

Ejercicio 4.2: Pruebas de homogeneidad de varianzas

Silvia Pineda

Carga de Datos y Librerías

```
library(openintro)
```

Warning: package 'openintro' was built under R version 4.3.3

Loading required package: airports

Loading required package: cherryblossom

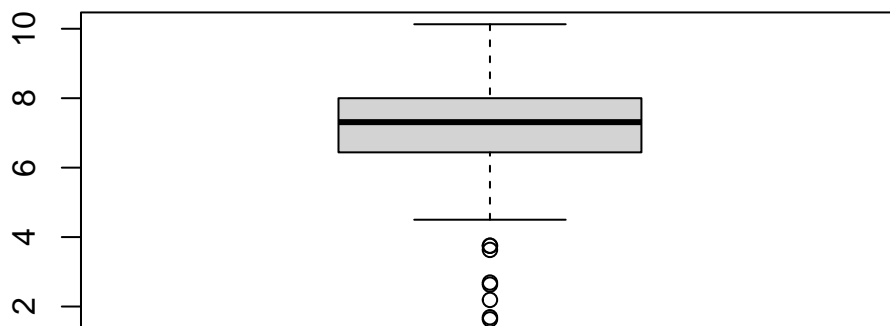
Loading required package: usdata

Warning: package 'usdata' was built under R version 4.3.3

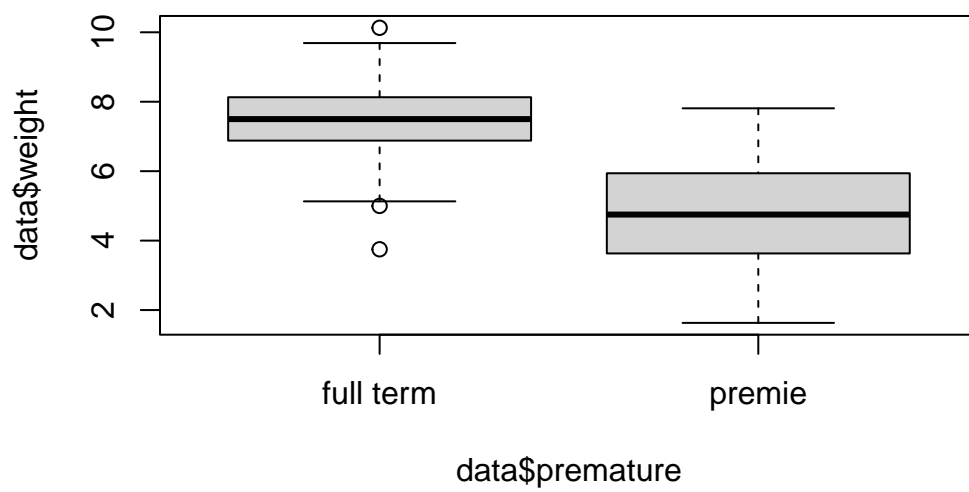
```
data(births)  
data<-births
```

Detectar Outliers

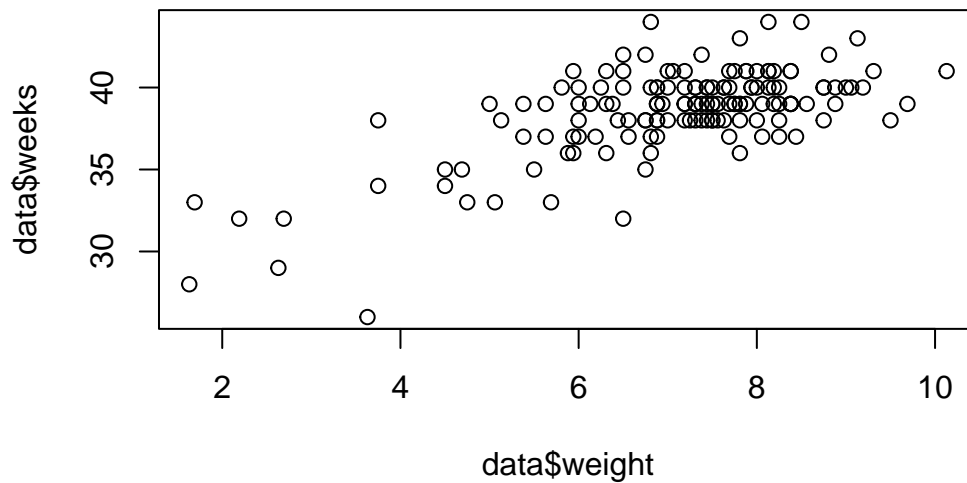
```
boxplot(data$weight)
```



```
boxplot(data$weight~data$premature)
```



```
plot(data$weight,data$weeks)
```

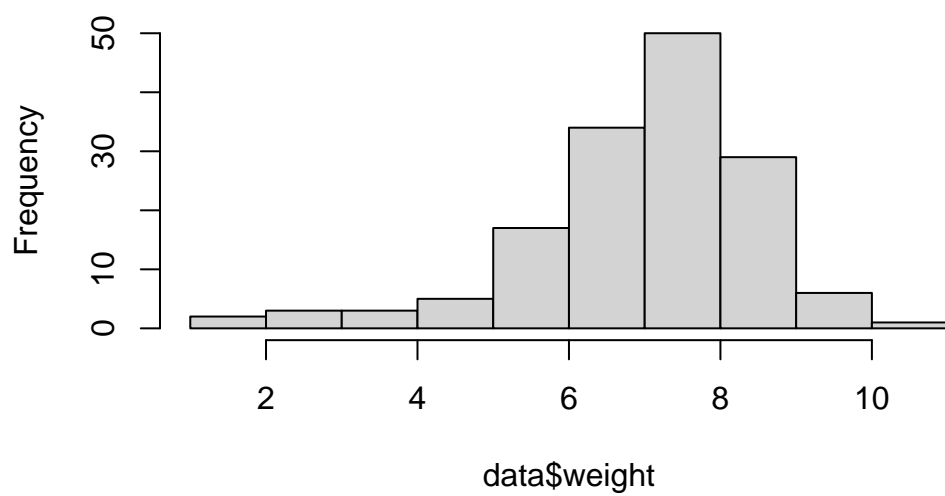


El boxplot de la variable weight muestra varios valores bajos que podrían considerarse outliers, pero cuando lo observamos con la variable premature ya vemos que corresponde a los bebés nacidos de forma prematura. Si lo hacemos con la variable weeks que representa las semanas de gestación, vemos efectivamente que la asociación corresponde a bebés prematuros. Por tanto no los consideramos outliers.

Normalidad

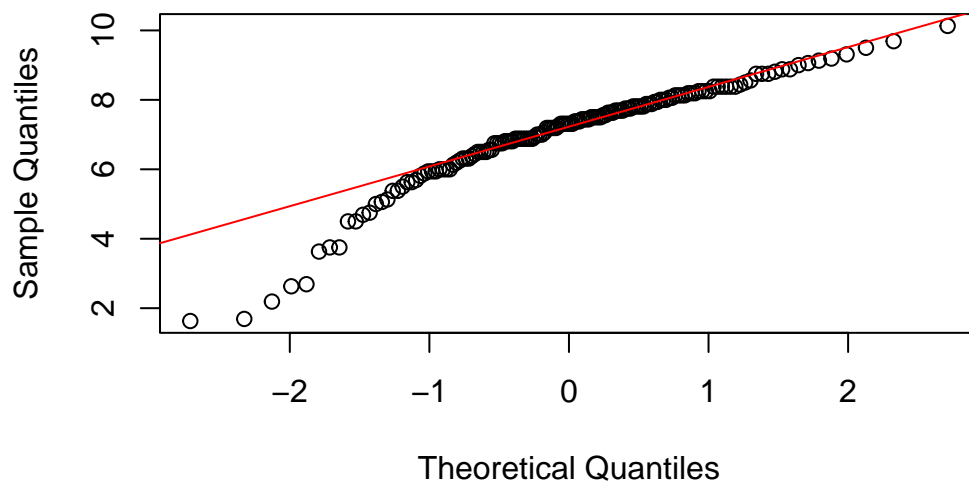
```
hist(data$weight)
```

Histogram of data\$weight



```
qqnorm(data$weight)
qqline(data$weight, col = "red")
```

Normal Q-Q Plot



```
shapiro.test(data$weight)
```

Shapiro-Wilk normality test

```
data: data$weight  
W = 0.91454, p-value = 9.58e-08
```

```
ks.test(data$weight, "pnorm", mean = mean(data$weight), sd = sd(data$weight))
```

Warning in ks.test.default(data\$weight, "pnorm", mean = mean(data\$weight), :
ties should not be present for the Kolmogorov-Smirnov test

Asymptotic one-sample Kolmogorov-Smirnov test

```
data: data$weight  
D = 0.12831, p-value = 0.01433  
alternative hypothesis: two-sided
```

La situación anterior, es decir que existen algunos bebés prematuros, pero en número son pocos, hace que observemos en la variable weight una asimetría a la izquierda que contrastamos en el qqplot y hace que ambos test rechacen la normalidad

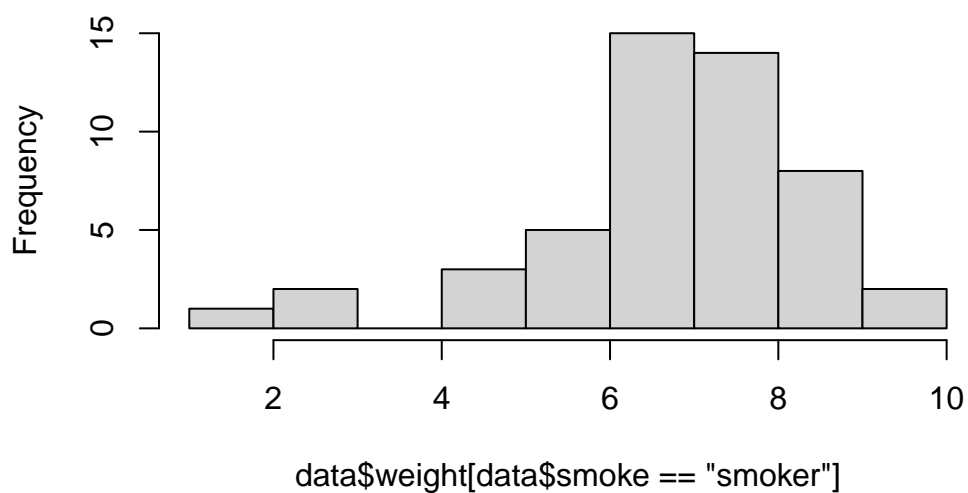
Homocedasticidad

Para estudiar la homocedasticidad, primero estudiaremos la normalidad en cada uno de los grupos de comparación:

Normalidad en Smokers

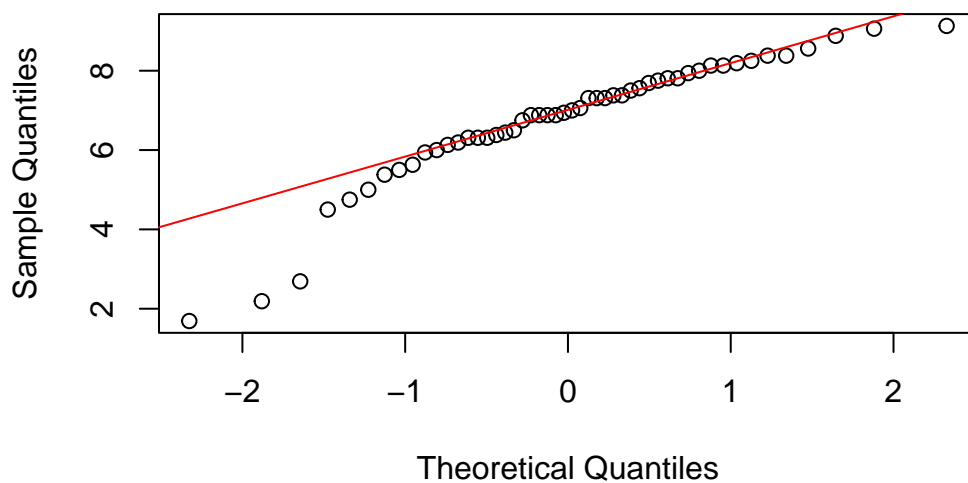
```
hist(data$weight[data$smoke=="smoker"])
```

Histogram of data\$weight[data\$smoke == "smoker"]



```
qqnorm(data$weight[data$smoke=="smoker"])  
qqline(data$weight[data$smoke=="smoker"], col = "red")
```

Normal Q-Q Plot



```
shapiro.test(data$weight[data$smoke=="smoker"])
```

Shapiro-Wilk normality test

```
data: data$weight[data$smoke == "smoker"]  
W = 0.89491, p-value = 0.0003276
```

```
ks.test(data$weight[data$smoke=="smoker"], "pnorm", mean = mean(data$weight[data$smoke=="smoker"])
```

Warning in ks.test.default(data\$weight[data\$smoke == "smoker"], "pnorm", : ties
should not be present for the Kolmogorov-Smirnov test

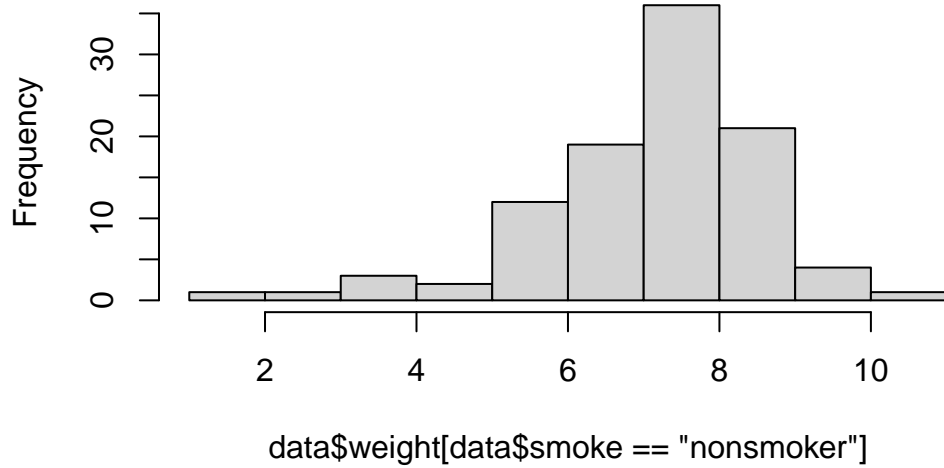
Asymptotic one-sample Kolmogorov-Smirnov test

```
data: data$weight[data$smoke == "smoker"]  
D = 0.12521, p-value = 0.4133  
alternative hypothesis: two-sided
```

Normalidad en Non-Smokers

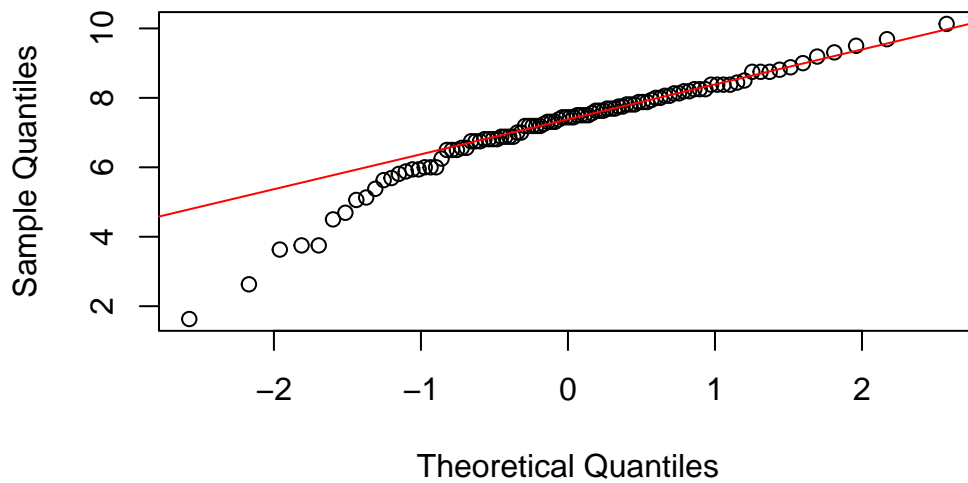
```
hist(data$weight[data$smoke=="nonsmoker"])
```

Histogram of data\$weight[data\$smoke == "nonsmoker"]



```
qqnorm(data$weight[data$smoke=="nonsmoker"])  
qqline(data$weight[data$smoke=="nonsmoker"], col = "red")
```

Normal Q-Q Plot




```
shapiro.test(data$weight[data$smoke=="nonsmoker"])
```

Shapiro-Wilk normality test

```
data: data$weight[data$smoke == "nonsmoker"]  
W = 0.92374, p-value = 2.234e-05
```

```
ks.test(data$weight[data$smoke=="nonsmoker"], "pnorm", mean = mean(data$weight[data$smoke=="nonsmoker"]))
```

```
Warning in ks.test.default(data$weight[data$smoke == "nonsmoker"], "pnorm", :  
ties should not be present for the Kolmogorov-Smirnov test
```

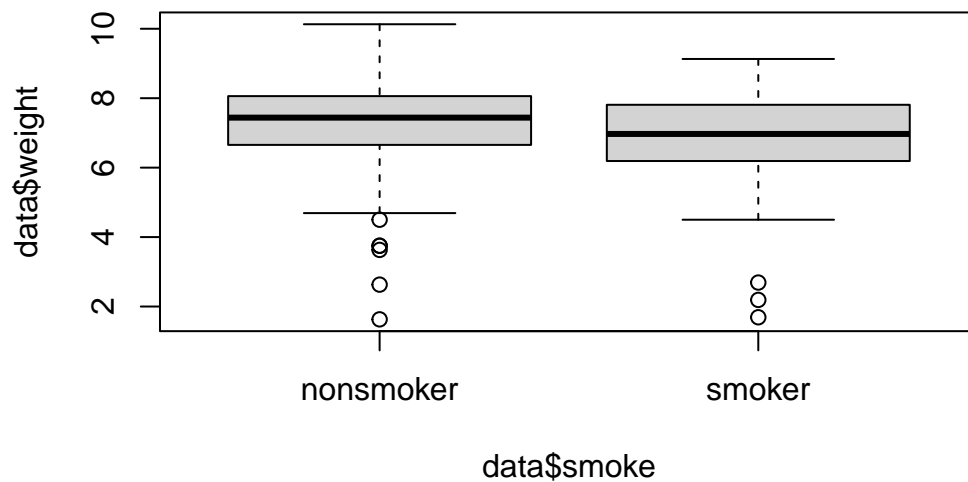
Asymptotic one-sample Kolmogorov-Smirnov test

```
data: data$weight[data$smoke == "nonsmoker"]  
D = 0.13229, p-value = 0.0604  
alternative hypothesis: two-sided
```

No podemos aceptar normalidad en ninguno de los dos grupos, por tanto no es adecuado ni el F-test ni el test de Bartlett, así que usaremos el test de Levenne:

Test de Levenne

```
boxplot(data$weight~data$smoke)
```



```
library(car)
```

Warning: package 'car' was built under R version 4.3.3

Loading required package: carData

Attaching package: 'car'

The following object is masked from 'package:openintro':

densityPlot

```
leveneTest(y = data$weight, group = data$smoke, center = "median")
```

Levene's Test for Homogeneity of Variance (center = "median")

	Df	F value	Pr(>F)
group	1	0.4442	0.5062
	148		

En el boxplot se aprecia que ambas distribuciones parecen tener dispersión similar, pero los outliers podrían inflar la varianza en los métodos sensibles.

El test de Levene no rechaza la igualdad de varianzas ($p > 0.05$), indicando que la variabilidad entre *smokers* y *nonsmokers* puede considerarse homogénea.

Dado que los datos no siguen una distribución normal ni globalmente ni en los grupos, y además contienen outliers, **el test apropiado para evaluar la homogeneidad de varianzas es el test de Levene con centro en la mediana**, que es robusto frente a estas violaciones de supuestos