



# Estadística aplicada en R

## *T-test*

-Febrero 2021-

Carlota Solano  
Álvaro Arredondo

## 2.1. T-test (test de Student)

### 1. ¿Qué es?

T-test es una prueba estadística que permite determinar si existe una diferencia estadísticamente significativa entre la media de dos grupos.

$$t = \frac{(\widehat{\beta}_a - \widehat{\beta}_b) - (\beta_a - \beta_b)}{s.e.(\widehat{\beta}_a) + s.e.(\widehat{\beta}_b)} \qquad t = \frac{\text{observado} - \text{esperado (si } \beta_a = \beta_b)}{\text{precisión de la media calculada respecto a la media real}}$$

### 2. ¿Cuándo se puede utilizar?

Cuando quieres comparar una variable continua entre dos niveles.

E.g.: Comparar la evapotranspiración de eucalipto en ladera de solana y umbría.

Comparar la cantidad de sal que echan griegos e italianos en la comida.

### 3. ¿Qué tipo de datos se necesitan?

Variable respuesta (dep.; y) → Numérica continua

Variable explicativa (indep.; x) → Categórica con dos niveles

## 2.1. T-test (test de Student)

### 4. ¿Qué asunciones tiene?

**Independencia** de las observaciones → Muestreo aleatorio

Distribución **normal** de la variable respuesta

Igualdad de **varianza** en los grupos a comparar (Homocedasticidad)

### 5. Matemáticamente, ¿cuál es la hipótesis?

H0: La media de dos grupos no difiere →  $\mu_1 = \mu_2$

Ha: La media de dos grupos difiere →  $\mu_1 \neq \mu_2$

## 2.1. T-test (test de Student)

6. ¿Cómo se corre en R?

```
R >t.test(respuesta ~ explicativa)
```

7. ¿Cómo se interpreta el resultado de R?

E.g.:

¿Cómo afecta la vitamina C en el crecimiento de dientes de cobayas?

Cobayas son tratadas con un suplemento (supp) de zumo de naranja (OJ) o con pastillas de vitamina C (VC), y se mide la longitud de sus dientes (len).

```
>t.test(ToothGrowth$len~ToothGrowth$supp)
```

t: diferencia entre grupos representado en unidades de error estándar (s.e.).

```
Welch Two Sample t-test
```

```
data: ToothGrowth$len by ToothGrowth$supp
```

```
t = 1.9153, df = 55.309, p-value = 0.06063
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
-0.1710156  7.5710156
```

```
sample estimates:
```

```
mean in group OJ mean in group VC
```

```
20.66333
```

```
16.96333
```

## 2.1. T-test (test

6. ¿Cómo se cor

R >t.test

7. ¿Cómo se inte

E.g.:

¿Cómo afecta la

Cobayas son tra

mide la longitud

>t.test(Toc

Welch

data: ToothG

t = 1.9153, d

alternative h

95 percent co

-0.1710156

sample estima

mean in group

20.66

	cum. prob	t <sub>.50</sub>	t <sub>.75</sub>	t <sub>.80</sub>	t <sub>.85</sub>	t <sub>.90</sub>	t <sub>.95</sub>	t <sub>.975</sub>	t <sub>.99</sub>	t <sub>.995</sub>	t <sub>.999</sub>	t <sub>.9995</sub>
one-tail		0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails		1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df												
1		0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2		0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3		0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4		0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5		0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6		0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7		0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8		0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9		0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10		0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11		0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12		0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13		0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14		0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15		0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16		0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17		0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18		0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19		0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20		0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21		0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22		0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23		0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24		0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25		0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26		0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27		0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28		0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29		0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30		0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40		0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60		0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80		0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100		0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000		0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
Z		0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
		0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
		Confidence Level										

$$= \frac{\hat{\beta} - \beta_0}{\text{s.e.}(\hat{\beta})}$$

tamina C (VC), y se

representado en  
ar (s.e.).

## 2.1. T-test (test de Student)

6. ¿Cómo se corre en R?

**R** >t.test(respuesta ~ explicativa)

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta_0}{\text{s. e.}(\hat{\beta})}$$

7. ¿Cómo se interpreta el resultado de R?

E.g.:

¿Cómo afecta la vitamina C en el crecimiento de dientes de cobayas?

Cobayas son tratadas con un suplemento (supp) de zumo de naranja (OJ) o con pastillas de vitamina C (VC), y se mide la longitud de sus dientes (len).

>t.test(ToothGrowht\$len~ToothGrowth\$supp)

**t**: diferencia entre grupos representado en unidades de error estándar (s.e.).

Welch Two Sample t-test

data: ToothGrowth\$len by ToothGrowth\$supp

t = 1.9153, df = 55.309, p-value = 0.06063

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.1710156 7.5710156

sample estimates:

mean in group OJ	mean in group VC
20.66333	16.96333

**df**: grados de libertad: cantidad de valores independientes que tenemos para calcular el modelo.



## 2.1. T-test (test de Student)

6. ¿Cómo se corre en R?

**R** >t.test(respuesta ~ explicativa)

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta_0}{\text{s. e.}(\hat{\beta})}$$

7. ¿Cómo se interpreta el resultado de R?

E.g.:

¿Cómo afecta la vitamina C en el crecimiento de dientes de cobayas?

Cobayas son tratadas con un suplemento (supp) de zumo de naranja (OJ) o con pastillas de vitamina C (VC), y se mide la longitud de sus dientes (len).

>t.test(ToothGrowth\$len~ToothGrowth\$supp)

**t**: diferencia entre grupos representado en unidades de error estándar (s.e.).

Welch Two Sample t-test

data: ToothGrowth\$len by ToothGrowth\$supp

t = 1.9153, df = 55.309, p-value = 0.06063

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.1710156 7.5710156

sample estimates:

mean in group OJ mean in group VC

20.66333

16.96333

**df**: grados de libertad: cantidad de valores independientes que tenemos para calcular el modelo.

**P-value**: Probabilidad de haber obtenido el resultado obtenido siendo H0 cierta

## 2.1. T-test (test de Student)

6. ¿Cómo se corre en R?

**R** >t.test(respuesta ~ explicativa)

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta_0}{\text{s.e.}(\hat{\beta})}$$

7. ¿Cómo se interpreta el resultado de R?

E.g.:

¿Cómo afecta la vitamina C en el crecimiento de dientes de cobayas?

Cobayas son tratadas con un suplemento (supp) de zumo de naranja (OJ) o con pastillas de vitamina C (VC), y se mide la longitud de sus dientes (len).

>t.test(ToothGrowth\$len~ToothGrowth\$supp)

**t:** diferencia entre grupos representado en unidades de error estándar (s.e.).

Welch Two Sample t-test

data: ToothGrowth\$len by ToothGrowth\$supp

t = 1.9153, df = 55.309, p-value = 0.06063

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.1710156 7.5710156

sample estimates:

mean in group OJ	mean in group VC
20.66333	16.96333

**df:** grados de libertad: cantidad de valores independientes que tenemos para calcular el modelo.

**P-value:** Probabilidad de haber obtenido el resultado obtenido siendo H0 cierta

**95%CI:** el rango de valores posibles de diferencia entre las medias de los grupos. → ¿Si contiene valor cero?



## 2.1. T-test (test de Student)

6. ¿Cómo se corre en R?

**R** >t.test(respuesta ~ explicativa)

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta_0}{\text{s. e.}(\hat{\beta})}$$

7. ¿Cómo se interpreta el resultado de R?

E.g.:

¿Cómo afecta la vitamina C en el crecimiento de dientes de cobayas?

Cobayas son tratadas con un suplemento (supp) de zumo de naranja (OJ) o con pastillas de vitamina C (VC), y se mide la longitud de sus dientes (len).

>t.test(ToothGrowth\$len~ToothGrowth\$supp)

**t:** diferencia entre grupos representado en unidades de error estándar (s.e.).

Welch Two Sample t-test

data: ToothGrowth\$len by ToothGrowth\$supp

t = 1.9153, df = 55.309, p-value = 0.06063

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.1710156 7.5710156

sample estimates:

mean in group OJ	mean in group VC
20.66333	16.96333

**df:** grados de libertad: cantidad de valores independientes que tenemos para calcular el modelo.

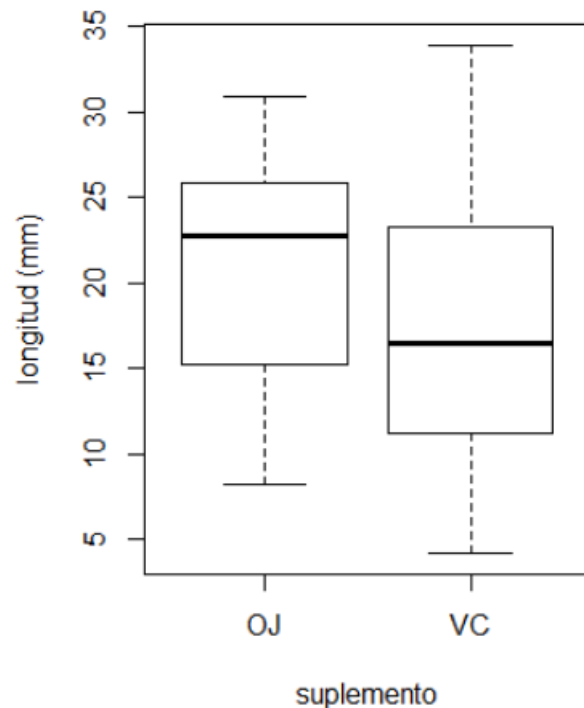
**P-value:** Probabilidad de haber obtenido el resultado obtenido siendo H0 cierta

**95%CI:** el rango de valores posibles de diferencia entre las medias de los grupos. → ¿Si contiene valor cero?

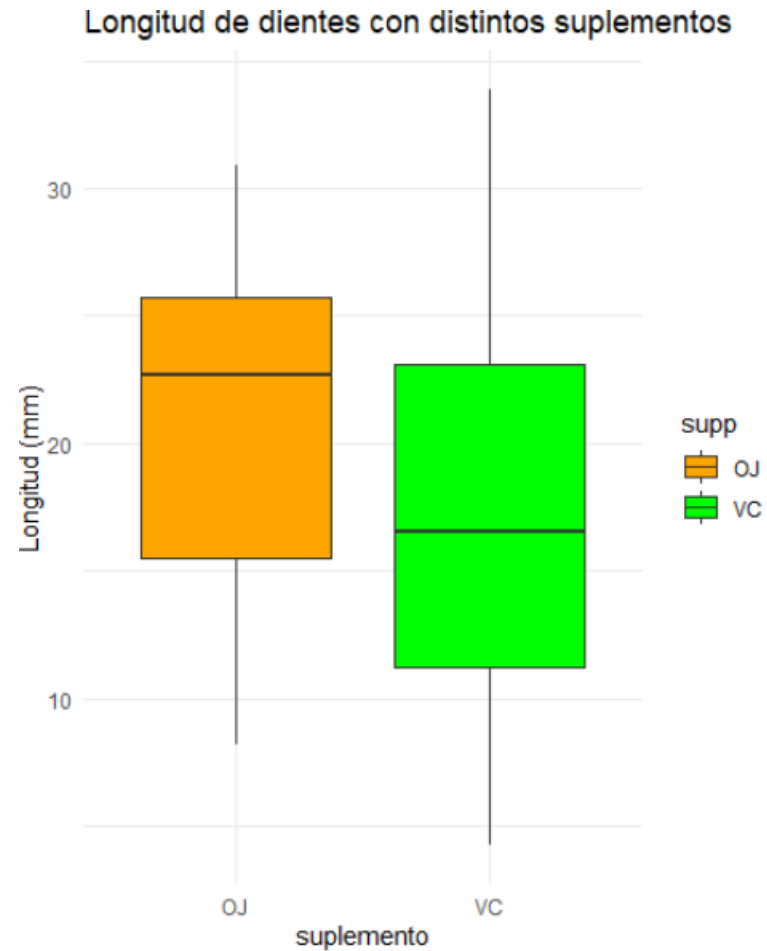
La **media** de la variable respuesta del grupo OJ y del grupo VC

## 2.1. T-test (test de Student)

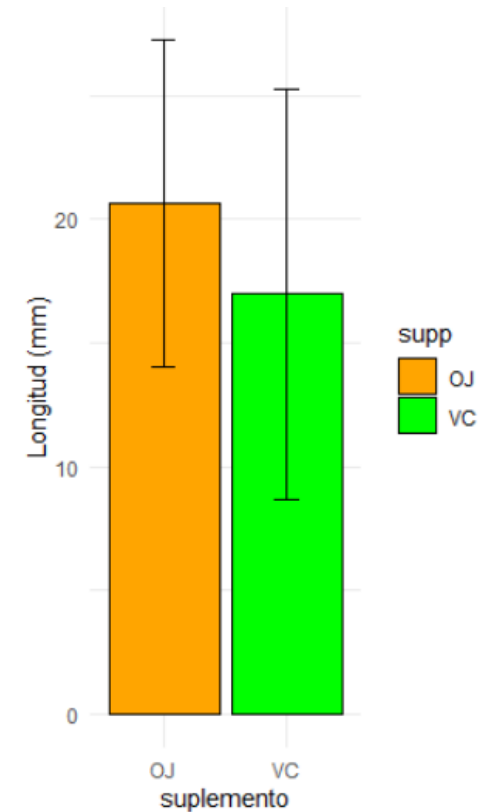
### 8. ¿Cómo se puede representar?



```
>plot(y~x)
```



```
>ggplot(data, aes(x,y))+  
geom_boxplot()
```



```
>data_summary()  
>ggplot(data,aes(x,y))+  
geom_bar(stat="identity")+  
geom_errorbar(aes(ymin,ymax))
```

## 2.2. T-test ejemplo

Base de datos con medidas de arrendajo azul (*Cyanocitta cristata*).

¿Difiere la masa de hembras y machos de arrendajo azul?

H0: La masa de hembras y machos no difiere

Ha: La masa de hembras y machos difiere



R

### 1º Entender la base de datos

```
> str(db)
'data.frame':   123 obs. of  9 variables:
 $ BirdID      : Factor w/ 123 levels "0000-00000","1142-05901",...: 1 2 3 4 5 6 7 8 9 10 ...
 $ KnownSex    : Factor w/ 2 levels "F","M": 2 2 2 1 2 1 2 2 1 1 ...
 $ BillDepth   : num  8.26 8.54 8.39 7.78 8.71 7.28 8.74 8.72 8.2 7.67 ...
 $ Billwidth   : num  9.21 8.76 8.78 9.3 9.84 9.3 9.28 9.94 9.01 9.31 ...
 $ BillLength  : num  25.9 25 26.1 23.5 25.5 ...
 $ Head        : num  56.6 56.4 57.3 53.8 57.3 ...
 $ Mass        : num  73.3 75.1 70.2 65.5 74.9 ...
 $ Skull       : num  30.7 31.4 31.2 30.3 31.9 ...
 $ Sex         : int   1 1 1 0 1 0 1 1 0 0 ...
```

## 2.2. T-test ejemplo

Base de datos con medidas de arrendajo azul (*Cyanocitta cristata*).

¿Difiere la masa de hembras y machos de arrendajo azul?

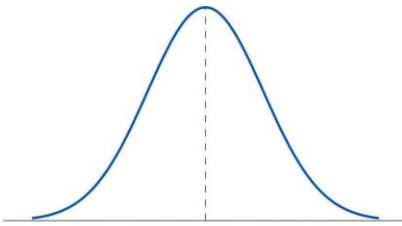
H0: La masa de hembras y machos no difiere

Ha: La masa de hembras y machos difiere

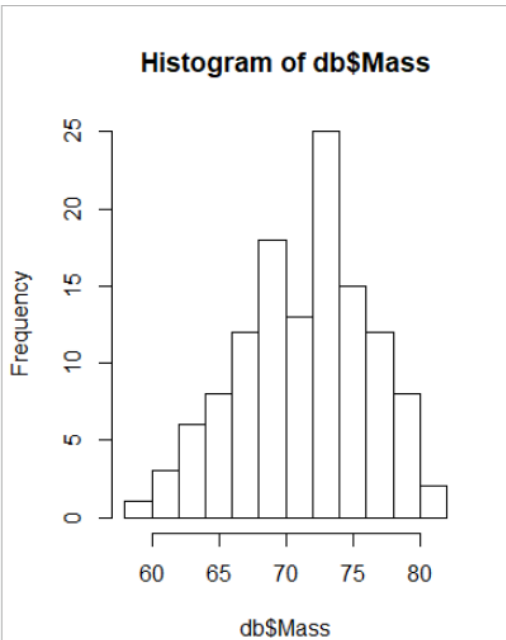
R

### 2º Asunciones

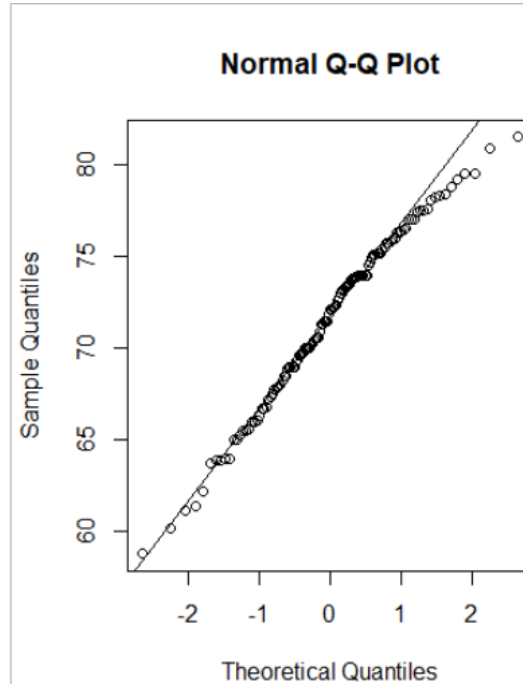
- Normalidad:



```
>hist(db$Mass)
```



```
>qqnorm(db$Mass)  
>qqline(db$Mass)
```



*Cyanocitta cristata*

```
>shapiro.test(db$Mass)
```

Shapiro-wilk normality test

data: db\$Mass

W = 0.98599, p-value = 0.2366



H0: datos con distribución normal

## 2.2. T-test ejemplo

Base de datos con medidas de arrendajo azul (*Cyanocitta cristata*).

¿Difiere la masa de hembras y machos de arrendajo azul?

H0: La masa de hembras y machos no difiere

Ha: La masa de hembras y machos difiere

R

### 2º Asunciones

- Homocedasticidad:

```
> library(car)
```

```
> leveneTest(db$Mass~db$KnownSex)
```

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

	Df	F value	Pr(>F)
group 1	4.4591	0.03677	*
121			

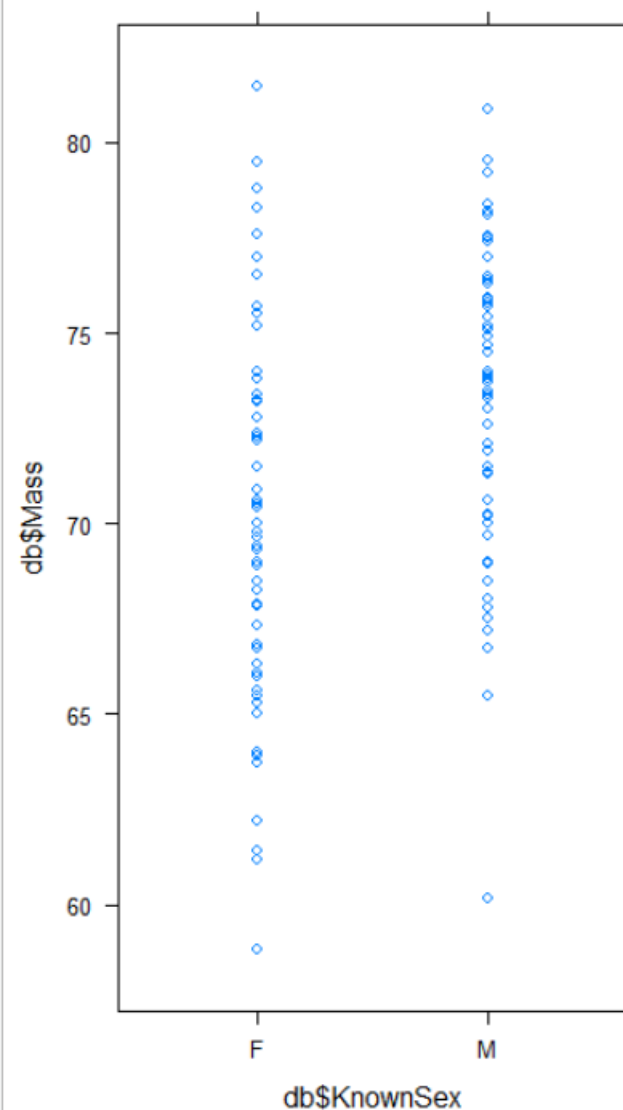
---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



H0: varianzas de ambos grupos son semejantes



## 2.2. T-test ejemplo

Base de datos con medidas de arrendajo azul (*Cyanocitta cristata*).

¿Difiere la masa de hembras y machos de arrendajo azul?

H0: La masa de hembras y machos no difiere

Ha: La masa de hembras y machos difiere

R

3º Test estadístico

```
> t.test(db$Mass~db$KnownSex)
```

```
Welch Two Sample t-test
```

```
data: db$Mass by db$KnownSex
```

```
t = -4.2179, df = 111.05, p-value = 5.051e-05
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
-5.024846 -1.812646
```

```
sample estimates:
```

```
mean in group F mean in group M
```

```
69.80633
```

```
73.22508
```



*Cyanocitta cristata*



## 2.2. T-test ejemplo

Base de datos con medidas de arrendajo azul (*Cyanocitta cristata*).

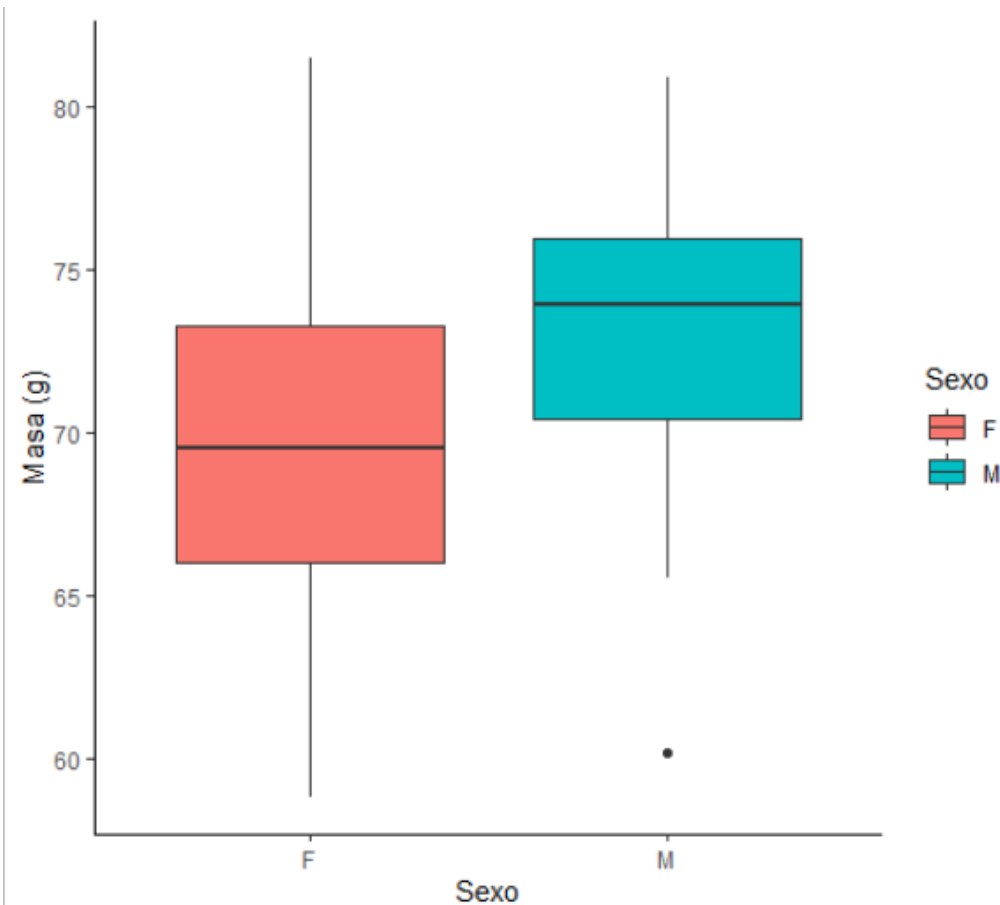
¿Difiere la masa de hembras y machos de arrendajo azul?

H0: La masa de hembras y machos no difiere

Ha: La masa de hembras y machos difiere

R

4º Graficar



## 2.2. T-test ejemplo

Base de datos con medidas de arrendajo azul (*Cyanocitta cristata*).

¿Difiere la masa de hembras y machos de arrendajo azul?

H0: La masa de hembras y machos no difiere

Ha: La masa de hembras y machos difiere

R

### 5º Interpretar matemática y biológicamente

```
> t.test(db$Mass~db$KnownSex)
```

Welch Two Sample t-test

```
data: db$Mass by db$KnownSex
```

```
t = -4.2179, df = 111.05, p-value = 5.051e-05
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
-5.024846 -1.812646
```

```
sample estimates:
```

```
mean in group F mean in group M
```

```
69.80633
```

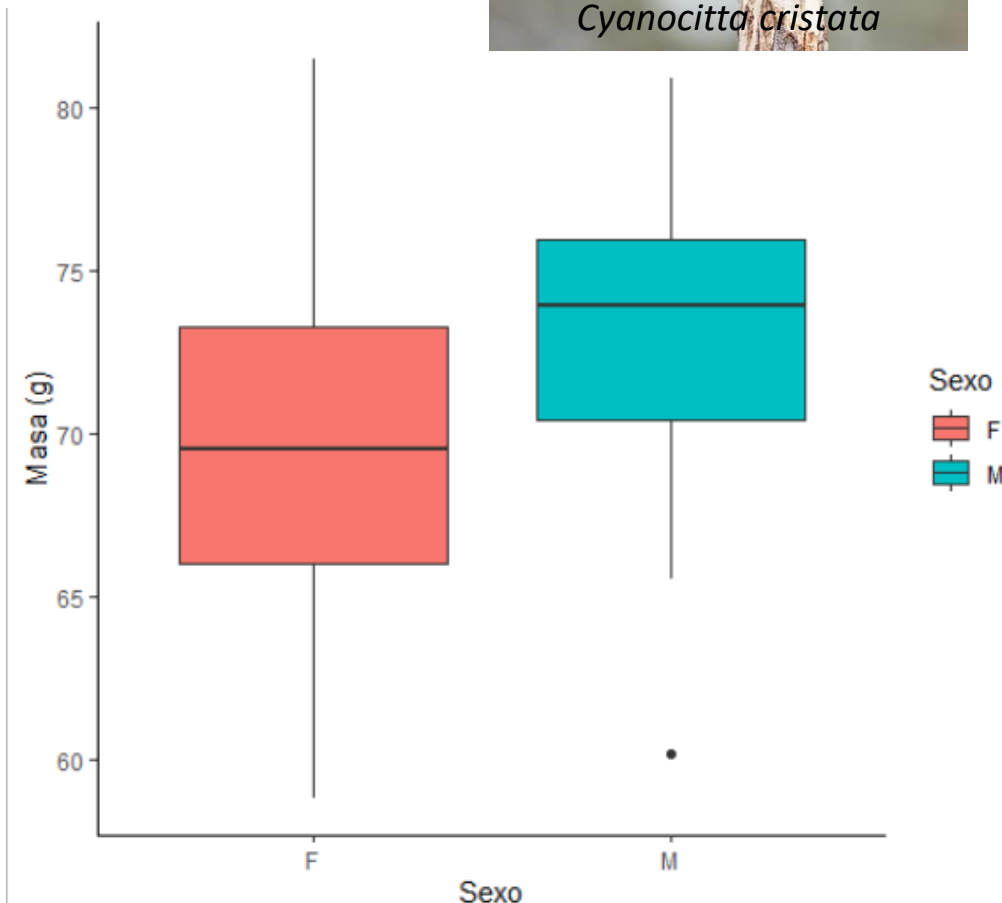
```
73.22508
```



**Effect size (Tamaño del efecto)**



*Cyanocitta cristata*



## 2.2. T-test ejemplo

Base de datos con medidas de arrendajo azul (*Cyanocitta cristata*).

¿Difiere la masa de hembras y machos de arrendajo azul?

**H0:** La masa de hembras y machos no difiere

**Ha:** La masa de hembras y machos difiere



### 6º Conclusión y comprobación de hipótesis

Las hembras (F) presentan una masa de 69.81 g, frente a los machos (M) que muestran una masa de 73.23g.

Estadísticamente, esta diferencia es significativa ( $t=-4.2179$ ,  $df=111.05$ ,  $p\text{-value} < 0.01$ ), por lo que podemos **rechazar la H0 y aceptar la Ha**, i.e. la masa de hembras y machos de arrendajo azul difiere significativamente, con las hembras presentando aproximadamente 3 gramos de masa menos que los machos.

## 2.3. T-test ejercicio

Ejercicios: 2.Ejer\_Ttest