

# Prueba de hipótesis y análisis de varianzas

GUALAPURO, Moisés

2023-04-25

## Prueba de hipótesis

Para una pregunta de investigación se plantea una proposición que se debe verificar. La prueba de la hipótesis es la verificación, con un nivel de confianza, de si se cumple o no la proposición inicial. ### Data Orange

La data `orange` tiene 35 registros del crecimiento de árboles de naranja. Las variables son: el número de árbol, la edad en días a partir del 31 de diciembre de 1968 y la circunferencia medido en milímetros.

```
data(Orange)
str(Orange)
```

```
## Classes 'nfnGroupedData', 'nfGroupedData', 'groupedData' and 'data.frame': 35 obs. of 3 variables
## $ Tree      : Ord.factor w/ 5 levels "3"<"1"<"5"<"2"<...: 2 2 2 2 2 2 2 4 4 4 ...
## $ age       : num 118 484 664 1004 1231 ...
## $ circumference: num 30 58 87 115 120 142 145 33 69 111 ...
## - attr(*, "formula")=Class 'formula' language circumference ~ age | Tree
## .. ..- attr(*, ".Environment")=<environment: R_EmptyEnv>
## - attr(*, "labels")=List of 2
## ..$ x: chr "Time since December 31, 1968"
## ..$ y: chr "Trunk circumference"
## - attr(*, "units")=List of 2
## ..$ x: chr "(days)"
## ..$ y: chr "(mm)"
```

El número de árboles es de tipo factor y se tiene 5 árboles. La edad y el diámetro son de tipo numérico.

```
summary(Orange)
```

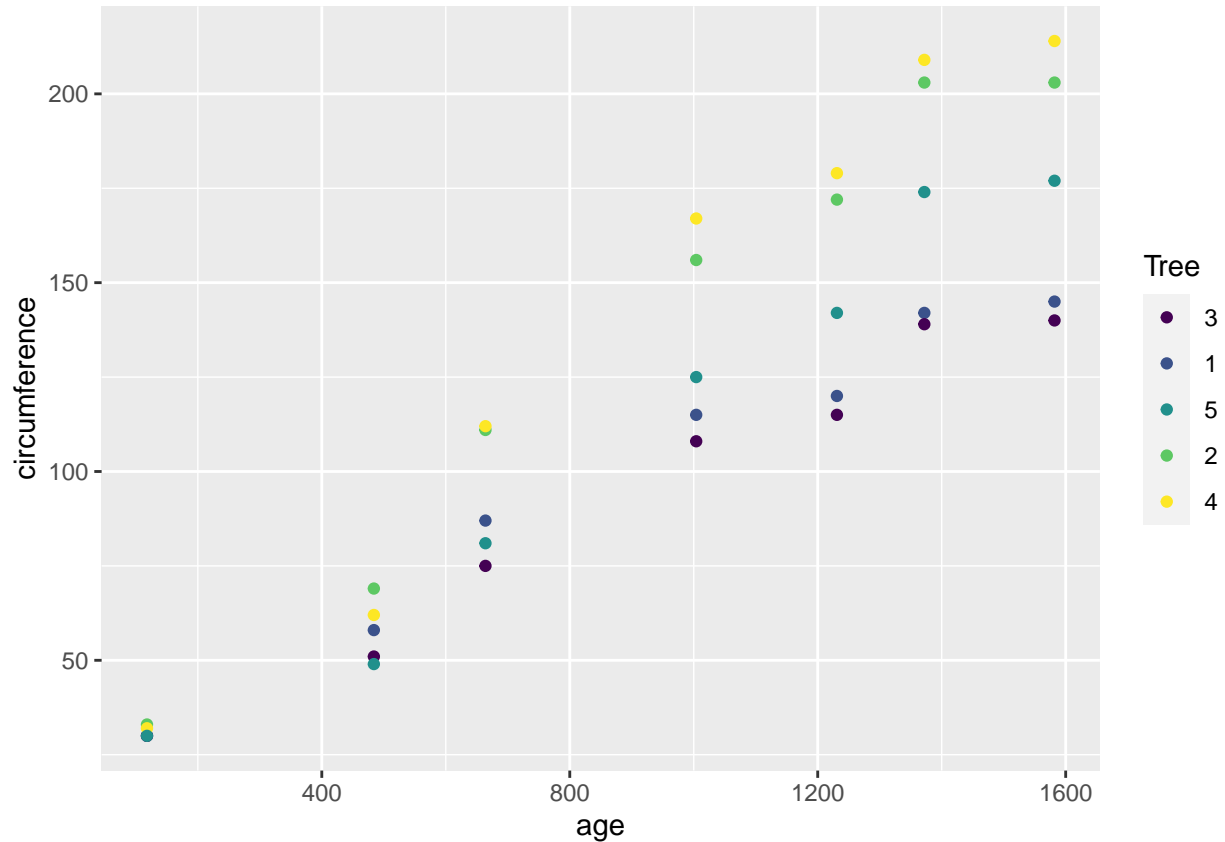
```
## Tree      age      circumference
## 3:7   Min.    : 118.0   Min.      : 30.0
## 1:7   1st Qu.: 484.0   1st Qu.: 65.5
## 5:7   Median :1004.0   Median :115.0
## 2:7   Mean    : 922.1   Mean      :115.9
## 4:7   3rd Qu.:1372.0   3rd Qu.:161.5
##      Max.    :1582.0   Max.      :214.0
```

## Visualización preliminar

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
ggplot(data = Orange, aes(x = age, y = circumference,  
                           col = Tree)) +  
  geom_point()
```



La figura muestra que a medida que pasa el tiempo (edad) el diámetro de los árboles incrementa. El diámetro de los árboles 2 y 4 es mayor que la de los árboles 1 y 3.

## Planteamiento de la Hipótesis

La hipótesis nula ( $H_0$ ) es aquella donde se espera que no exista efecto. Es la más fácil de contradecir.

$H_{01}$ : La edad del árbol no incide en el tamaño del diámetro.  $H_{02}$ : Los árboles tienen el mismo diámetro a medida que pasa el tiempo.

```
h01 <- lm(circumference ~ age, data = Orange)  
summary(h01)
```

```
##  
## Call:
```

```
## lm(formula = circumference ~ age, data = Orange)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -46.310 -14.946  -0.076   19.697   45.111
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.399650   8.622660   2.018   0.0518 .
## age         0.106770   0.008277  12.900 1.93e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.74 on 33 degrees of freedom
## Multiple R-squared:  0.8345, Adjusted R-squared:  0.8295
## F-statistic: 166.4 on 1 and 33 DF,  p-value: 1.931e-14
```

Con un nivel de confianza del 90% se estima que el intercepto (el valor del diámetro sin efecto de la edad) es de 17.39 mm.

Con un nivel de confianza del 99.9% se estima que la pendiente (el cambio en mm del diámetro por cada día que pasa) es de  $0.106 \frac{mm}{dia}$ .

Con un nivel de confianza del 90% la relación entre el diámetro de los árboles de naranja y la edad es:

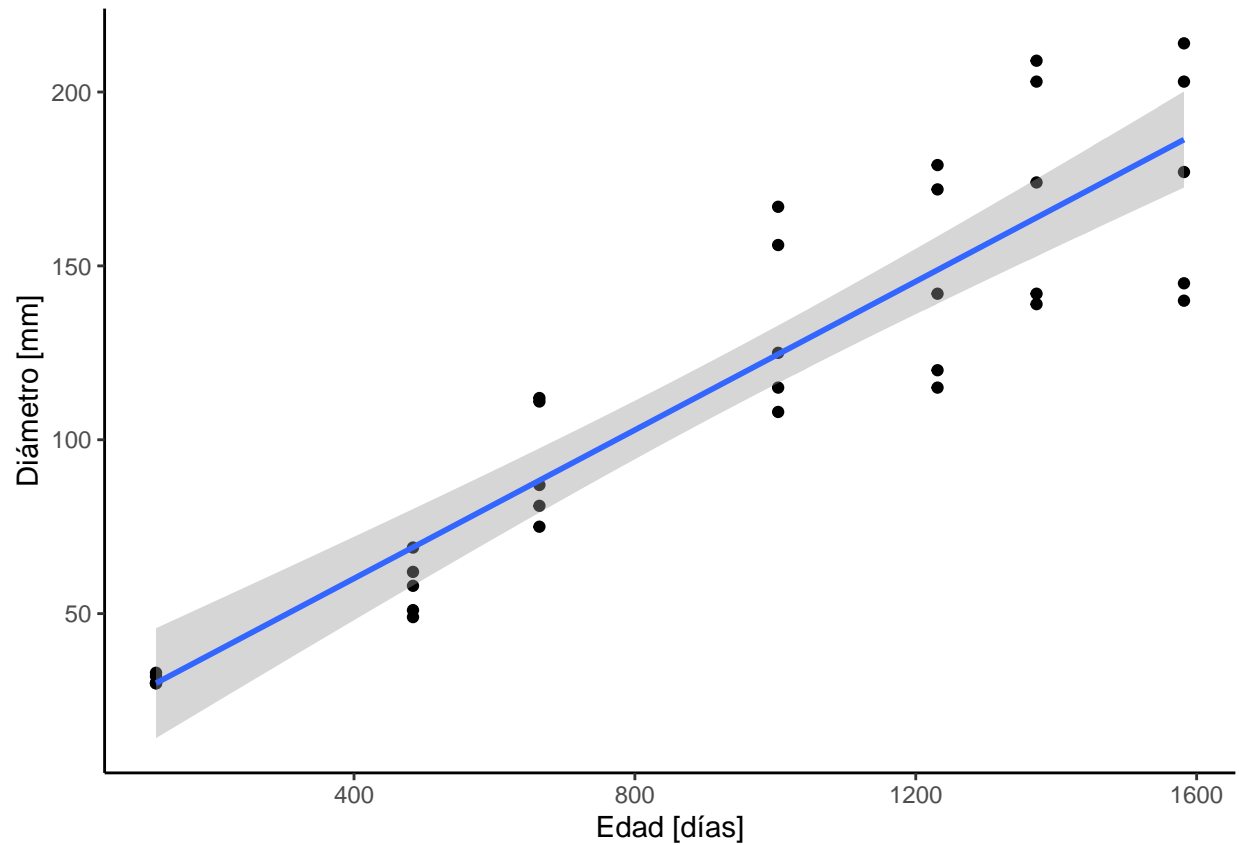
$$D = 17.39 + 0.106 \cdot E$$

Donde D es el diámetro en milímetros y E es la edad en días.

## Visualización de la regresión

```
ggplot(data = Orange, aes(x = age, y = circumference)) +
  geom_point() + geom_smooth(method = "lm") + theme_classic() +
  labs(x = "Edad [días]", y = "Diámetro [mm]")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



## Verificación la hipótesis 2

```
ggplot(data = Orange, aes(x = age, y = circumference, col = Tree)) +
  geom_point() + geom_smooth(method = "lm") + theme_classic() +
  labs(x = "Edad [días]", y = "Diámetro [mm]")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

