

# Asignación 2: Resumen de datos

Dr. Marco Aurelio González Tagle

15 /02/ 2020

## Índice

<b>Objetivo</b>	<b>6</b>
<b>Tablas de frecuencia: <i>table(data)</i></b>	<b>6</b>
Ejercicio 1 . . . . .	6
Ejercicio 2 . . . . .	8
Ejercicio 3 . . . . .	9
Ejercicio 4 . . . . .	9
<b>Histogramas: <i>hist(data)</i></b>	<b>10</b>
Ejercicio 5 . . . . .	10
<b>Referencias</b>	<b>11</b>

## Objetivo

El objetivo de la siguiente asignación la interpretación gráfica y en cuadros de diversas fuentes de datos obtenidas en el sector forestal. Las tablas de frecuencia se pueden realizar con la función *table* y adicionalmente instalando librería *plyr* y usando la función *count*.

## Tablas de frecuencia: *table(data)*

Realice las actividades indicadas en los siguientes apartados de la actividad 1. Recuerde que el script debe estar 100 % funcional y compilado en formato *pdf*.

### Ejercicio 1

Los datos de trabajo provienen del libro *Introductory probability & statistics, applications for forestry & natural sciences* de Kozak et al. (2008). El Cuadro 1 muestra el número de accidentes por mes reportados en un aserradero para los últimos 20 meses.

Cuadro 1: Muestra el número de accidentes por mes reportados en un aserradero para los últimos 20 meses.

0	1	0	2	2	1	4	3	0	1
5	1	2	3	4	0	1	1	3	4

- Construir una tabla de frecuencias (absoluta y relativa) para el número de accidentes como el Cuadro 2.

Cuadro 2: Proporción absoluta y relativa de accidentes que se representan en el aserradero.

#	Accidentes	F absoluta	F relativa
0			
1			
2			
3			
4			
5			

- ¿Cuál es el promedio de accidentes al mes?
- ¿Qué número de accidentes reporta la mayor proporción (%)?
- Construir un barplot con la frecuencia de accidentes.

Ejemplo solución usando la librería *plyr* función *count*

```
library(plyr)
accidentes <- c(0,1,0,2,2,1,4,3,0,1,5,1,2,3,4,0,1,1,3,4)
acc <- count(accidentes)
acc
```

```
##   x freq
## 1 0    4
## 2 1    6
## 3 2    3
## 4 3    3
## 5 4    3
## 6 5    1
```

Agregamos a la BD acc una nueva columna (`acc$rfr`) con la fórmula de frecuencia relativa (`acc$freq/sum(acc$freq)*100`)

```
acc$rfr <- acc$freq/sum(acc$freq)*100
```

Finalmente graficamos mediante la función *barplot*

```
barplot(acc$freq, names.arg = acc$x, main=" Accidentes en el aserradero",
        col="lightgreen")
```

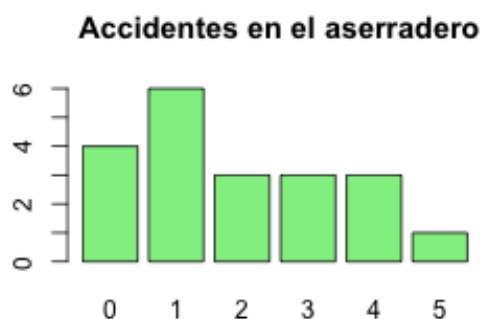


Figura 1: Frecuencia de meses con 0 a 5 accidentes

## Ejercicio 2

En un sitio permanente de muestreo de la costa Oeste de Columbia Británica se registraron la especie de los 24 árboles presentes. Especies: C: Cedro Rojo (Western red cedar); F: Douglasia verde (Douglas fir); H: Tsuga heterófila (western hemlock)

EL código de R para ingresar datos cualitativos es como sigue:

```
especies <- c("F", "H", "F", "C", "F", "A", "H", "F",
              "H", "C", "A", "C", "F", "H", "H", "H",
              "F", "H", "A", "C", "F", "H", "H", "F")
```

- Construya una tabla de frecuencias
- Determina la frecuencia relativa de cada especie
- ¿Qué especie presenta mayor proporción (Anote su %)?

Cuadro 3: Frecuencia absoluta y relativa de las especies registrada en un sitio permanente de muestreo en la Comumbia Británica.

Especies	F absoluta	F relativa
A		
C		
F		
H		

### Ejercicio 3

Los datos pueden descargarse del servidor de *dropbox* utilizando la paquetería *repmis* utilizando el siguiente código

```
library(repmis)

## Registered S3 method overwritten by 'R.oo':
##   method      from
##   throw.default R.methodsS3

conjunto <- source_data("https://www.dropbox.com/s/hmsf07bbayxv6m3/cuadro1.csv?dl=1")

## Downloading data from: https://www.dropbox.com/s/hmsf07bbayxv6m3/cuadro1.csv?dl=1

## SHA-1 hash of the downloaded data file is:
## 2bdde4663f51aa4198b04a248715d0d93498e7ba
```

- Encontrar la frecuencia absoluta para la variable *vecinos* y *Especie*

Cuadro 4: Distribución de frecuencias bivariada (Vecinos y Especie)

Vecinos	Cedro	Douglas-fir	Hemlock	Total
0				
1				
2				
3				
4				
5				
6				
Total				

### Ejercicio 4

Con el conjunto de datos anterior (n=50) determinar la frecuencia para la variable Altura (observar el ejemplo en la variable diámetro) Intervalos definidos= 7.5 - 22.5.

1. Extraer los datos de Diametro en la variable *dbh*
2. Determinar el rango de la variable *dbh*

```
dbh <- conjunto$Diametro
range(dbh)
```

```
## [1] 7.7 22.7
```

3. Divida el rango en subintervalos no superpuestos definiendo una secuencia de puntos de ruptura de igual distancia. Si redondeamos los puntos finales del intervalo (7.7, 22.7). Iniciamos con 7.5 y terminamos en 25.5.

```
Intervalo <- seq(7.5, 25.5, by=5)
Intervalo
```

```
## [1] 7.5 12.5 17.5 22.5
```

```
dbh.table <- cut(dbh, Intervalo)
table(dbh.table)
```

```
## dbh.table
## (7.5,12.5] (12.5,17.5] (17.5,22.5]
##          6          27          16
```

```
dbh.prop <- cbind(table(dbh.table))
dbh.per <- round(prop.table(dbh.prop)*100,2)
```

Cuadro 5: Frecuencias determinada por intervalos definidos por el usuario

Intervalo	F. absoluta	F. relativa (%)
7.5-12.5		
12.5-17.5		
17.5-22.5		

## Histogramas: hist(data)

### Ejercicio 5

Con la variable *Diametro* crear el histograma con los datos sin procesar (Figura A). Crear un intervalo de datos menor usando la secuencia: *seq(7.5, 25.5, by=2.5)*.

```
## [1] 7.5 10.0 12.5 15.0 17.5 20.0 22.5 25.0
```

Arbol	Fecha	Especie	Clase	Vecinos	Diametro	Altura
1	12	F	C	4	15.3	14.78
2	12	F	D	3	17.8	17.07
3	9	C	D	5	18.2	18.28
4	9	H	S	4	9.7	8.79
5	7	H	I	6	10.8	10.18
6	10	C	I	3	14.1	14.90

Ambos histogramas deberán ser similares a la figura 2.

```
par(mfrow=c(1,2))  
hist(conjunto$Diametro, main = "Sin modificar", xlab="Diámetro")  
hist(conjunto$Diametro, breaks = inter, main = "Datos Intervalos")  
par(mfrow=c(1,1))
```

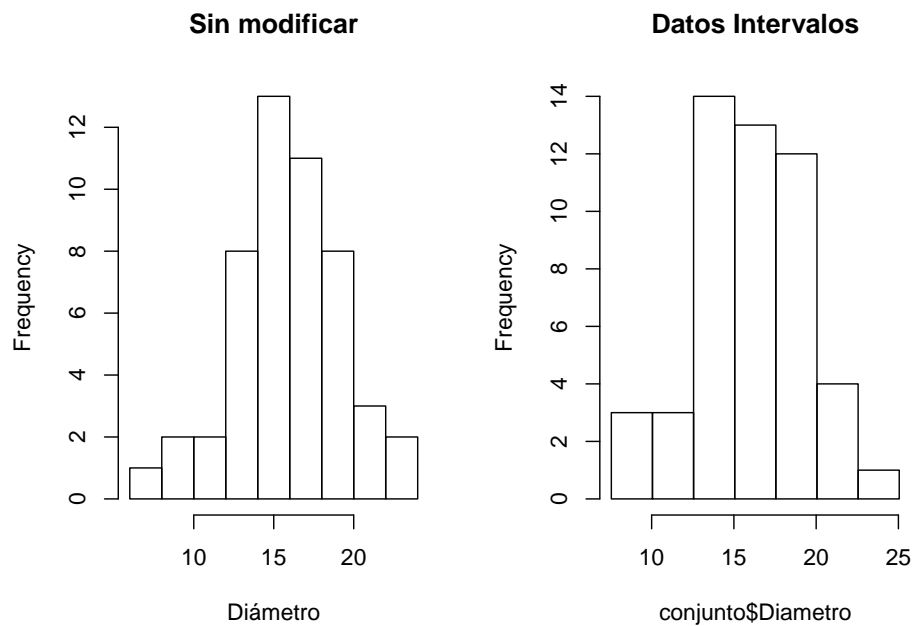


Figura 2: Histogramas

## Referencias

Kozak, A, RA Kozak, CL Staudhammer, and SB Watte. 2008. *Introductory Probability & Statistics. Applications for Forestry & the Natural Sciences*. Cambridge: Cambridge University Press.