

# MAE: ENTREGA I

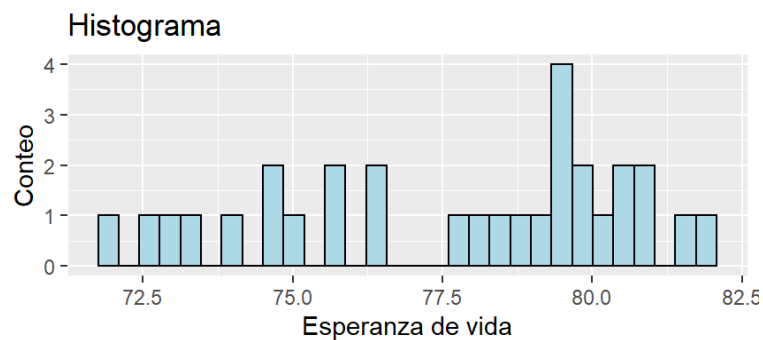
Jon Zorrilla Gamboa

## 1. Ejercicio 1

Para este ejercicio, dejaré el código usado y las figuras obtenidas.

### 1.1.

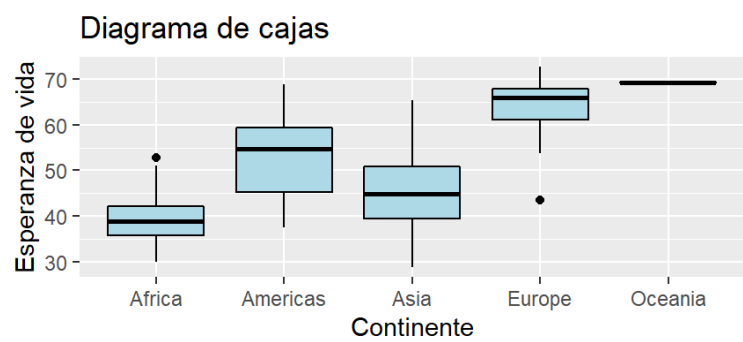
```
install.packages("gapminder")  
library("gapminder")  
library("dplyr")  
library("ggplot2")  
gapminder %>%  
  filter(year == 2007, continent == "Europe") %>%  
  ggplot() +  
  geom_histogram(aes(lifeExp), bins = 30, fill="lightblue", col="black") +  
  labs(x="Esperanza de vida", y="Conteo", title="Histograma")
```



**Figura 1:** Histograma esperanzas de vida de los países de Europa.

## 1.2.

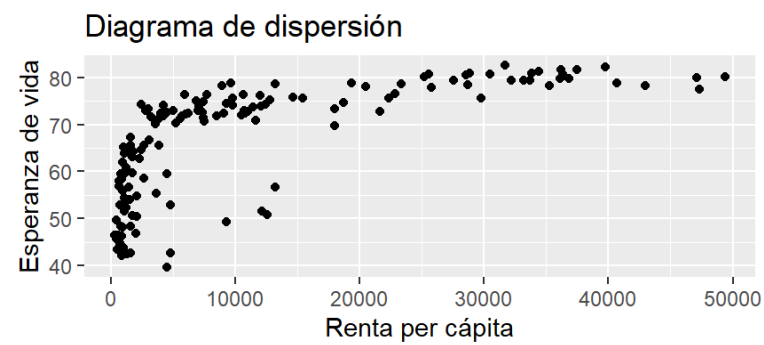
```
gapminder %>%  
filter(year == 1952) %>%  
ggplot() +  
geom_boxplot(aes(x=continent, y=lifeExp), fill="lightblue", col="black") +  
labs(x = "Continente", y = "Esperanza de vida", title = "Diagrama de cajas")
```



**Figura 2:** Diagrama de cajas de esperanzas de vida por continentes.

### 1.3.

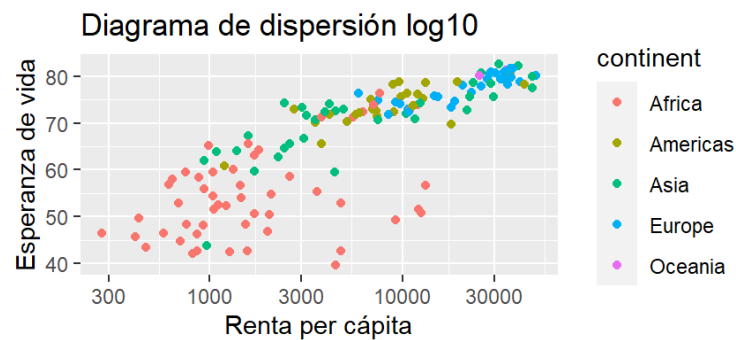
```
gapminder %>%  
filter(year==2007) %>%  
ggplot() +  
geom_point(aes(x=gdpPercap, y=lifeExp)) +  
labs(x = "Renta per cápita", y = "Esperanza de vida", title="Diagrama de dispersión")
```



**Figura 3:** Diagrama de dispersión renta per cápita en función de esperanza de vida.

## 1.4.

```
gapminder %>%  
  filter(year==2007) %>%  
  ggplot() +  
  geom_point(aes(x=gdpPercap, y=lifeExp, col=continent)) +  
  labs(x = "Renta per cápita", y = "Esperanza de vida", title = "Diagrama de dispersión") +  
  scale_x_log10()
```



**Figura 4:** Diagrama de dispersión logarítmico renta per cápita en función de la esperanza de vida para distintos continentes.

## 2. Ejercicio 2

### 2.1.

Dados  $\hat{p}_1$  y  $\hat{p}_2$ , definimos el siguiente estimador:

$$\hat{p} = 0,4 \cdot \hat{p}_1 + 0,6 \cdot \hat{p}_2 \quad (1)$$

$p_1$  y  $p_2$  seguirán distribuciones binomiales. Para demostrar que este estimador es insesgado, hallamos la esperanza:

$$E[\hat{p}] = 0,4E[p_1] + 0,6E[p_2] = 0,4\mu + 0,6\mu = \mu \quad (2)$$

### 2.2.

Calculamos la varianza teniendo en cuenta que  $p_1$  y  $p_2$  siguen distribuciones binomiales, cuya varianza viene determinada por:  $Var(X) = np(1-p)$ . Entonces, dado que estimaremos el número de personas enfermas sobre el número de personas totales:

$$\hat{p} = 0,4 \cdot \frac{\hat{p}_1}{n_1} + 0,6 \cdot \frac{\hat{p}_2}{n_2} \quad (3)$$

La varianza será:

$$Var(\hat{p}) = \frac{0,4^2 Var(p_1)}{n_1^2} + \frac{0,6^2 Var(p_2)}{n_2^2} = \frac{0,4^2 p_1(1-p_1)}{n_1} + \frac{0,6^2 p_2(1-p_2)}{n_2} \quad (4)$$

### 2.3.

Dado  $p_1 = p_2$ , nuestro estimador será el número de enfermos entre el número total de personas. Nuestro estimador será entonces:

$$\hat{p} = \frac{N_{enf}}{N} \quad (5)$$

Donde, como antes, el número de enfermos seguirá una distribución binomial. De esta manera, podemos calcular la varianza como sigue:

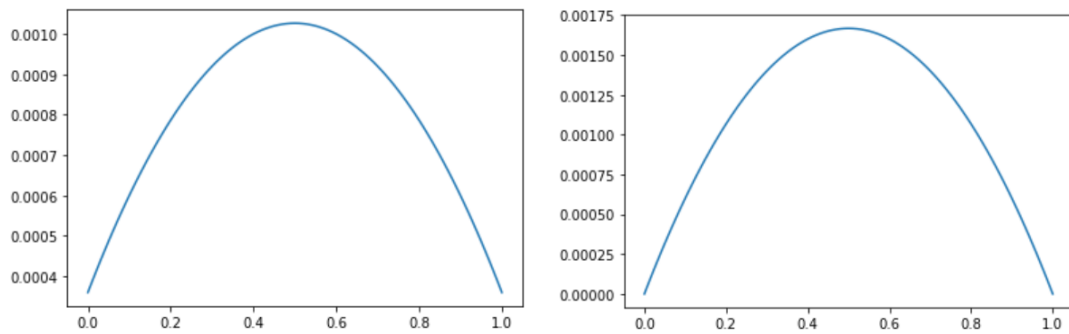
$$Var(\hat{p}) = \frac{Var(N_{enf})}{N^2} = \frac{Np(1-p)}{N^2} = \frac{p(1-p)}{N} \quad (6)$$

Dependiendo de los valores de  $p$ , la varianza del caso estratificado puede ser menor a la varianza del caso simple, donde no se han separado dos clases diferentes.

De hecho, si hacemos  $p_1 = p_2$  en la ecuación correspondiente al apartado anterior, obtendremos la misma ecuación.

## 2.4.

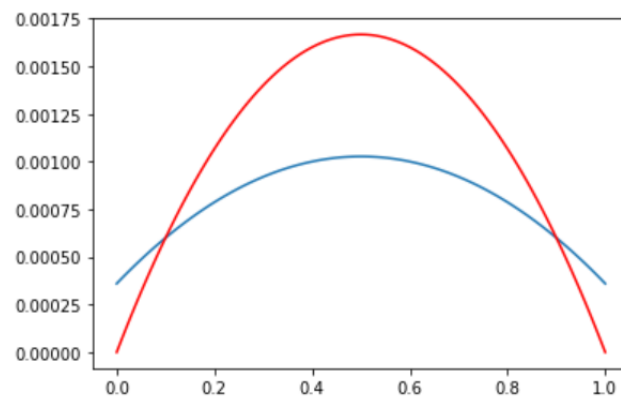
Ahora, representaremos los valores de las varianzas en ambos casos en función de  $p_1$ , dado que  $p_2 = 0,1$ .



(a)  $\vec{E}$  perpendicular al plano de incidencia. (b) Varianza en función de  $p_1$  para el caso estratificado.

**Figura 5:** Varianza en función de  $p_1$  para el caso no estratificado.

De hecho, si hacemos la representación de las dos varianzas simultáneamente:



**Figura 6:** Varianzas en función de  $p_1$  en ambos casos.

Como se puede observar, la varianza del caso simple será mayor para los valores de  $p_1$  mayores a 0.1 y menores a 0.9, por lo que será mejor utilizar el caso estratificado en esos casos, ya que obtendremos una mayor eficiencia.

## 3.

### 3.1.

Para hallar el estimador de máxima verosimilitud, usaremos lo siguiente:

```
optimize(function(theta) -sum(dcauchy(muestra, location=theta, log=TRUE)), c(-100,100))
```

Obteniendo un valor mínimo de 9.842954. Hemos hecho uso de ese intervalo ya que sabemos que este estimador debe encontrarse por el centro de la distribución.

Para aproximar la varianza del estimador de máxima verosimilitud, podemos calcular el estimador muchas veces, y guardando los resultados obtenidos para cada iteración, obtener la varianza.

Usando el siguiente código, obtenemos una varianza  $Var(\theta) = 42,32521$ .

```
set.seed(123)
theta <- 10
n <- 100
resultados <- c()
f <- function(muestra,theta) {
  tmp <- 0
  for (val in muestra) {tmp=tmp+(2*(val-theta))/(1+(val-theta)^2)}
  return(tmp)}
for (i in 1:1000) {
  muestra <-rt(n,1) + theta
  E<- optimize(f,c(-100,100),muestra=muestra, maximum=TRUE)
  resultado_nuevo=E$maximum
  resultados <- c(resultados, resultado_nuevo)
}
Varianza_final <- var(resultados)
Varianza_final
```