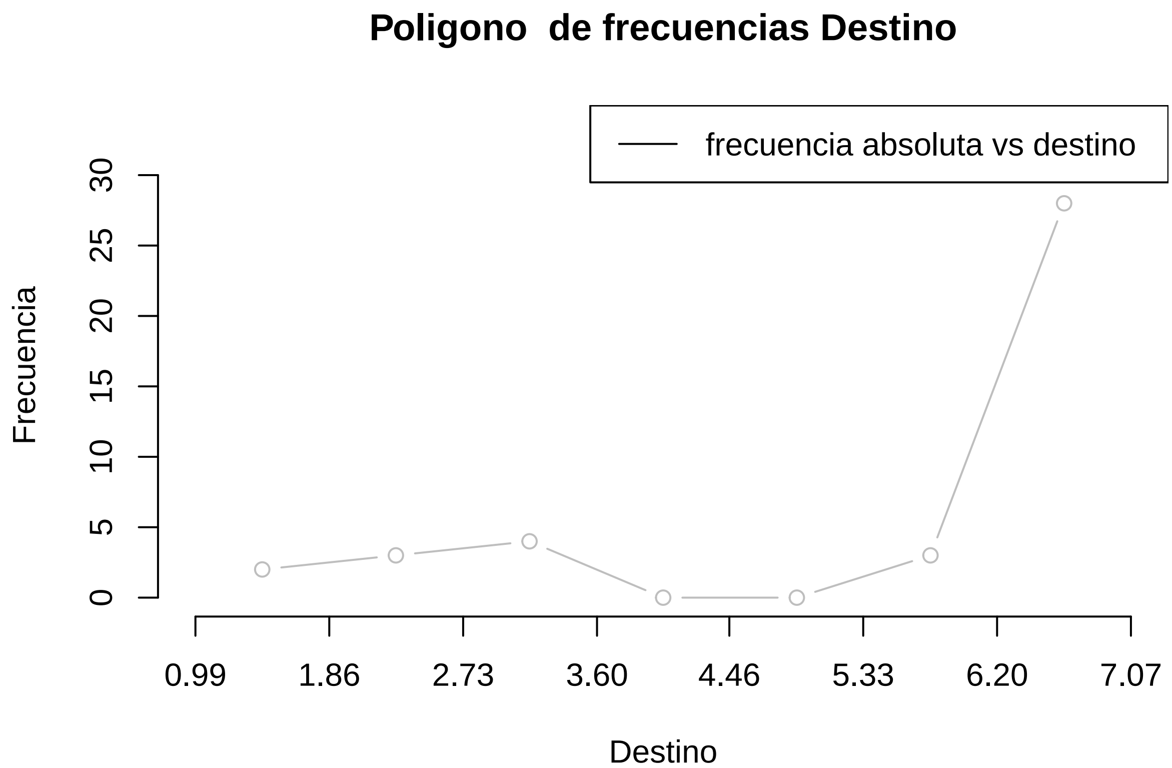
```
##   Class limits  f   rf rf(%) cf cf(%)
##    [0.99,1.859)  2 0.05   5.0  2   5.0
##   [1.859,2.727)  3 0.07   7.5  5  12.5
##   [2.727,3.596)  4 0.10  10.0  9  22.5
##   [3.596,4.464)  0 0.00   0.0  9  22.5
##   [4.464,5.333)  0 0.00   0.0  9  22.5
##   [5.333,6.201)  3 0.07   7.5 12  30.0
##    [6.201,7.07) 28 0.70  70.0 40 100.0
```

las graficas son las siguientes

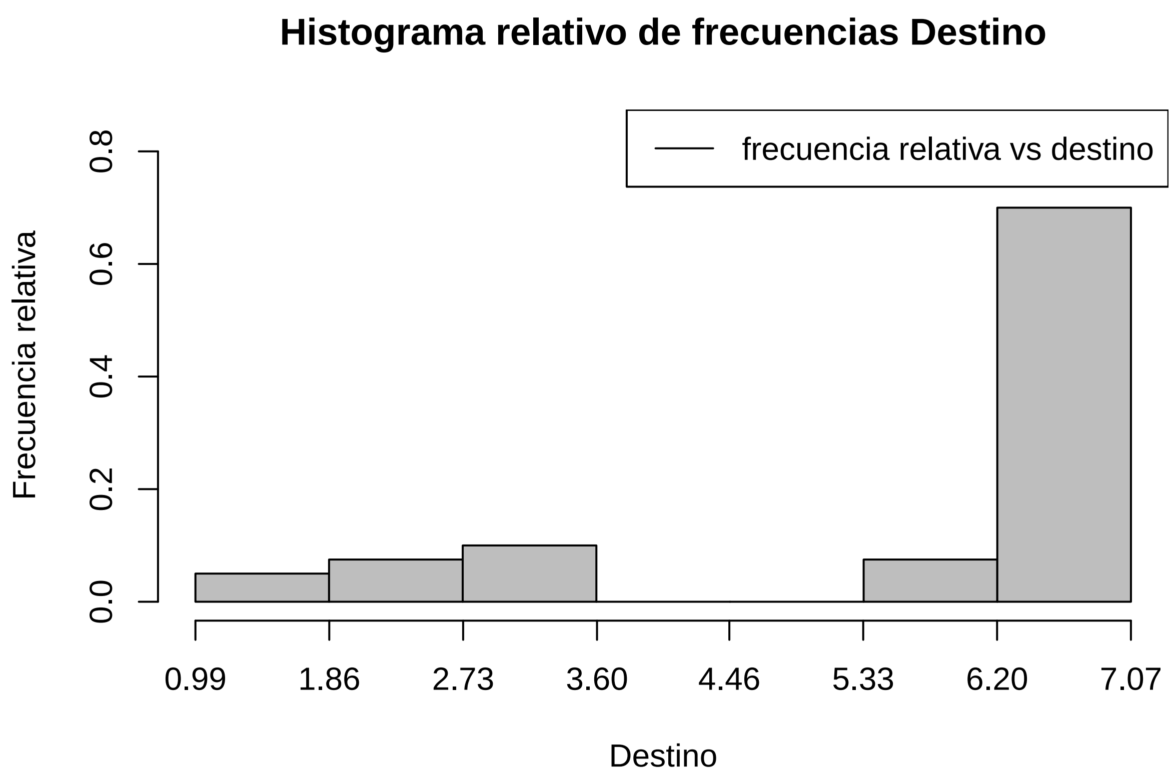
```r
#plot(clinicaDB$edad,frecEdad$table$f, type = 'h')
plot(frecDestino, type = "fh", main = "Histograma de frecuencias Destino",
     xlab = 'Destino', ylab = 'Frecuencia ')
  legend('topright',
       legend = c("frecuencia absoluta vs destino  "),
       col = c("black"),
       lty = 1)
```




```
plot(frecDestino, type = "fp", main = "Poligono  de frecuencias Destino",
     xlab = 'Destino', ylab = 'Frecuencia ')
  legend('topright',
       legend = c("frecuencia absoluta vs destino  "),
       col = c("black"),
       lty = 1)
```




```r
plot(frecDestino, type = "rfh", main = "Histograma relativo de frecuencias Destino",
     xlab = 'Destino', ylab = 'Frecuencia relativa')
  legend('topright',
      legend = c("frecuencia relativa vs destino"),
      col = c("black"),
      lty = 1)
```



Para ver cuales es el valor maximo, minimo, media, mediana y el 3 cuartil.

```
summary(destino)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    6.00    7.00    5.85    7.00    7.00
```

```
# con los percentiles
quantile(destino, 0.25)
```

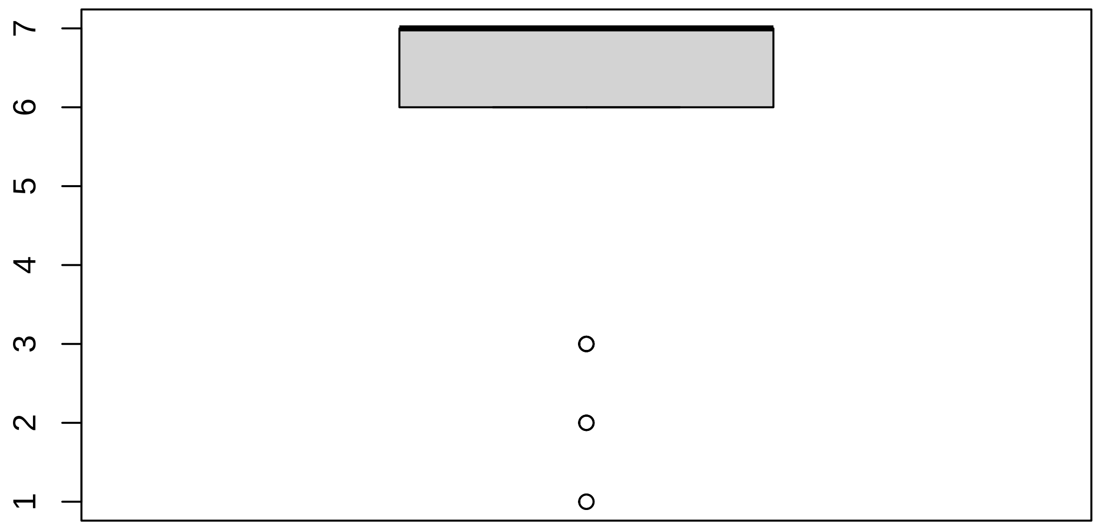
```
## 25%
##   6
```

```
quantile(destino, c(0.1,0.2,0.3,0.4))
```

```
## 10% 20% 30% 40%
## 2.0 3.0 6.7 7.0
```

Segun el diagrama de cajas, se puede observar que hay datos atipicos entre 1,2 y 3 ya que la concentracion de los datos esta en 7. Para un analisis mas acorde a la realidad se recomienda tratar estos datos atipicos por aparte y manejar los datos restantes.

```
#para generar el diagrama de cajas
boxplot(destino)
```



Variables cualitativas

Sexo y Diagnóstico

Para hacer una tabla cruzada con las variables cualitativas para ver la relacion que comparten, y asi observar la frecuencia con una combinación particular de valores que ocurre en los datos. Estos valores se organizan en filas y columnas para crear una matriz de observaciones.

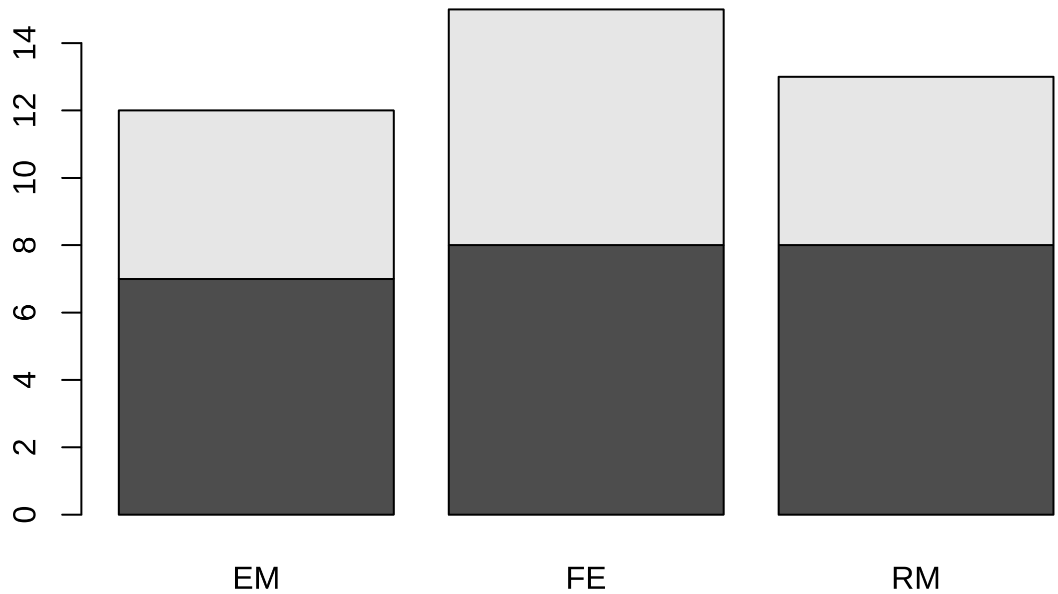
```
tabla <- table(clinicaDB$sexo, clinicaDB$diagnositico)
tabla
```

```
##
##     EM FE RM
##   F  7  8  8
##   M  5  7  5
```

Se puede observar en la tabla cruzada que del sexo femenino se tiene a personas con enfermo mental de 7, fisicamente enfermo de 8 y retraso mental de 8. Por parte del sexo masculino se tiene a personas con enfermo mental de 5, fisicamente enfermo de 7 y retraso mental de 5.

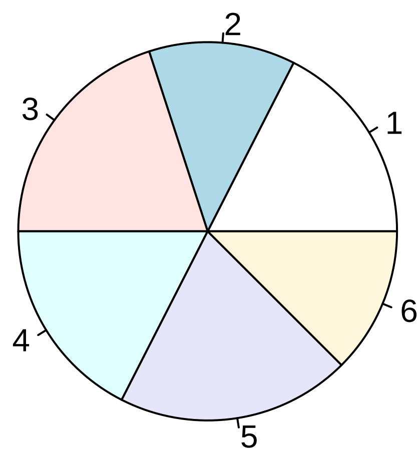
graficamos para ver su comportamiento

```
barplot(tabla)
```



generamos un diagrama circular

```
pie(tabla)
```



Punto 2

Entonces, para calcular la probabilidad de que exactamente 5 de los siguientes 20 pacientes intervenidos sobrevivan, aplicamos la ecuacion:$P(X = x) = (nCx) * p^x * q^(n - x)$ y con los datos ingresados queda de la siguiente forma $P(X = 5) = (20C5) * (0.9)^5 * (0.1)^15$.

```
# Definimos los parámetros de la distribución binomial
n = 20 # número total de pacientes intervenidos
x = 5 # número de pacientes que sobreviven
p = 0.9 # probabilidad de que un paciente sobreviva

# Calculamos la probabilidad usando dbinom
probabilidad = dbinom(x, n, p)

# Imprimimos el resultado

probabilidad
```

```
## [1] 9.154957e-12
```

La probabilidad de que exactamente 5 de los siguientes 20 pacientes intervenidos sobrevivan es $9.154957e-12$.

Punto 3

Para resolver este problema, vamos a utilizar la distribución binomial ya que se trata de una situación en la que se realizan múltiples ensayos independientes y cada uno de ellos tiene dos resultados posibles (aprobar o desaprobar el consumo diario de marihuana).

Como primer paso se calcula la probabilidad de que un estudiante desapruebe el consumo diario de marihuana es $p = 0.7$, y la probabilidad de que apruebe es $q = 1 - p = 0.3$.

a) cualquiera entre 7 y 9

Para calcular la probabilidad de que el número de estudiantes que desaprueban el consumo diario de marihuana sea cualquiera entre 7 y 9, podemos calcular la probabilidad acumulada para $k = 6; k = 10$ y restarlos:

```
p_a = sum(sapply(7:9, function(k) dbinom(k, 12, 0.7)))
p_a
```

```
## [1] 0.6293359
```

La probabilidad de que el número de estudiantes que desaprueban el consumo diario de marihuana sea entre 7 y 9 es del 62.9%.

b) > 5

Para calcular la probabilidad de que más de 5 estudiantes desaprueben el consumo diario de marihuana, se calcula la probabilidad acumulada hasta k=5 y restarla de 1:

```
p_b = 1 - sum(sapply(0:5, function(k) dbinom(k, 12, 0.7)))
p_b
```

```
## [1] 0.9613992
```

La probabilidad de que más de 5 estudiantes desaprueben el consumo diario de marihuana es del 96.1%.

c) < 8

Para calcular la probabilidad de que menos de 8 estudiantes desaprueben el consumo diario de marihuana, se calcula la probabilidad acumulada hasta k=7:

```
p_c = sum(sapply(0:7, function(k) dbinom(k, 12, 0.7)))
p_c
```

```
## [1] 0.2763445
```

La probabilidad de que menos de 8 estudiantes desaprueben el consumo diario de marihuana es del 27%.

d) tabla de distribucion y distribucion acumalada

Para realizar la tabla de distribución y distribución acumulada, se utiliza las mismas funciones utilizadas anteriormente:

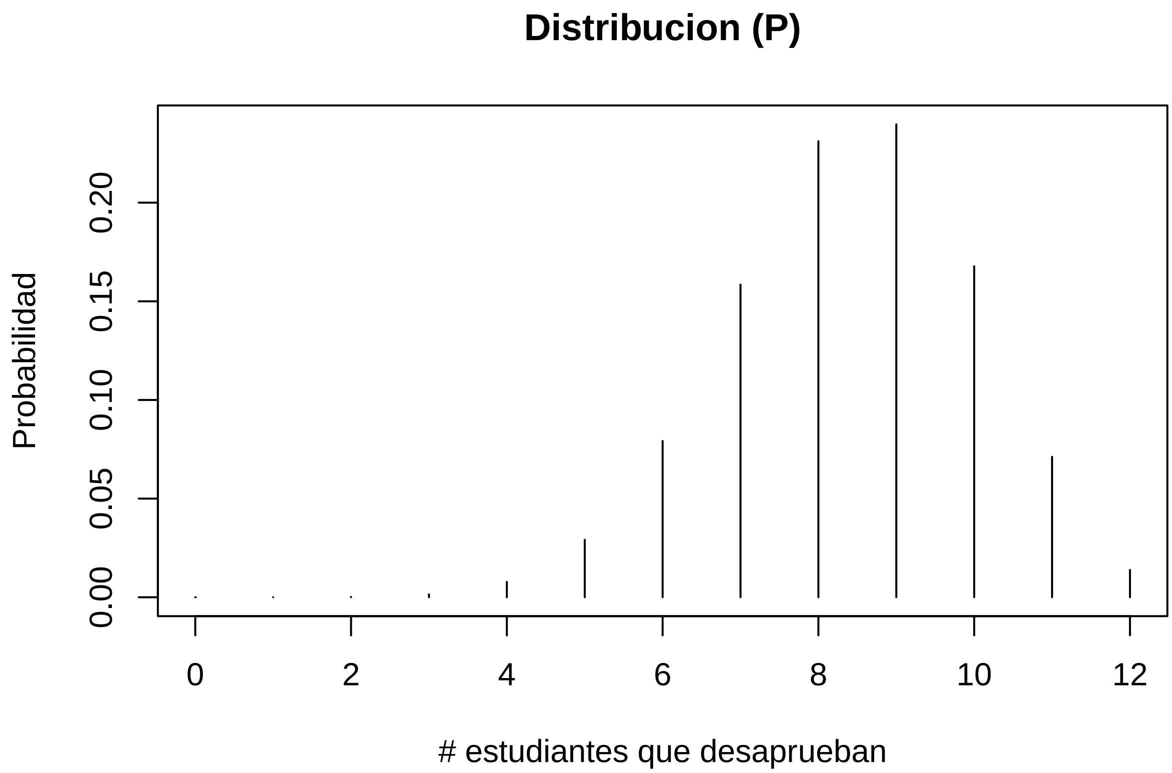
```
distribucion = data.frame(k=0:12, P=sapply(0:12,
                                            function(k) dbinom(k, 12, 0.7)))
distribucion$Pa = cumsum(distribucion$P)

distribucion # ver tabla completa
```

```
##     k            P           Pa
## 1   0 5.314410e-07 5.314410e-07
## 2   1 1.488035e-05 1.541179e-05
## 3   2 1.909645e-04 2.063763e-04
## 4   3 1.485279e-03 1.691655e-03
## 5   4 7.797716e-03 9.489371e-03
## 6   5 2.911147e-02 3.860084e-02
## 7   6 7.924790e-02 1.178487e-01
## 8   7 1.584958e-01 2.763445e-01
## 9   8 2.311397e-01 5.074842e-01
## 10  9 2.397004e-01 7.471847e-01
## 11 10 1.677903e-01 9.149750e-01
## 12 11 7.118376e-02 9.861587e-01
## 13 12 1.384129e-02 1.000000e+00
```

e) graficos

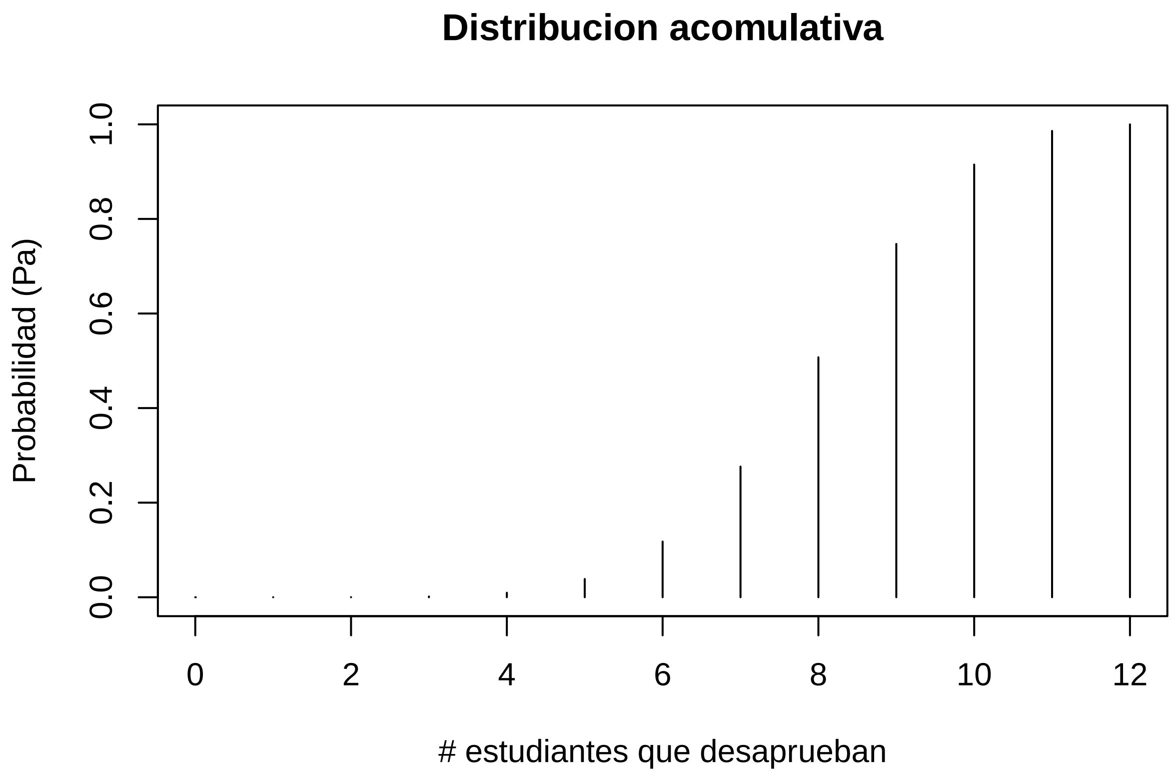
Para realizar los graficos, se realizan con el siguiente codigo El histograma para distribucion normal

```
plot(distribucion$k, distribucion$P, type = 'h',
     xlab = '# estudiantes que desaprueban', ylab = 'Probabilidad',
     main = 'Distribucion (P)')
```



El histograma para distribucion acomulativa

```
plot(distribucion$k, distribucion$Pa, type = 'h',
     xlab = '# estudiantes que desaprueban',
     ylab = 'Probabilidad (Pa)', main = 'Distribucion acomulativa')
```



Punto 4

Para este ejercicio se tiene que hallar una distribucion de poisson y dependiendo del ejercicio se va a utilizar las funciones

a) P(x = 6)

```
media = 3.7
prob_accidentes_6 = dpois(x = 6, lambda = media)
prob_accidentes_6
```

```
## [1] 0.08810251
```

La probabilidad es de aproximadamente 8.8%.

b) P(x < 2)

Para calcular la probabilidad de que en un año haya menos de 2 accidentes, podemos utilizar la función ppois(), que nos da la probabilidad acumulada hasta ese punto.

```r
prob_menos_2_accidentes = ppois(q = 1, lambda = media)
prob_menos_2_accidentes
```

```
## [1] 0.1162006
```

La probabilidad es de aproximadamente 11%.

c) P(X > 8)

Para calcular la probabilidad de que en un año haya más de 8 accidentes, se utiliza la misma función ppois(), pero esta vez necesitaremos la probabilidad acumulada desde el otro extremo. Es decir, estamos buscando $P(X > 8) = 1 - P(X <= 8)$.

```r
prob_mas_8_acciden <- 1 - ppois(q = 8, lambda = media)
prob_mas_8_acciden
```

```
## [1] 0.01370281
```

La probabilidad es de aproximadamente 1.3%.

d) P(X <= X_max) = 0.9

Para calcular el número máximo de accidentes con una probabilidad mayor o igual a 0.9, podemos utilizar la función qpois(), que nos da el cuantil correspondiente a una determinada probabilidad acumulada. Estamos buscando el valor de X tal que $P(X <= X_max = 0.9)$.

```r
maximo_prob_90 = qpois(p = 0.9, lambda = media)
maximo_prob_90
```

```
## [1] 6
```

El número máximo de accidentes con una probabilidad mayor o igual a 0.9 es de 6.

e) Accidentes en 20 años

Para simular el número anual de accidentes en un periodo de 20 años, se utiliza la función rpois(), que nos permite generar valores aleatorios de una distribución de Poisson. Simplemente necesitamos llamar a la función especificando el número de valores que queremos generar (20, en este caso) y la media de la distribución (que ya habíamos definido antes, como $media = 3.7$).

```r
simulacion_accidentes <- rpois(n = 20, lambda = media)
simulacion_accidentes
```

```
##  [1] 3 2 3 1 3 5 2 2 1 3 5 4 3 3 3 2 4 1 1 6
```

Se genera un vector de 20 números aleatorios con distribución de Poisson con $media = 3.7$.

Punto 5

Vamos a modelar el número de artículos con burbujas por cada 8000 como una distribución de Poisson con media 8, ya que para calcular la probabilidad de menos de 7 artículos con burbujas necesitamos conocer la probabilidad acumulada de la Poisson hasta $x = 6$.

La probabilidad de que haya menos de 7 artículos con burbujas en una muestra de 8000 es:

```
prob_menos_7 = ppois(6, lambda = 8)
prob_menos_7
```

```
## [1] 0.3133743
```

La probabilidad es de aproximadamente 0.3133743.