

Estadística Descriptiva

Paúl Arévalo, Esteban Vizhñay

2024-06-28

```
X <- c(120, 121, 120, 119, 121, 120, 120, 119, 120, 121, 120, 120, 122, 120, 121, 120, 119, 122, 120, 1
```

La Frecuencia Absoluta y Frecuencia Relativa se calcula de la siguiente manera:

```
table(X)
```

```
## X
## 119 120 121 122
##   4  10   4   2
```

```
table(X) / length(X)
```

```
## X
## 119 120 121 122
## 0.2 0.5 0.2 0.1
```

```
addmargins(table(X))
```

```
## X
## 119 120 121 122 Sum
##   4  10   4   2  20
```

```
addmargins(table(X) / length(X))
```

```
## X
## 119 120 121 122 Sum
## 0.2 0.5 0.2 0.1 1.0
```

Frecuencias Acumuladas

```
cumsum(table(X))
```

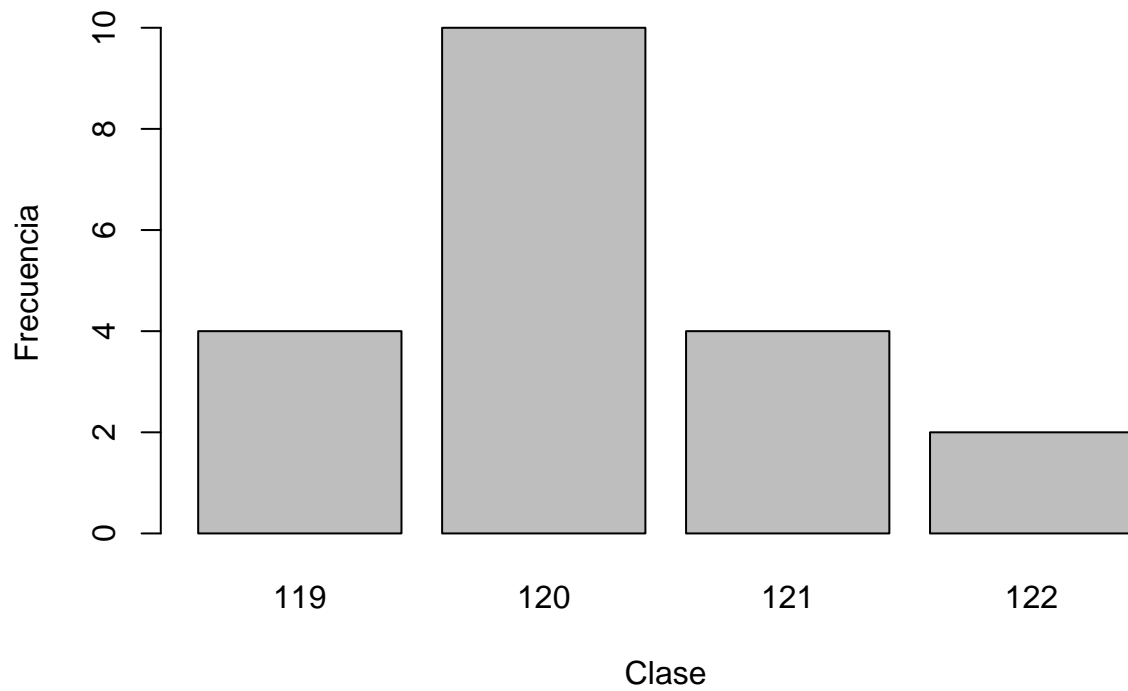
```
## 119 120 121 122
##   4  14  18  20
```

```
cumsum(table(X) / length(X))
```

```
## 119 120 121 122
## 0.2 0.7 0.9 1.0
```

Gráficos para clases de tipo Cualitativo

```
barplot(table(X), xlab = "Clase", ylab = "Frecuencia")
```



```
barplot(table(X) / length(X), xlab = "Clase", ylab = "Frecuencia")
```

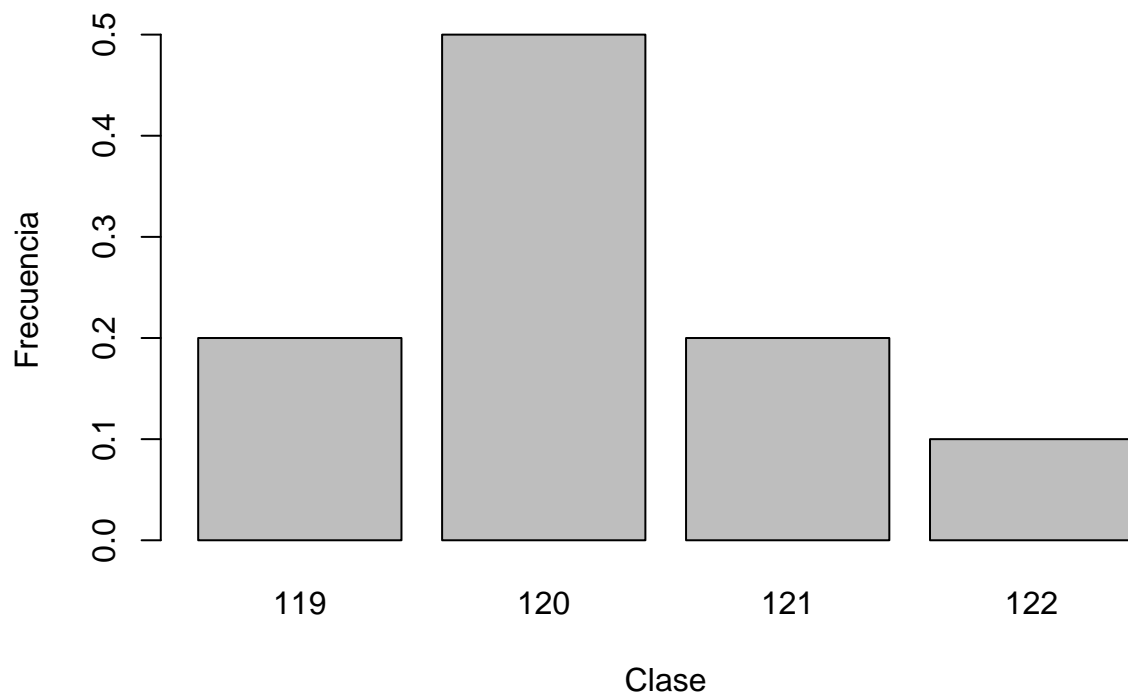
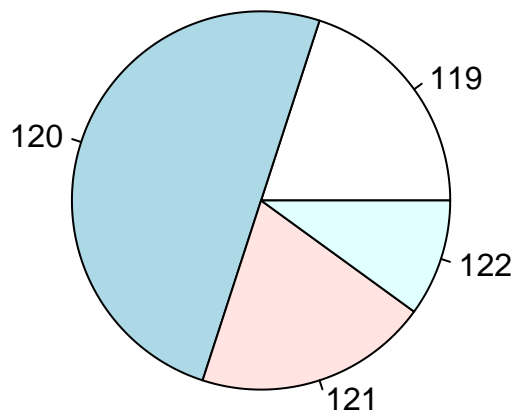
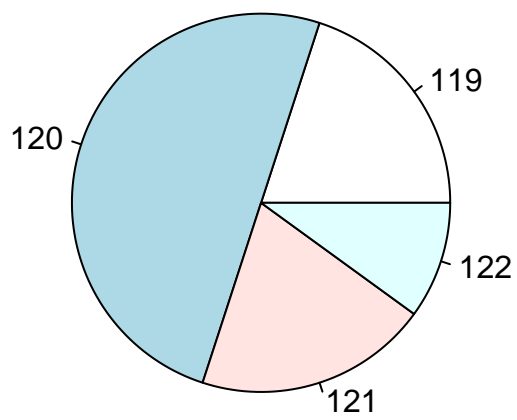


Diagrama de sectores

```
pie(table(X))
```



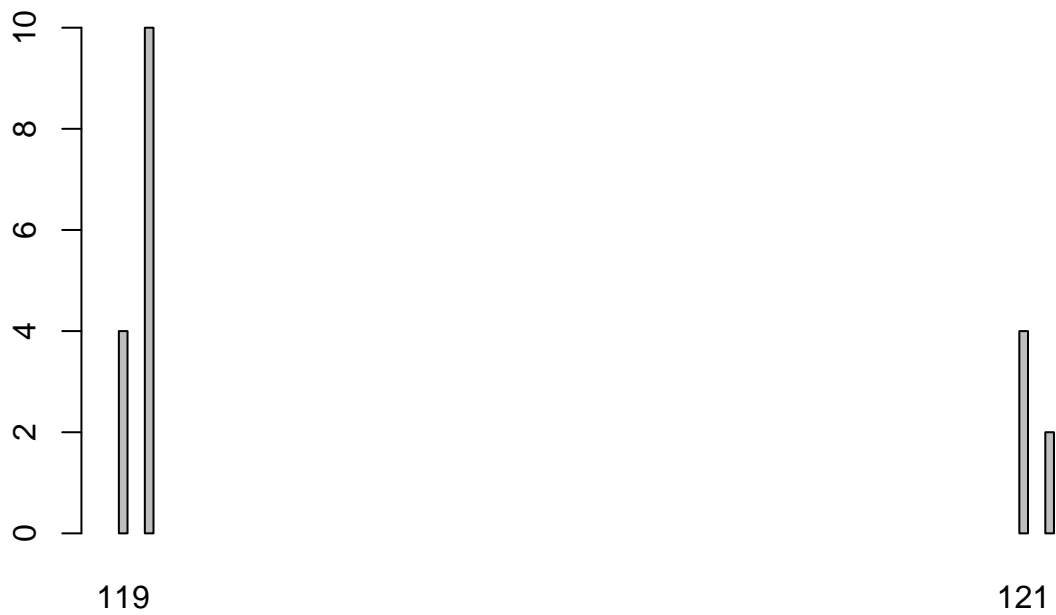
```
pie(table(X) / length(X))
```



Gráficos para Clases Cuantitativas

Diagrama de Barras

```
barplot(table(X), space = c(100, 2))
```

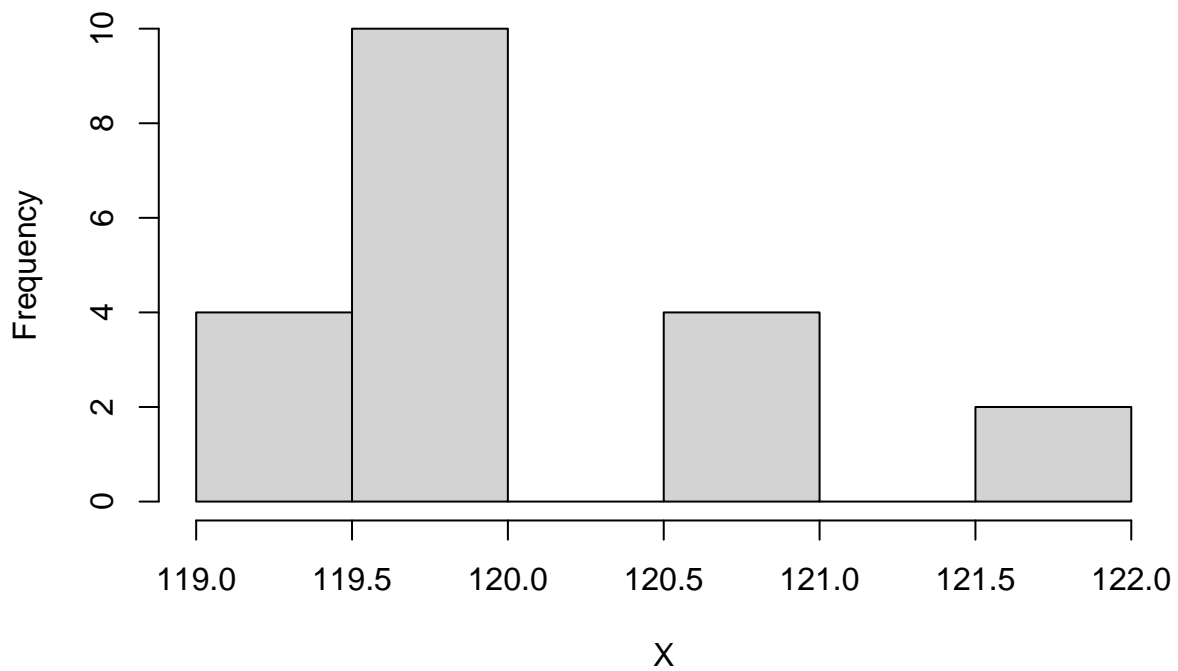


```
barplot(table(X) / length(X), space = c(100, 2))
```



```
hist(X)
```

Histogram of X



Polígono de Frecuencias

Tallo y Hoja Procedimiento semigráfico útil cuando se tiene menos de 50 datos.

```
stem(X)
```

```
##
## The decimal point is at the |
```

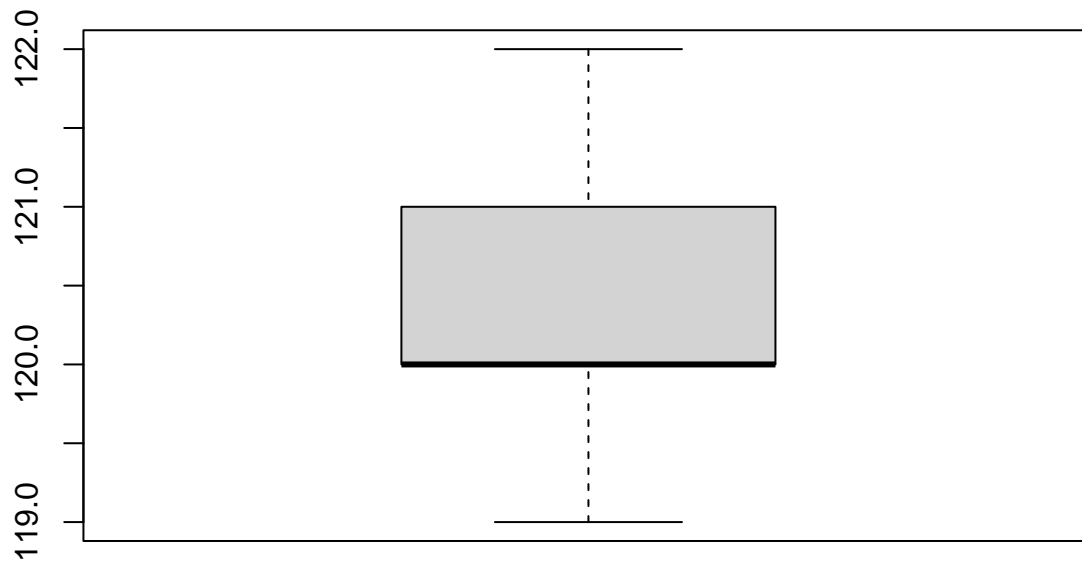
```
##
## 119 | 0000
## 120 | 0000000000
## 121 | 0000
## 122 | 00

ingresos <- c(66814.195, 42144.338, 25697.767, 35976.874, 39060.606, 13362.839, 61674.641, 53451.356)
stem(ingresos)
```

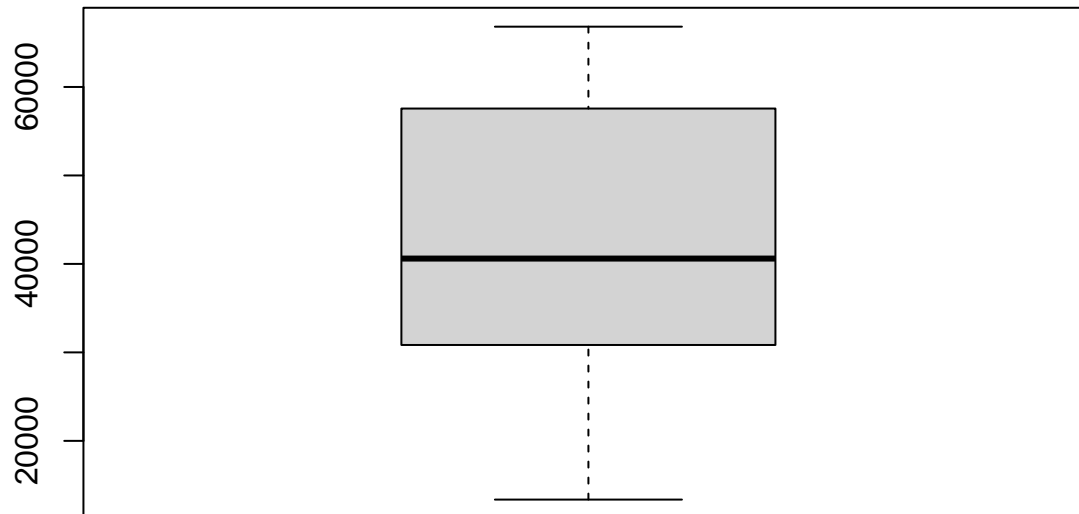
```
##
## The decimal point is 4 digit(s) to the right of the |
##
## 0 | 3
## 2 | 669
## 4 | 23
## 6 | 27
```

Box Plot

```
boxplot(X)
```



```
boxplot(ingresos)
```



Agrupación de Datos Cuantitativos

```
datos <- c(4, 8, 4, 6, 8, 6, 7, 7, 7, 8, 10, 9, 7, 6, 10, 8, 5, 9, 6, 3, 7, 6, 4, 7, 6, 9, 7, 4, 7, 6, 8)
```

Longitud de Datos

```
n <- length(datos)
n
```

```
## [1] 44
```

Recomendaciones para construir la tabla de frecuencias

Rango R

```
mayorValor <- max(datos)
menorValor <- min(datos)
R <- mayorValor - menorValor
```

extraer el max y min en un nuevo vector, puede ser alternativo

```
rango <- range(datos)
```

Número de clases de acuerdo a la recomendación

```
k <- 6
```

Amplitud

```
Amplitud <- R / k
```

por simplicidad se redefine a la amplitud como 1

```
Amplitud <- round(Amplitud)
```

Limites de los intervalos

```
seq(menorValor, mayorValor, length = k)
```

```
## [1] 3.0 4.6 6.2 7.8 9.4 11.0
```

Construcción de los intervalos mediante la función cut():

```
intervalosDatos <- cut(datos, breaks = seq(menorValor, mayorValor, length = k), include.lowest = TRUE)
```

frecuencia absoluta

```
frecuencia <- table(intervalosDatos)
frecuencia
```

```
## intervalosDatos
## [3,4.6] (4.6,6.2] (6.2,7.8] (7.8,9.4] (9.4,11]
##      5      11      11      12      5
```

Frecuencia Relativa fr

```
fr <- frecuencia / n # fr=prop.table(frecuencia)
fr
```

```
## intervalosDatos
## [3,4.6] (4.6,6.2] (6.2,7.8] (7.8,9.4] (9.4,11]
## 0.1136364 0.2500000 0.2500000 0.2727273 0.1136364
```

Frecuencia Acumulada F

```
F <- cumsum(frecuencia)
F
```

```
## [3,4.6] (4.6,6.2] (6.2,7.8] (7.8,9.4] (9.4,11]
##      5      16      27      39      44
```

Frecuencia Acumulada Relativa Fra

```
Fra <- cumsum(fr)
```

Tabla final de frecuencias

```
tablafinal <- cbind(frecuencia, fr, F, Fra) # juntamos todo
tablafinal
```

```
##      frecuencia      fr  F      Fra
## [3,4.6]      5 0.1136364  5 0.1136364
## (4.6,6.2]     11 0.2500000 16 0.3636364
## (6.2,7.8]     11 0.2500000 27 0.6136364
## (7.8,9.4]     12 0.2727273 39 0.8863636
## (9.4,11]      5 0.1136364 44 1.0000000
```

Medidas Estadísticas

Medidas de Tendencia Central

Media Aritmética

X

```
## [1] 120 121 120 119 121 120 120 119 120 121 120 120 122 120 121 120 119 122 120
## [20] 119
```

```
mean(X)
```

```
## [1] 120.2
```

ingresos

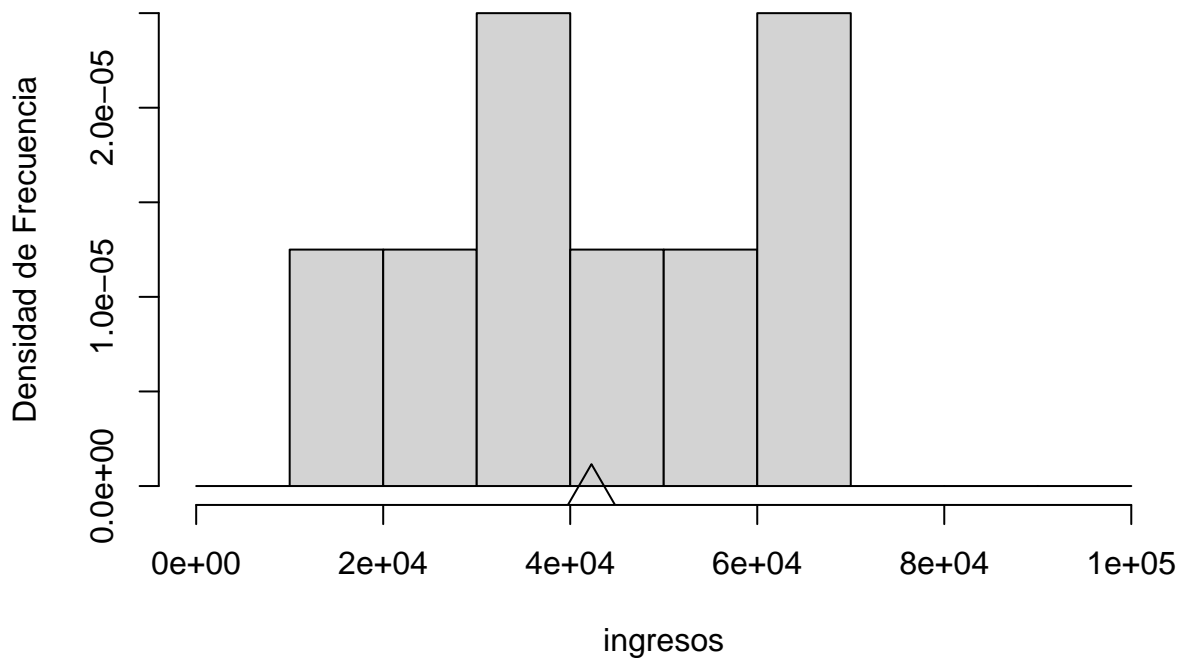
```
## [1] 66814.20 42144.34 25697.77 35976.87 39060.61 13362.84 61674.64 53451.36
```

```
mean(ingresos)
```

```
## [1] 42272.83
```

```
hist(ingresos, breaks = seq(0, 100000, by = 10000), freq = FALSE, main = "Histograma de Ingresos", ylab =  
points(mean(ingresos), -0.0000005, pch = 24, cex = 2.8))
```

Histograma de Ingresos



Mediana

```
median(X)
```

```
## [1] 120
```

Moda

```
moda <- function(x) {  
  t <- table(x)
```



```
    return(as.numeric(names(t)[t == max(t)]))
}
```

```
moda(X)
```

```
## [1] 120
```

Medidas de Dispersión

```
X <- c(120, 121, 120, 119, 121, 120, 120, 119, 120, 121, 120, 120, 122, 120, 121, 120, 119, 122, 120, 1
```

Rango

```
range(X)
```

```
## [1] 119 122
```

Varianza Muestral

```
var(X)
```

```
## [1] 0.8
```

Desviación Estándar Muestral (desviación típica)

```
sd(X)
```

```
## [1] 0.8944272
```

Desviación Media

Coefficiente de Variación

Medidas de Dispersión

Cuartiles

```
quantile(X, c(0.25, 0.50, 0.75))
```

```
## 25% 50% 75%
```

```
## 120 120 121
```

Percentiles

```
quantile(X, c(0.01, 0.02, 0.9))
```

```
## 1% 2% 90%
```

```
## 119.0 119.0 121.1
```

Deciles

```
quantile(X, c(0.1, 0.20, 0.30))
```

```
## 10% 20% 30%
```

```
## 119.0 119.8 120.0
```

Resumen

```
summary(X)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    119.0   120.0   120.0   120.2   121.0   122.0
```