

TALLER I – II PARCIAL

NRC 6307

Odalys Benavidez

Melany Polo

Jonathan Cecrado

Shadem Grijalva

Damián Quishpe

5/8/2020

Librerías A utilizar

```
# install.packages("pracma")
# install.packages("xlsx")
# install.packages("fdth")
# install.packages("prettyR")
# install.packages("gmodels")
# install.packages("tidyverse")
# install.packages("gridExtra")
# install.packages("funModeling")
# install.packages("reshape")

library(pracma)
library(fdth)
library(prettyR)
library(ggplot2)
library(gmodels)
library(tidyverse)
library(gridExtra)
library(funModeling)
library(reshape)
```

Ejercicio 1

Crear la base de datos con el nombre (datos.estudiantes) con las variables: Altura, Peso, Número de hermanos, Sexo y Ciudad de nacimiento, generar vectores con por lo menos 30 datos por cada variables, obtener tablas de frecuencias absolutas y relativas de las variables cualitativas, y genere una tabla doble con las variables indicadas, elabore un diagrama de barras y el respectivo pie. Con las variables cuantitativas elabore las tablas de frecuencias absolutas y relativas, construya los gráficos respectivos y determine las medidas de tendencia central y de dispersión as como genere un boxplot por cada variable. Interprete los resultados.

Creación de la base de datos

```

set.seed(1234)
altura <- round(runif(50,min = 1.5,max = 2),2)

set.seed(1234)
peso <- round(runif(50,min = 50,max = 80),1)

hermanos <-sample(0:4, size=50, replace= T, prob= NULL )

set.seed(1234)
sexo <- factor(sample(1:2, size=50, replace= TRUE, prob= NULL ),
               levels = c(1,2),labels = c("Masculino","Femenino"))

set.seed(1234)
ciudad <- factor(sample(1:6, size=50, replace= TRUE, prob= NULL ),
                 levels = c(1:6),labels = c("Quito","Guayaquil","Cuenca",
                                             "Ambato","Riobamba","Manta"))

#Creación de la base de datos
datos.estudiantes <- data.frame(altura,peso,hermanos,sexo,ciudad)
datos.estudiantes

```

##	altura	peso	hermanos	sexo	ciudad
## 1	1.56	53.4	3	Femenino	Ambato
## 2	1.81	68.7	1	Femenino	Guayaquil
## 3	1.80	68.3	2	Femenino	Manta
## 4	1.81	68.7	1	Femenino	Riobamba
## 5	1.93	75.8	4	Masculino	Ambato
## 6	1.82	69.2	0	Femenino	Quito
## 7	1.50	50.3	2	Masculino	Riobamba
## 8	1.62	57.0	0	Masculino	Manta
## 9	1.83	70.0	4	Masculino	Ambato
## 10	1.76	65.4	0	Femenino	Guayaquil
## 11	1.85	70.8	0	Femenino	Manta
## 12	1.77	66.3	1	Femenino	Guayaquil
## 13	1.64	58.5	0	Femenino	Manta
## 14	1.96	77.7	2	Masculino	Manta
## 15	1.65	58.8	1	Femenino	Ambato
## 16	1.92	75.1	2	Femenino	Manta
## 17	1.64	58.6	0	Femenino	Manta
## 18	1.63	58.0	2	Masculino	Manta
## 19	1.59	55.6	0	Femenino	Ambato
## 20	1.62	57.0	1	Femenino	Ambato
## 21	1.66	59.5	4	Femenino	Riobamba
## 22	1.65	59.1	0	Femenino	Ambato
## 23	1.58	54.8	2	Femenino	Cuenca
## 24	1.52	51.2	2	Femenino	Ambato
## 25	1.61	56.6	1	Femenino	Riobamba
## 26	1.91	74.3	4	Femenino	Guayaquil
## 27	1.76	65.8	1	Masculino	Riobamba
## 28	1.96	77.4	3	Femenino	Guayaquil
## 29	1.92	74.9	3	Femenino	Manta
## 30	1.52	51.4	0	Femenino	Cuenca
## 31	1.73	63.7	4	Masculino	Ambato
## 32	1.63	58.0	2	Femenino	Ambato

```
## 33  1.65 59.1      4 Masculino  Cuenca
## 34  1.75 65.2      2 Masculino  Quito
## 35  1.59 55.4      3 Masculino  Cuenca
## 36  1.88 72.8      3 Femenino  Manta
## 37  1.60 56.0      0 Masculino  Ambato
## 38  1.63 57.8      3 Femenino  Guayaquil
## 39  2.00 79.8      3 Femenino  Cuenca
## 40  1.90 74.2      2 Masculino  Guayaquil
## 41  1.78 66.6      0 Femenino  Riobamba
## 42  1.82 69.4      3 Masculino  Manta
## 43  1.66 59.4      0 Femenino  Quito
## 44  1.81 68.7      1 Femenino  Manta
## 45  1.66 59.9      4 Femenino  Cuenca
## 46  1.75 65.1      4 Masculino  Manta
## 47  1.84 70.3      4 Masculino  Quito
## 48  1.74 64.5      0 Masculino  Riobamba
## 49  1.62 57.3      1 Masculino  Quito
## 50  1.88 73.0      1 Femenino  Quito
```

```
glimpse(datos.estudiantes)
```

```
## Rows: 50
## Columns: 5
## $ altura    <dbl> 1.56, 1.81, 1.80, 1.81, 1.93, 1.82, 1.50, 1.62, 1.83, 1.76...
## $ peso      <dbl> 53.4, 68.7, 68.3, 68.7, 75.8, 69.2, 50.3, 57.0, 70.0, 65.4...
## $ hermanos  <int> 3, 1, 2, 1, 4, 0, 2, 0, 4, 0, 0, 1, 0, 2, 1, 2, 0, 2, 0, 1...
## $ sexo      <fct> Femenino, Femenino, Femenino, Femenino, Masculino, Femenin...
## $ ciudad    <fct> Ambato, Guayaquil, Manta, Riobamba, Ambato, Quito, Riobamb...
```

Análisis de las variables Cualitativas

```
attach(datos.estudiantes)
freq1 <- datos.estudiantes %>%
  group_by(sexo) %>%
  summarise(Frecuencia = n()) %>%
  mutate(Frec.Relativa=Frecuencia/sum(Frecuencia))%>%
  mutate(Frec.Acumulada = cumsum(Frecuencia))%>%
  mutate(Frec.Acumulada = cumsum(Frecuencia))%>%
  mutate(Frec.Relativa_Acum=Frec.Acumulada/sum(Frecuencia))
freq1
```

```
## # A tibble: 2 x 5
##   sexo      Frecuencia Frec.Relativa Frec.Acumulada Frec.Relativa_Acum
##   <fct>      <int>      <dbl>      <int>      <dbl>
## 1 Masculino      18      0.36      18      0.36
## 2 Femenino      32      0.64      50      1
```

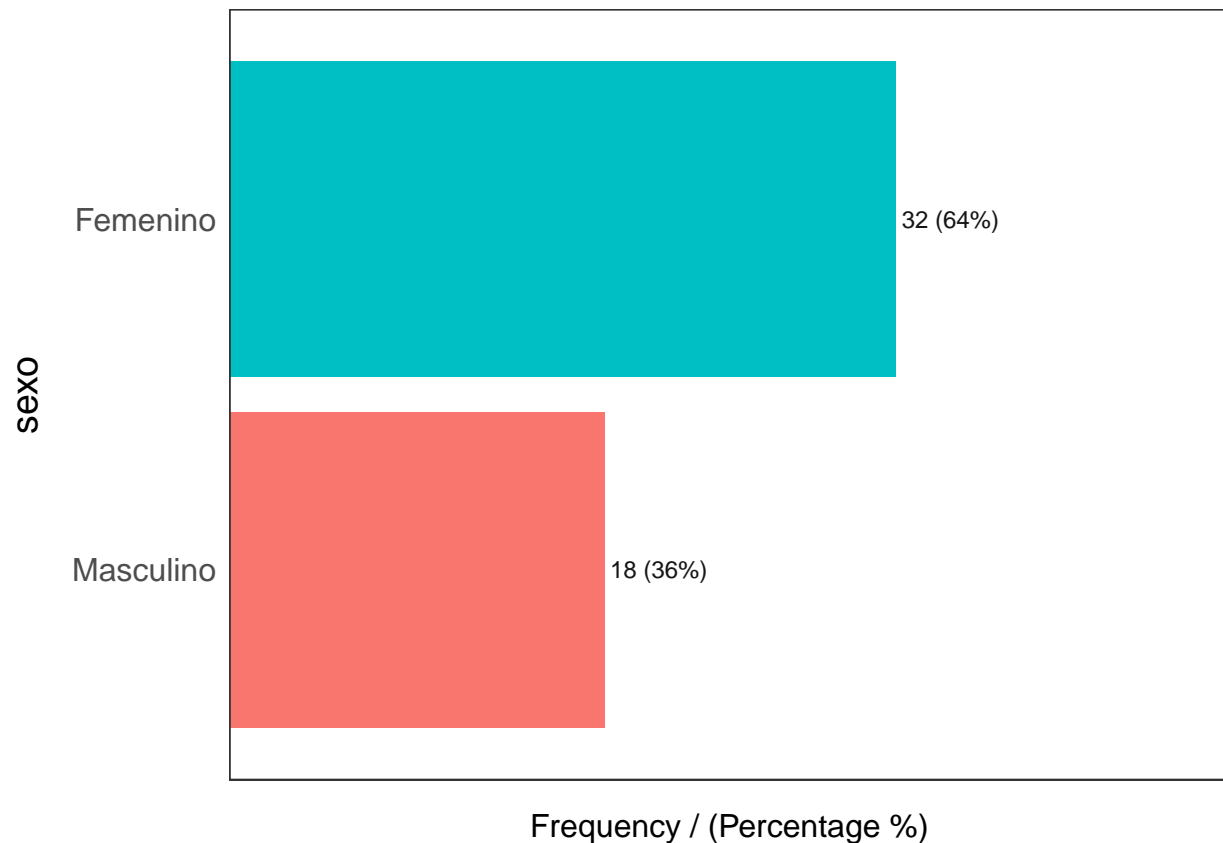
```
freq2 <- datos.estudiantes %>%
  group_by(ciudad) %>%
  summarise(Frecuencia = n()) %>%
```

```
mutate(Frec.Relativa=Frecuencia/sum(Frecuencia))%>%
mutate(Frec.Acumulada = cumsum(Frecuencia))%>%
mutate(Frec.Acumulada = cumsum(Frecuencia))%>%
mutate(Frec.Relativa_Acum=Frec.Acumulada/sum(Frecuencia))
freq2
```

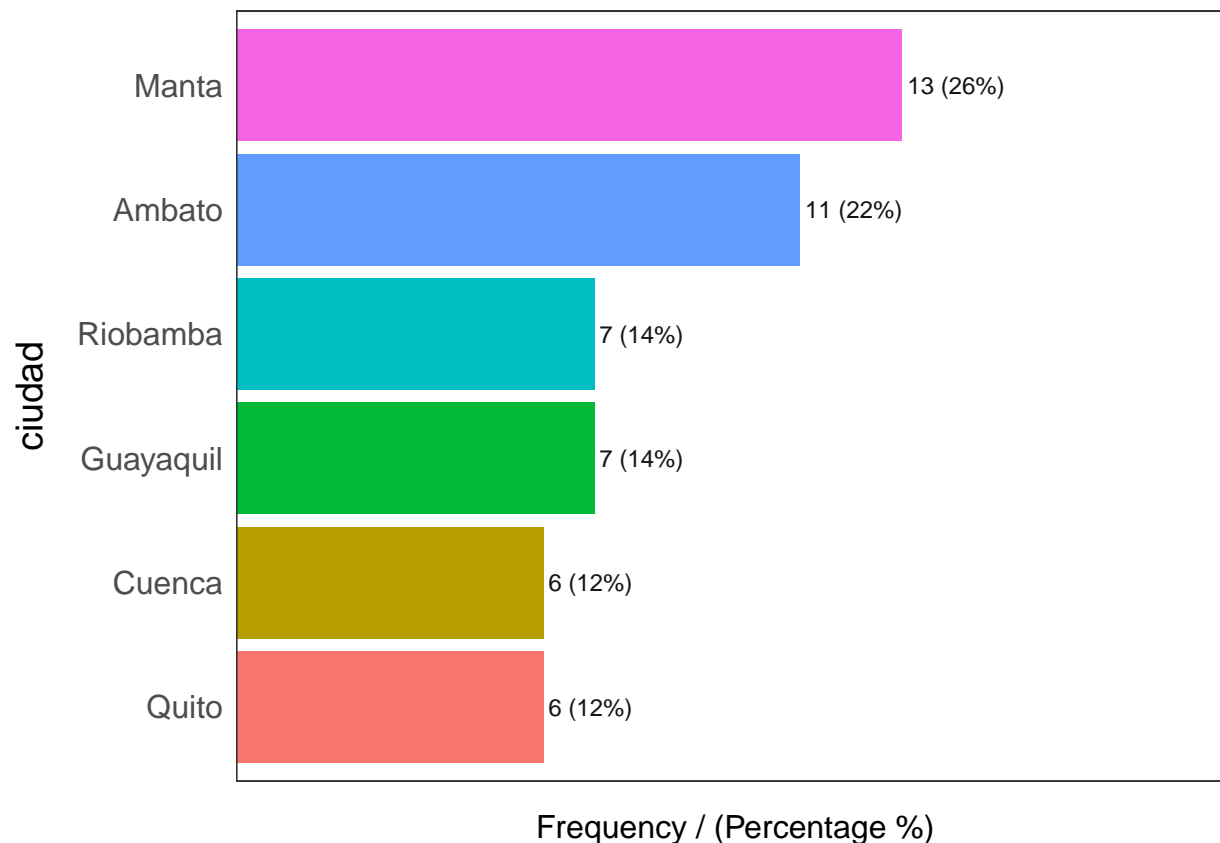
```
## # A tibble: 6 x 5
##   ciudad      Frecuencia Frec.Relativa Frec.Acumulada Frec.Relativa_Acum
##   <fct>         <int>         <dbl>         <int>         <dbl>
## 1 Quito             6          0.12             6          0.12
## 2 Guayaquil          7          0.14            13          0.26
## 3 Cuenca             6          0.12            19          0.38
## 4 Ambato            11          0.22            30          0.6
## 5 Riobamba           7          0.14            37          0.74
## 6 Manta             13          0.26            50          1
```

#Gráficos

```
freq(datos.estudiantes)
```



```
##      sexo frequency percentage cumulative_perc
## 1 Femenino      32         64             64
## 2 Masculino     18         36            100
```

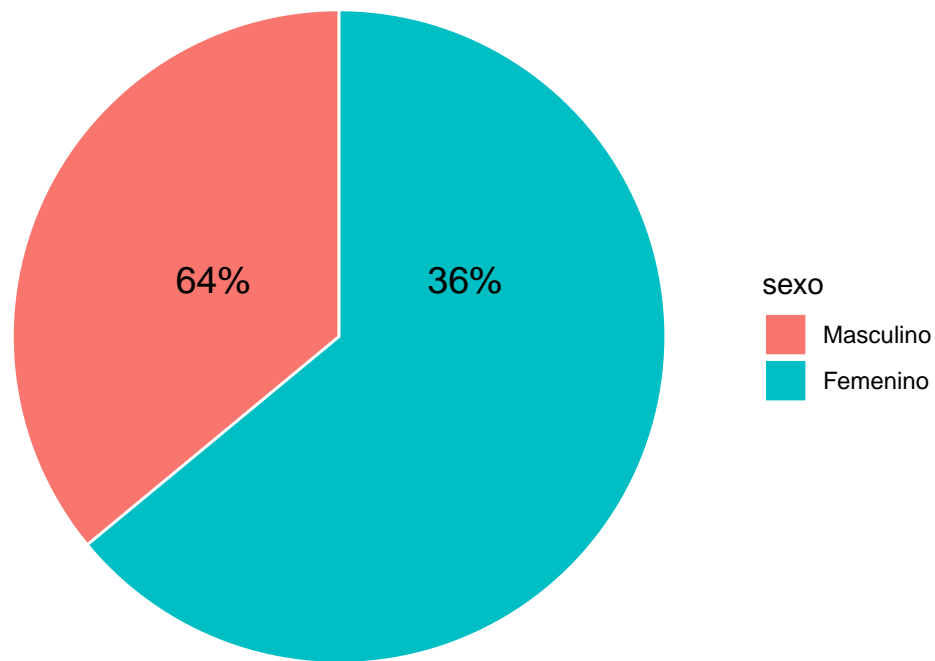


```
##      ciudad frequency percentage cumulative_perc
## 1      Manta         13         26             26
## 2      Ambato         11         22             48
## 3  Guayaquil          7         14             62
## 4      Riobamba        7         14             76
## 5        Quito         6         12             88
## 6        Cuenca         6         12            100
```

```
## [1] "Variables processed: sexo, ciudad"
```

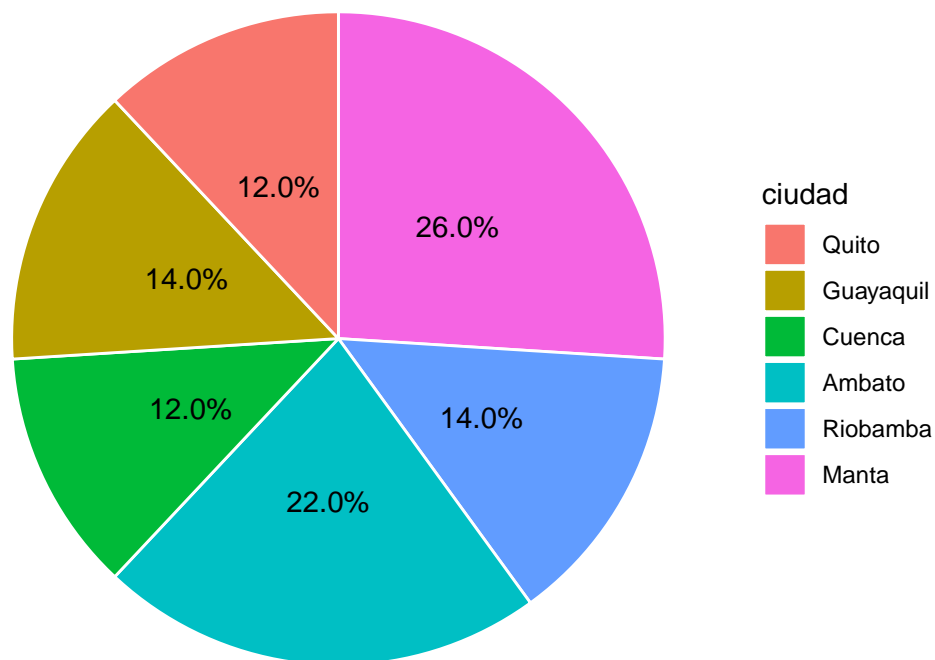
```
ggplot(freq1, aes(x="", y=Frec.Relativa, fill=sexo)) +
  geom_bar(stat="identity", width=1, color="white") +
  coord_polar("y", start=0) +
  theme_void() +
  geom_text(aes(label=scales::percent(Frec.Relativa)), vjust=-5.5, size=5) +
  labs(title="Gráfico Pie chart de la variable sexo")
```

Gráfico Pie chart de la variable sexo



```
posicion <- freq2 %>%  
  arrange(desc(ciudad)) %>%  
  mutate(lab.ypos = cumsum(Frec.Relativa) - 0.5*Frec.Relativa)  
  
ggplot(posicion, aes(x="", y=Frec.Relativa, fill=ciudad)) +  
  geom_bar(stat="identity", width=1, color="white") +  
  coord_polar("y", start=0) +  
  geom_text(aes(y=lab.ypos, label = scales::percent(Frec.Relativa)), color = "black")+  
  labs(title="Gráfico Pie chart de la variable ciudad de nacimiento")+  
  theme_void()
```

Gráfico Pie chart de la variable ciudad de nacimiento



El aula NRC 6307 está conformado por 32 mujeres y 18 hombres, que representa el 64% y 36% respectivamente

El aula está conformado en su mayoría por estudiantes de la ciudad de Manta, que representa el 26%, le sigue la ciudad de Ambato con 22%, ambas acumulan el 48% de los estudiantes. El restante 52% se encuentra distribuido en las ciudades de: Guayaquil y Riobamba con 7 estudiantes cada una alcanzado conjuntamente un 28%, y las Ciudades de Quito y Cuenca con un 12% cada una.

```
#Tabla cruzada
table1 <- CrossTable(ciudad, sexo, prop.chisq = FALSE)
```

```
##
##
##   Cell Contents
## |-----|
## |               N |
## |   N / Row Total |
## |   N / Col Total |
## |   N / Table Total |
## |-----|
##
##
## Total Observations in Table:  50
##
##
##           | sexo
```

##	ciudad	Masculino	Femenino	Row Total
##	Quito	3	3	6
##		0.500	0.500	0.120
##		0.167	0.094	
##		0.060	0.060	
##	Guayaquil	1	6	7
##		0.143	0.857	0.140
##		0.056	0.188	
##		0.020	0.120	
##	Cuenca	2	4	6
##		0.333	0.667	0.120
##		0.111	0.125	
##		0.040	0.080	
##	Ambato	4	7	11
##		0.364	0.636	0.220
##		0.222	0.219	
##		0.080	0.140	
##	Riobamba	3	4	7
##		0.429	0.571	0.140
##		0.167	0.125	
##		0.060	0.080	
##	Manta	5	8	13
##		0.385	0.615	0.260
##		0.278	0.250	
##		0.100	0.160	
##	Column Total	18	32	50
##		0.360	0.640	
##				
##				
##				

El aula NRC 6307 está conformada en un 48% por estudiantes de las ciudades de: Manta y Ambato. Los alumnos de la Ciudad de Mannta se dividen en 61.5% por mujeres y un 38.5% de hombres. Cabe señalar que la prevalencia del sexo femenino en el aula es notoria con una 64%.

Análisis de variables cuantitativas

```
freq3 <- fdt(altura, start=1.5, end=2.0, h=0.1)
freq3
```

```
## Class limits  f  rf rf(%) cf cf(%)
## [1.5,1.6)  7 0.14  14  7  14
## [1.6,1.7) 16 0.32  32 23  46
## [1.7,1.8)  8 0.16  16 31  62
```



```
##      [1.8,1.9) 11 0.22    22 42    84
##      [1.9,2)   7 0.14    14 49    98
```

```
freq4 <- fdt(peso, start=50, end=80, h=5)
freq4
```

```
## Class limits  f   rf rf(%) cf cf(%)
##      [50,55)  5 0.10    10  5    10
##      [55,60) 18 0.36    36 23    46
##      [60,65)  2 0.04     4 25    50
##      [65,70) 12 0.24    24 37    74
##      [70,75)  8 0.16    16 45    90
##      [75,80)  5 0.10    10 50   100
```

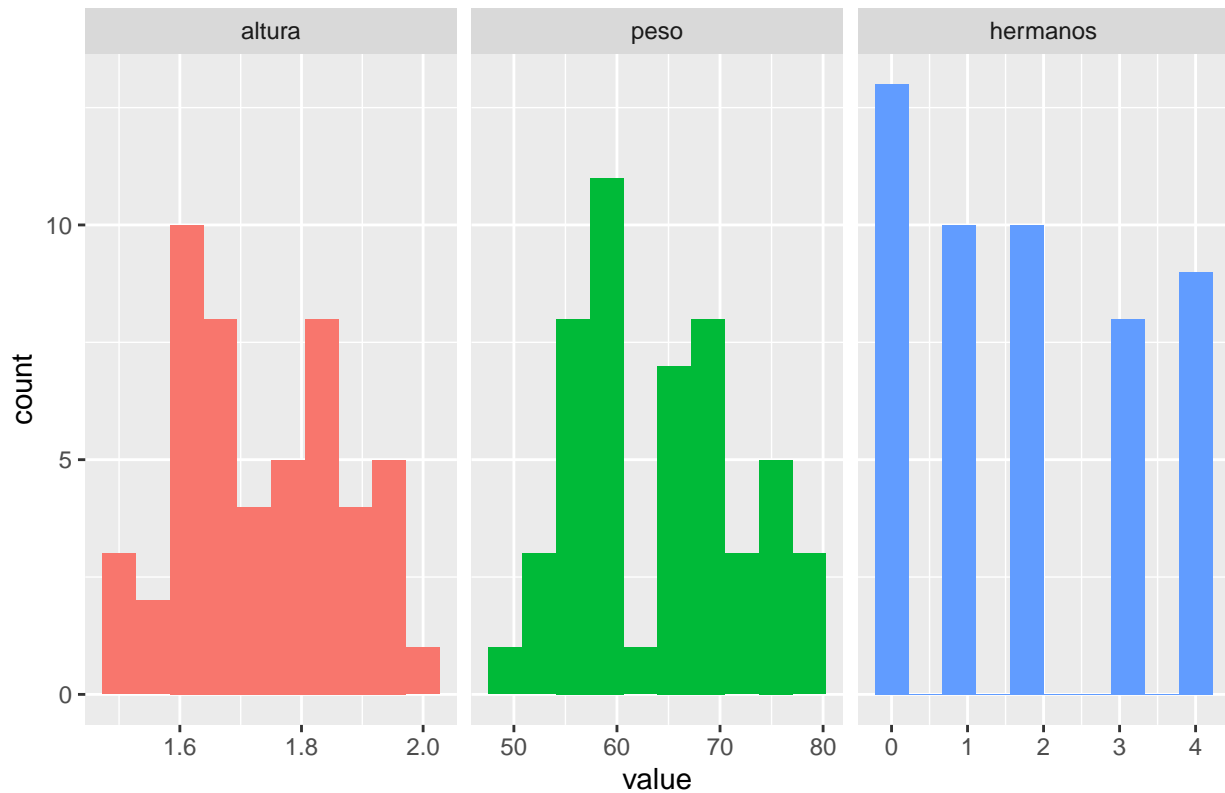
```
freq5 <- fdt(hermanos, start=0, end=5, h=1)
freq5
```

```
## Class limits  f   rf rf(%) cf cf(%)
##      [0,1) 13 0.26    26 13    26
##      [1,2) 10 0.20    20 23    46
##      [2,3) 10 0.20    20 33    66
##      [3,4)  8 0.16    16 41    82
##      [4,5)  9 0.18    18 50   100
```

```
# Función para crear los gráficos
plot1 <- function (data, bins = 10, path_out = NA)
{
  wide_data = suppressMessages(melt(data))
  p = ggplot(data = wide_data, mapping = aes(x = value)) +
    geom_histogram(bins = bins, na.rm = T) +
    facet_wrap(~variable,
               scales = "free_x") + aes(fill = variable) + guides(fill = FALSE)+
    labs(title = "Histogramas de las variables cuantitativas")
  if (!is.na(path_out)) {
    export_plot(p, path_out, "histograms")
  }
  plot(p)
}

plot1 <- plot1(datos.estudiantes)
```

Histogramas de las variables cuantitativas



```
# Medidas de tendennncia central y dispersión
profiling_num(datos.estudiantes)
```

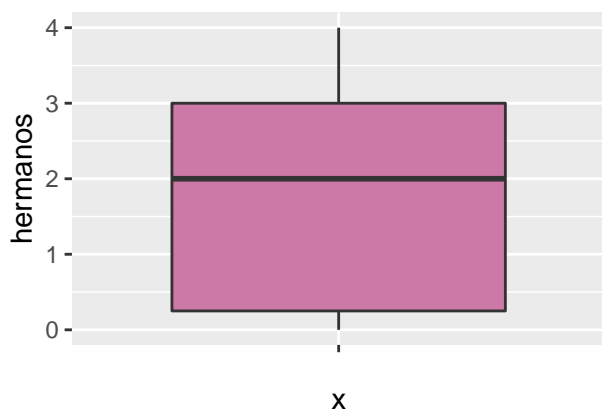
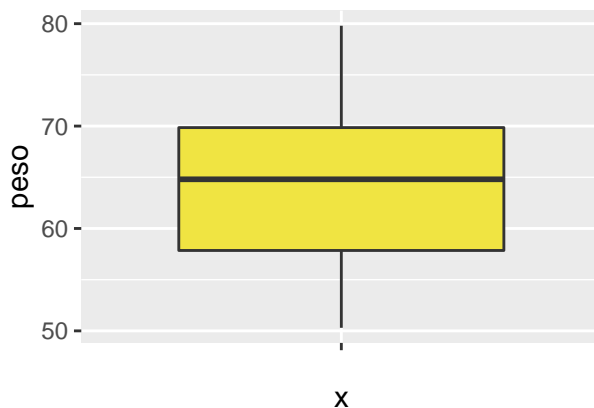
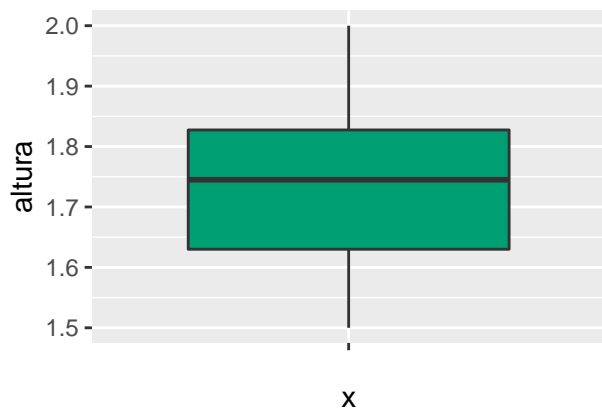
```
##  variable    mean  std_dev variation_coef    p_01    p_05  p_25  p_50  p_75
## 1  altura  1.7344 0.1324211    0.0763498  1.5098  1.538  1.63  1.745  1.8275
## 2  peso  64.0880 7.9102785    0.1234284 50.7410 52.300 57.85 64.800 69.8500
## 3 hermanos 1.8000 1.4568627    0.8093682  0.0000  0.000  0.25  2.000  3.0000
##      p_95  p_99  skewness kurtosis    iqr      range_98      range_80
## 1  1.9465  1.9804 0.1787503 1.940980  0.1975 [1.5098, 1.9804] [1.589, 1.92]
## 2 76.6800 78.7710 0.1705408 1.916352 12.0000 [50.741, 78.771] [55.34, 74.92]
## 3  4.0000  4.0000 0.1920116 1.701183  2.7500          [0, 4]          [0, 4]
```

```
p1 <- ggplot(datos.estudiantes,aes(x="",y=altura))+
  geom_boxplot(fill= "#009E73")

p2 <- ggplot(datos.estudiantes,aes(x="",y=peso,fill="blue"))+
  geom_boxplot(fill= "#F0E442")

p3 <- ggplot(datos.estudiantes,aes(x="",y=hermanos,fill="green"))+
  geom_boxplot(fill= "#CC79A7")

grid.arrange(p1,p2,p3, widths= c(2,2))
```



Altura

Los resultados nos muestran que la estatura se encuentra concentrada en el intervalo de 1.6 y 1.7 metros aproximadamente, esto se ve reflejado en el valor promedio y la mediana que son: 1.73 y 1.74 metros. Los datos no muestran un comportamiento de sesgo, al no estar tan dispersos, pues su desviación estándar apenas es de 0.13. Además, el diagrama de caja y bigotes nos permite observar que no existen valores atípicos en la altura.

Peso

El peso promedio de los estudiantes es de 64.08 kg, esta medida nos puede dar un indicio de la presencia de sobrepeso en el aula, es decir si se compara con la altura promedio para obtener un índice de masa corporal (IMC) promedio del aula se obtiene un 21.3 como resultado, que se encuentra dentro de los límites normales. Se obtiene así que, cerca del 50% del aula no sufre problemas de sobrepeso. A pesar de esto cabe señalar que existe una desviación estándar de 7.91, y esta desviación se ve reflejado en el gráfico que nos muestra una alta concentración de personas que tienen un peso sobre los 65 kg y que podrían sufrir problemas de sobrepeso.

Número de hermanos

El número de hermanos promedio es de 2 aproximadamente, en esta variable no se encuentran valores que estén por fuera de lo normal, además se puede observar que posee una distribución muy uniforme.

Ejercicio 2

Se analiza una muestra de 25 pacientes, si la probabilidad de que tenga una determinada enfermedad es del 7%, calcule la probabilidad que:

a) Ninguno de los pacientes tenga esa enfermedad

```
n2a <- 25
p2a <- 0.07
x2a <- 0
dbinom(x2a,n2a,p2a)
```

```
## [1] 0.1629573
```

b) De que entre 15 y 20 pacientes adolezcan de la enfermedad

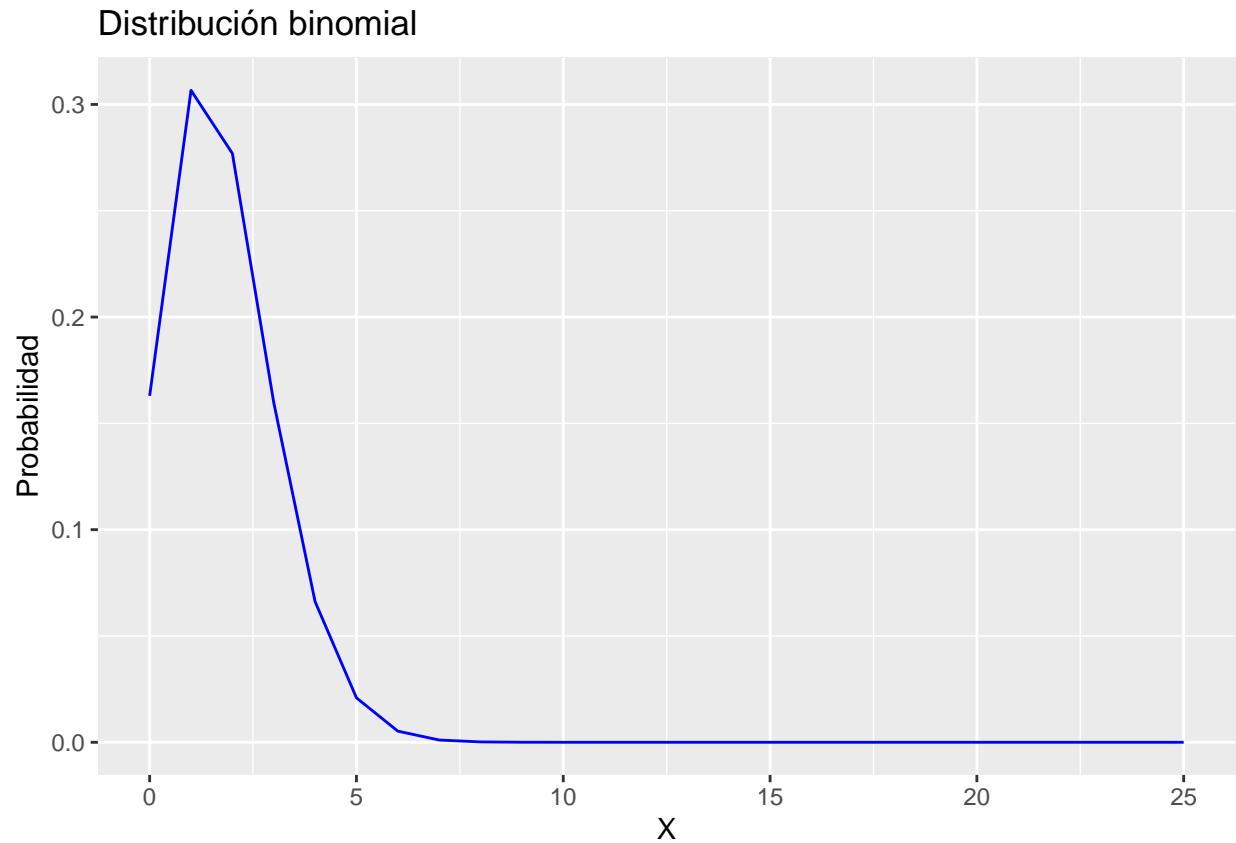
```
n2b <- 25
p2b <- 0.07
x2b <- c(14,20)
pbinom(x2b,n2b,p2b)[2]-pbinom(x2b,n2b,p2b)[1]
```

```
## [1] 7.878587e-12
```

Graficar la distribución de probabilidad

```
x_bin <- seq(0,25, by=1)
y_dbin <- dbinom(x_bin, size = 25, prob=0.07)
distr_binom <- data.frame(x_bin, y_dbin)

#Visualización de la distribución binomial
ggplot(distr_binom, aes(x=x_bin, y=y_dbin)) +
  geom_line(colour="blue")+
  ggtitle("Distribución binomial")+
  labs(x="X",
       y="Probabilidad")
```



Ejercicio 3

Supongamos que $\mu = 20$, por lo que el número esperado de entradas en nuestra página web en una hora es 20. Calcule la probabilidad de que:

a) Nadie ingrese a la página web

```
u3a <- 20
x3a <- 0
dpois(x3a,u3a)
```

```
## [1] 2.061154e-09
```

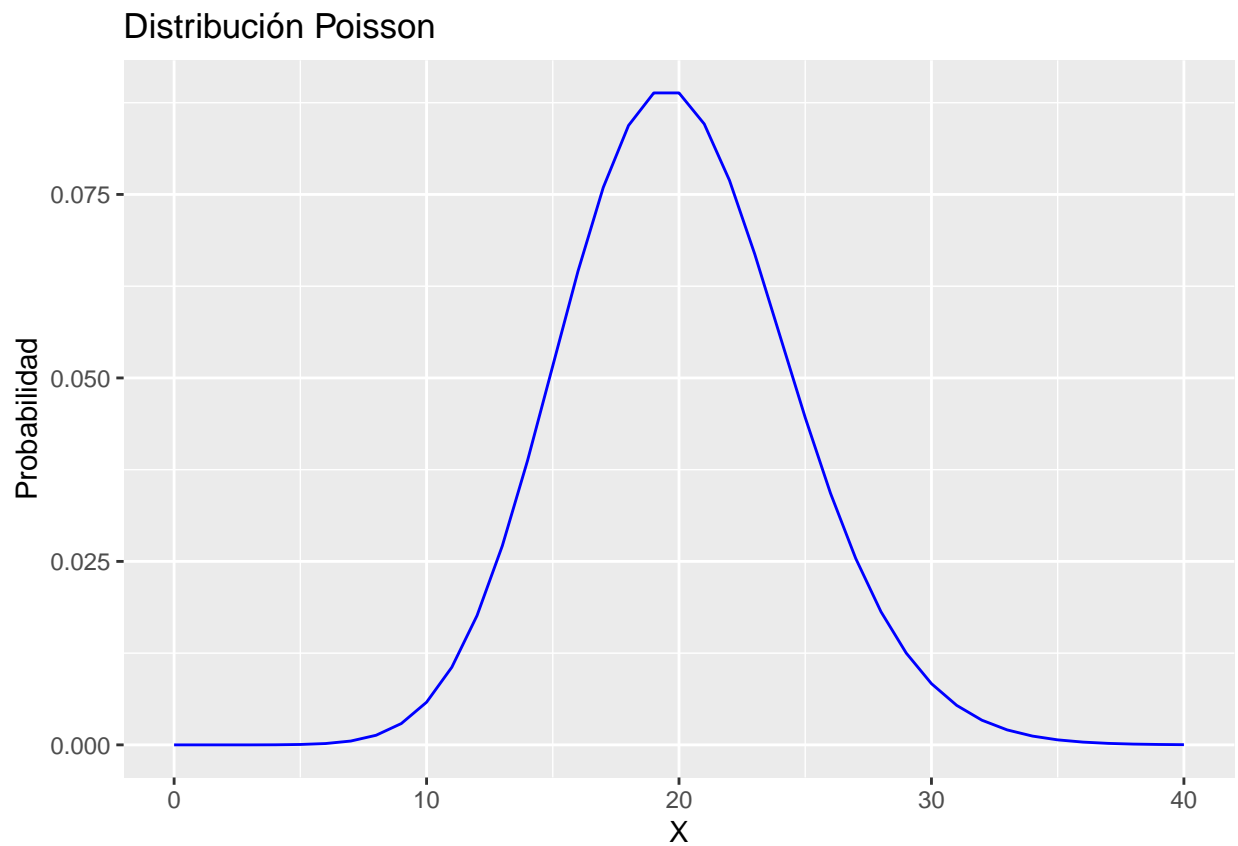
b) De que haya entre 25 y 40 entradas

```
u3b <- 20
x3b <- c(24,40)
ppois(x3b,u3b)[2] - ppois(x3b,u3b)[1]
```

```
## [1] 0.1567472
```

c) Graficar la distribución de probabilidad

```
x_pois <- seq(0,40, by=1)
y_dpois <- dpois(x_pois, lambda = 20)
distr_pois <- data.frame(x_pois, y_dpois)
#Visualización de la distribución Poisson
ggplot(distr_pois, aes(x=x_pois, y=y_dpois)) +
  geom_line(colour="blue")+
  ggtitle("Distribución Poisson")+
  labs(x="X",
       y="Probabilidad")
```



Ejercicio 4

Una población consta de 15 elementos, 10 de los cuales son aceptables. En una muestra de 4 elementos, ¿cuál es la probabilidad de que exactamente 3 sean aceptables? Suponga que las muestras se toman sin reemplazo.

```
n4 <- 15
m4 <- 10
k4 <- 4
x4 <- 3
dhyper(x4,m4 ,n4-m4, k4)
```

```
## [1] 0.4395604
```

a) Ninguno se aceptable

```
n4a <- 15
m4a <- 10
k4a <- 4
x4a <- 0
dhyper(x4a,m4a ,n4a-m4a, k4a)
```

```
## [1] 0.003663004
```

b) De que por lo menos 3 sean aceptables

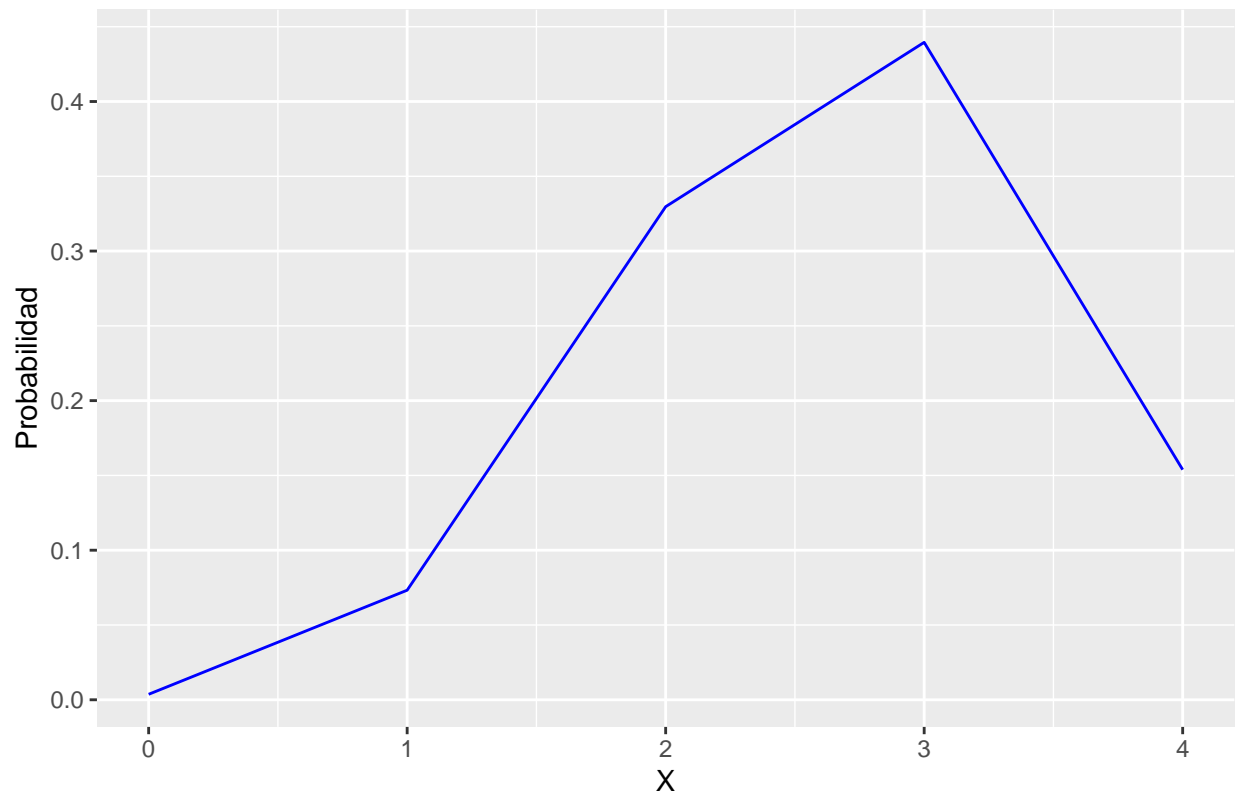
```
n4b <- 15
m4b <- 10
k4b <- 4
x4b <- c(3,4)
sum(dhyper(x4b,m4b ,n4b, k4b))
```

```
## [1] 0.1588933
```

c) Graficar la distribución de probabilidad

```
x_hiper <- seq(0,4, by=1)
y_dhiper <- dhyper(x_hiper,10,15-10,4)
distr_hiper <- data.frame(x_hiper, y_dhiper)
#Visualización de la distribución Hipergeométrica
ggplot(distr_hiper, aes(x=x_hiper, y=y_dhiper)) +
  geom_line(colour="blue")+
  ggtitle("Distribución Hipergeométrica")+
  labs(x="X",
       y="Probabilidad")
```

Distribución Hipergeométrica



#Ejercicio 5

El diámetro de los puntos producidos por una impresora matricial tiene distribución normal con un diámetro promedio de 0.002 pulgadas y una desviación estándar de 0.0004 pulgadas

a) Cuál es la probabilidad de que el diámetro de un punto exceda 0,0026 pulgadas.

```
u5a <- 0.002
s5a <- 0.0004
x5a <- 0.0026
pnorm(x5a, u5a, s5a)
```

```
## [1] 0.9331928
```

b) Cual es la probabilidad de que el diámetro de un punto mida entre 0.0014 y 0.0026 pulgadas.

```
u5b <- 0.002
s5b <- 0.0004
x5b <- c(0.0014, 0.0026)
pnorm(x5b, u5b, s5b)[2] - pnorm(x5b, u5b, s5b)[1]
```

```
## [1] 0.8663856
```


c) Grafique la distribución de probabilidad.

```
n<-seq(0.0014,0.0026,0.0001)
distr_norm <- data.frame(n=n,
                        distribucion=pnorm(n, mean=0.002, sd=0.0004))
#Visualización de la distribución Normal
ggplot(distr_norm, aes(x=n, y=distribucion )) +
  geom_line(colour="blue") +
  ggtitle ("Distribución de probabilidad normal")+
  labs(x="X",
       y="Probabilidad")
```

