

# ANOVA. Análisis de regresión y modelo lineal

## Métodos de contraste de hipótesis y diseño de experimentos

00RTeam

Marzo 2020

- 1 Comparaciones múltiples
- 2 ANOVA
- 3 Modelo de regresión lineal
- 4 Referencias y bibliografía

## Comparaciones múltiples

## Comparaciones múltiples. Problemas

## Aumento del error de tipo I

Si contrastamos diferentes hipótesis nulas (independientes) a la vez, con  $\alpha = 0.05$  hay más de un 5% de probabilidades de obtener un resultado significativo por azar.

Si realizamos tres contrastes, la probabilidad de no cometer error de tipo I es **0.95 para cada test**. Si son independientes:

- La probabilidad de **no** cometer error de tipo I es:  $(0.95)^3 = 0.857375$ .
- La probabilidad de cometer error de tipo I es:  $1 - (0.95)^3 = 0.142625$

Cuanto más comparaciones hagamos más crece el error de tipo I. Para fijar el error general en 0.05 hemos de tomar en cada test  $\alpha = 1 - \sqrt[3]{0.95} = 0.01695$

# ANOVA

## Ejemplo tratamiento. Enunciado

Efectividad de un tratamiento en un conjunto de enfermos de una enfermedad rara. Tres pruebas de falta de capacidad de raciocinio: mes cero (antes), mes uno y mes tres. ¿Cómo contrastar el raciocinio antes del tratamiento ( $m_0$ ) según la raza?

```
trat <- read.table( "files/tratamiento.csv", sep = ";",  
                    head = TRUE )  
head( trat, 5 )
```

##	id	genero	raza	m0	m1	m3
## 1	1	1	3	35	25	16
## 2	2	2	1	37	23	12
## 3	3	2	1	36	22	14
## 4	4	1	2	34	21	13
## 5	5	2	3	60	43	22

## Introducción al ANOVA



# Introducción al ANOVA

El ANOVA (análisis de la varianza) sirve para comparar dos o más medias. Es una generalización de la prueba  $t$  de Student. Conocida también como “Anova de Fisher” o “análisis de varianza de Fisher” por utilizar la distribución  $F$  de Fisher en el contraste. Según los factores que tengamos:

- ANOVA de una vía (factor entre sujetos)
- ANOVA de dos vías (varios factores entre sujetos)
- ANOVA para medidas repetidas (factor intra sujetos)
- ANOVA mixto (factores entre e intra sujetos)

## ANOVA según variables independientes o factores (vías)

En ANOVA, a la variable categórica que define los grupos que deseamos comparar la llamamos **variable independiente** o **factor** y a la variable cuantitativa en la que deseamos comparar los grupos la llamamos **variable dependiente**.

Los factores pueden variar entre sujetos (between subjects) o dentro de los sujetos (within subjects).

- Los **factores entre sujetos** (between) son los que no se miden dos (o más) veces para un mismo sujeto. Ejemplo: edad, raza, género, etc.
- Los **factores dentro de los sujetos** (within) son los que se miden varias veces para el mismo sujeto. Ejemplo: muestras tomadas en varios momentos, peso antes y después, etc.

## Ejemplo tratamiento. Lectura de datos

```
str( trat )
```

```
## 'data.frame':    18 obs. of  6 variables:
## $ id      : int  1 2 3 4 5 6 7 8 9 10 ...
## $ genero: int  1 2 2 1 2 1 1 2 2 2 ...
## $ raza   : int  3 1 1 2 3 3 2 2 3 2 ...
## $ m0     : int  35 37 36 34 60 54 50 60 48 47 ...
## $ m1     : int  25 23 22 21 43 46 46 47 35 32 ...
## $ m3     : int  16 12 14 13 22 26 23 25 20 19 ...
```

## Ejemplo tratamiento. Factores

*# Ejercicio: transforma en factores*

## ANOVA de una vía

# ANOVA de una vía. Introducción

El ANOVA de una vía (*one-way ANOVA*) o *ANOVA de un factor* examina la igualdad de las medias de la población para un resultado cuantitativo y **una única variable categórica con dos o más niveles**.

La hipótesis nula  $H_0$  es que no hay diferencia entre las medias y la alternativa,  $H_1$ , es que al menos una de las medias difiere del resto.

# ANOVA de una vía. Supuestos

- Supuestos:
  - **Independencia** de las observaciones
  - **Normalidad** (robusto). Alternativa: Kruskal-Wallis
  - **Homocedasticidad** (robusto si muestras balanceadas). Alternativa: prueba de Welch

## Ejemplo tratamiento. Normalidad

*# Ejercicio: comprobar normalidad (Shapiro-Wilk)*



## Ejemplo tratamiento. Homocedasticidad

*# Ejercicio: comprobar homocedasticidad (Bartlett)*

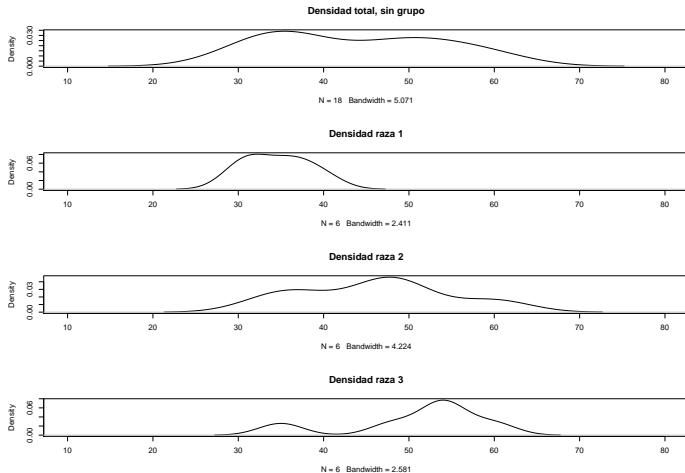
## Ejemplo tratamiento. ANOVA de una vía

La hipótesis nula  $H_0$  es que no hay diferencia entre las medias y la alternativa,  $H_1$ , que al menos una de las medias difiere del resto.

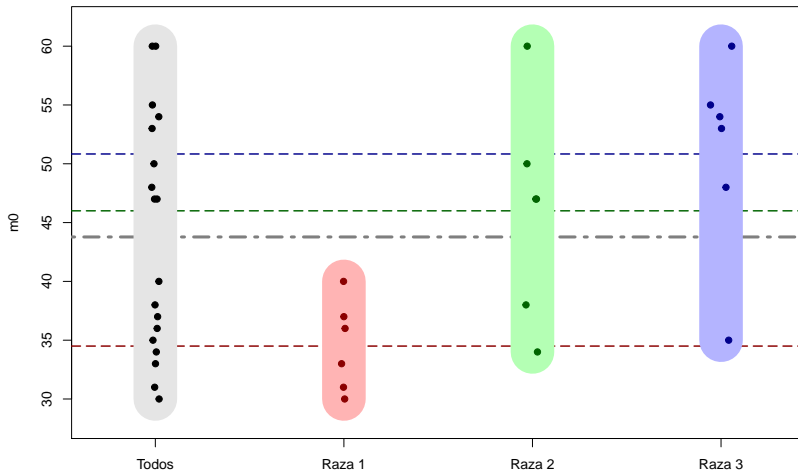
```
fitTrat <- aov( m0 ~ raza, data = trat )
summary( fitTrat )
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## raza           2   844.8    422.4     7.28 0.00617 **
## Residuals     15   870.3     58.0
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Ejemplo tratamiento. Suma de cuadrados



## Ejemplo tratamiento. Suma de cuadrados



## Ejemplo tratamiento. ANOVA

- Estadístico  $F$

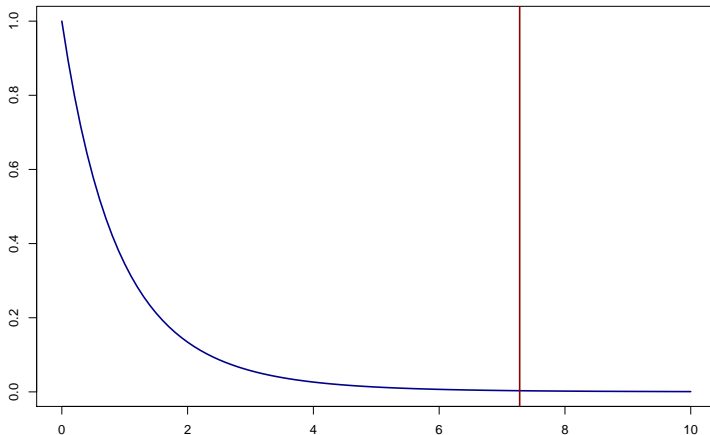
$$F = \frac{\text{varianza explicada}}{\text{varianza inexplicada}}$$

- P-valor

Si es menor que 0.05 se rechaza a la hipótesis nula, es decir, el factor tiene un efecto significativo en el experimento.

# Distribución $F$ de Snedecor

Distribución F de Snedecor con grados de libertad 2 y 15



## Ejemplo tratamiento. ANOVA y regresión

```
str( fitTrat )  
fitTrat$ # tabulador para ver opciones  
fitTrat$coefficients  
lm( m0 ~ raza, data = trat )
```

## Ejemplo tratamiento. No HOV, test de Welch

```
oneway.test( m0 ~ raza, data = trat )
```

```
##  
## One-way analysis of means (not assuming  
## equal variances)  
##  
## data:  m0 and raza  
## F = 10.516, num df = 2.0000, denom df =  
## 8.4967, p-value = 0.005032
```



## Después del ANOVA. Contrastes post-hoc

¿Qué grupos son los que difieren? Contrastes dos a dos corrigiendo el nivel de significación.

- Contrastes *post-hoc* o no planificados: sin idea previa
  - Menos de 6 niveles
    - Bonferroni
    - Holm
  - 6 o más niveles
    - LSD Fisher
    - HSD Tukey
- Comparaciones planificadas: con idea previa

## Post-hoc. Pruebas $t$ con ajuste de Bonferroni

- Bonferroni, Holm, LSD Fisher

```
# ?p.adjust  
pairwise.t.test( trat$m0, trat$raza, p.adj = "holm")
```

- HSD Tukey

```
fitTrat <- aov( m0 ~ raza, data = trat )  
TukeyHSD( fitTrat )
```

## Ejercicio

El conjunto de datos de R `PlantGrowth` muestra el peso que tiene un tipo de planta al no ser expuestas a ningún tipo de tratamiento (control) y al ser expuestas a dos tipos de tratamiento diferentes (`trt1` y `trt2`).

¿Hay diferencias entre los tratamientos? ¿Entre cuáles?

## Alternativa no paramétrica al Anova de 1 vía

- Prueba de **Kruskal-Wallis**
- Compara más de dos grupos independientes
- Requiere homogeneidad de varianzas

```
kruskal.test( variable ~ factor, data = datos )  
pairwise.wilcox.test( datos$variable,  
                      datos$factor,  
                      p.adjust = "bonferroni" )
```

## ANOVA de dos vías

## ANOVA de dos vías. Introducción

El ANOVA de dos vías (*two-way ANOVA*) o *ANOVA de dos factores* examina la igualdad de las medias de la población para un resultado cuantitativo y **dos variables categóricas o factores**. El modelo ANOVA de dos vías evalúa, además de los efectos de los factores sobre la variable independiente, los efectos de la interacción entre ellas.

## ANOVA de dos vías. Supuestos

Mismos supuestos que el ANOVA de una vía para ambos factores:

- Supuestos:
  - **Independencia** de las observaciones
  - **Normalidad** (robusto)
  - **Homocedasticidad** (robusto si muestras balanceadas)

## Ejemplo tratamiento 2. Datos

En la base de datos `tratamiento.csv` ¿Cómo contrastar el raciocinio antes del tratamiento ( $m_0$ ) según la raza y el género?

```
head( trat )
```

```
##    id genero raza m0 m1 m3
## 1  1      1    3 35 25 16
## 2  2      2    1 37 23 12
## 3  3      2    1 36 22 14
## 4  4      1    2 34 21 13
## 5  5      2    3 60 43 22
## 6  6      1    3 54 46 26
```



## Ejemplo tratamiento 2. Supuestos

Es necesario comprobar la homocedasticidad para ambos factores: *género* y *raza*, y comprobar la normalidad para todos los niveles de cada factor.

En este ejemplo, faltarían los supuestos para la variable *género* (los de *raza* se han realizado anteriormente).

```
bartlett.test( trat$m0 ~ trat$genero )  
shapiro.test( trat$m0[ trat$genero==1 ] )  
shapiro.test( trat$m0[ trat$genero==2 ] )
```

## Ejemplo tratamiento 2. ANOVA de dos vías

```
fitTrat2 <- aov( m0 ~ raza * genero, data = trat )
summary( fitTrat2 )
```

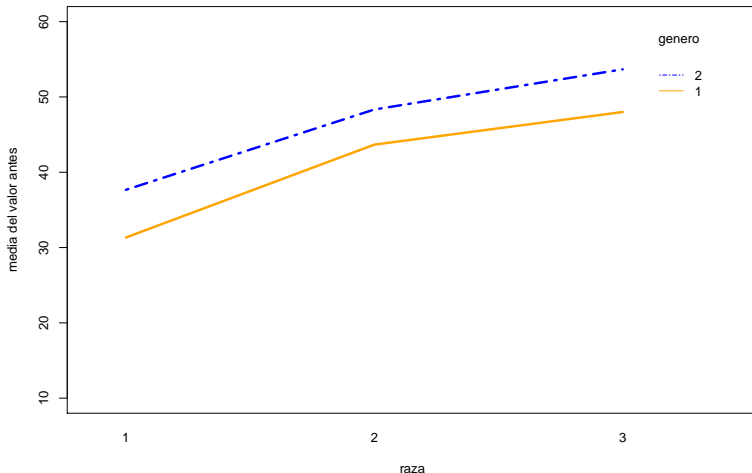
```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## raza          2   844.8    422.4     6.950 0.00989 **
## genero         1   138.9    138.9     2.285 0.15649
## raza:genero    2     2.1     1.1     0.017 0.98281
## Residuals     12   729.3     60.8
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La variable *raza* tiene un efecto significativo, el *género* no y no hay interacción entre ambas, esto es, se acepta la hipótesis nula.

## Ejemplo tratamiento 2. Interaction plot

```
interaction.plot( trat$raza, trat$genero, trat$m0,  
                 ylim = c( 10, 60 ),  
                 col = c( "orange", "blue" ),  
                 lty = c( 1, 12 ),  
                 lwd = 3,  
                 ylab = "media del valor antes",  
                 xlab = "raza", trace.label = "genero" )
```

## Ejemplo tratamiento 2. Interaction plot



## ANOVA de dos vías. ¿Importa el orden?

```
aov( m0 ~ raza * genero, data = trat )  
aov( m0 ~ genero * raza, data = trat )
```

Si los datos son **balanceados** no importa el orden. Si no son balanceados, **sí**. Esto se debe al tipo de suma de cuadrados.

¿Son balanceados nuestros datos?

```
table( trat$genero )
```

```
##
```

```
## 1 2
```

```
## 9 9
```

```
table( trat$raza )
```

## ANOVA de dos vías. ¿Importa el orden?

¿Y si mis datos no son balanceados? Lo recomendable es utilizar la función `Anova()` del paquete `car` con suma de cuadrados tipo III.

```
options( contrasts = c("contr.sum", "contr.poly") )  
fitDatos <- aov( m0 ~ genero * raza, data = trat )  
library( car )  
Anova( fitDatos, type = "III" )
```

## ANOVA para medidas repetidas

# ANOVA para medidas repetidas. Introducción

Se utiliza cuando se tienen **varias medidas para el mismo sujeto** en condiciones diferentes. No hay independencia. Una ventaja es que se tienen más datos con los mismos sujetos. Se controla mejor la variabilidad interna.

- Supuestos:
  - **Esfericidad:** varianzas de las diferencias iguales. Test de Mauchly ( $H_0$ : hay esfericidad)

Si no hay esfericidad: corrección de Greenhouse–Geisser y corrección de Huynh–Feldt



# ANOVA para medidas repetidas. Reestructurar los datos

Antes:

sujeto	n1	n2	n3	n4
S1	7	5	6	2
S2	1	0	3	0
S3	8	8	6	1
S4	4	3	1	2

# ANOVA con medidas repetidas. Reestructurar los datos

Ahora:

sujeto	nivel	medida
S1	n1	7
S1	n2	1
S1	n3	8
S1	n4	4
S2	n1	5
S2	n2	0
...	...	...
S4	n4	2

## Ejemplo tratamiento 3. Reestructurar los datos

Se utiliza la función `melt` del paquete `reshape2`.

```
library( reshape2 )  
tratRe <- melt( trat, id = c( "id", "genero", "raza" ),  
               measure = c( "m0", "m1", "m3" ),  
               variable.name = "mes",  
               value.name = "faltaRac" )  
  
head( tratRe )
```

```
##   id genero raza mes faltaRac  
## 1  1      1    3  m0        35  
## 2  2      2    1  m0        37  
## 3  3      2    1  m0        36  
## 4  4      1    2  m0        34  
## 5  5      2    3  m0        60  
## 6  6      1    3  m0        54
```

## Ejemplo tratamiento 3. Función aov

```
summary( aov( faltaRac ~ mes + Error( id / mes ),  
           data = tratRe ) )
```

```
##  
## Error: id  
##           Df Sum Sq Mean Sq F value Pr(>F)  
## Residuals 17   3247      191  
##  
## Error: id:mes  
##           Df Sum Sq Mean Sq F value Pr(>F)  
## mes         2   5684      2842   142.1 <2e-16 ***  
## Residuals 34    680       20  
## ---  
## Signif. codes:
```

## Ejemplo tratamiento 3. Función ezANOVA del paquete ez

La función `ezANOVA()` del paquete `ez` además del modelo ANOVA lleva a cabo el análisis de supuestos previo:

```
library( ez )  
options( contrasts = c( "contr.sum", "contr.poly" ) )  
ezANOVA( data = tratRe, dv = faltaRac,  
          wid = id, within = mes,  
          type = 3 )
```

## Ejemplo tratamiento 3. Función ezANOVA del paquete ez

```
## Registered S3 methods overwritten by 'lme4':
##   method                                from
##   cooks.distance.influence.merMod      car
##   influence.merMod                     car
##   dfbeta.influence.merMod              car
##   dfbetas.influence.merMod             car

## $ANOVA
##   Effect DFn DFd      F      p p<.05      ges
## 2     mes    2  34 142.0733 3.093681e-17 * 0.5913851
##
## $`Mauchly's Test for Sphericity`
##   Effect      W      p p<.05
## 2     mes 0.523408 0.005632782 *
##
## $`Sphericity Corrections`
##   Effect      GGe      p[GG] p[GG]<.05      HFe      p[HF] p[HF]<.05
## 2     mes 0.6772352 2.298217e-12 * 0.7152351 6.12003e-13 *
```

# ANOVA de medidas repetidas con R. ezANOVA() correcciones

Si el test de Mauchly nos da significativo quiere decir que no podemos asumir esfericidad y hay que considerar las correcciones.  
Pautas:

- Greenhouse–Geisser calcula  $\varepsilon \in \left(\frac{1}{k-1}, 1\right)$  ( $k$  número de grupos)
- Si  $\varepsilon > 0.75$  mirar Huynh–Feld
- Si  $\varepsilon < 0.4$  **no** son válidos los resultados (la violación de la esfericidad afecta a los p-valores)

# Alternativa no paramétrica al Anova de 1 vía con medidas repetidas

- **Test de Friedman** o **ANOVA de Friedman**
- Compara más de dos grupos dependientes

```
friedman.test( Variable ~ Tiempo | id , data = datos )  
pairwise.wilcox.test( df$variable,  
                      df$Tiempo,  
                      p.adjust = "bonferroni",  
                      exact = FALSE,  
                      paired = TRUE )
```



## ANOVA mixto

## ANOVA mixto. Introducción

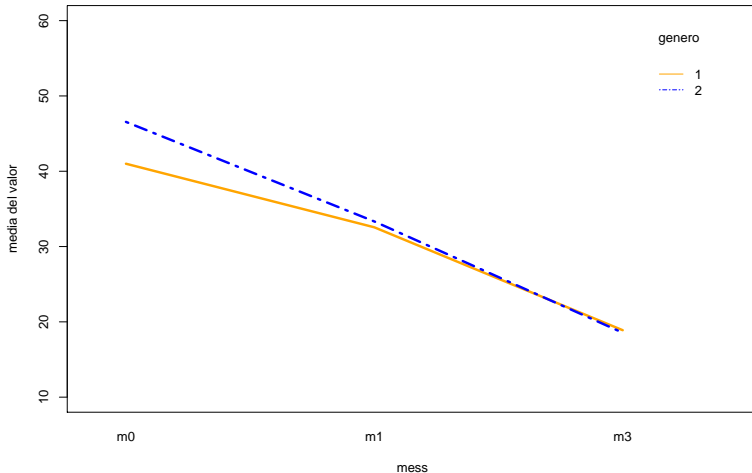
Se utiliza cuando tenemos varios tipos de factores: tanto entre sujetos como intra sujetos.

```
summary( aov( faltaRac ~ genero * mes + Error( id / mes ),  
            data = tratRe ) )
```

## Ejemplo tratamiento 4. Interaction plot

```
interaction.plot( tratRe$mes,  
                  tratRe$genero,  
                  tratRe$faltaRac ,  
                  ylim = c( 10, 60 ),  
                  col = c( "orange", "blue" ),  
                  lty = c( 1, 12 ),  
                  lwd = 3,  
                  ylab = "media del valor",  
                  xlab = "mess", trace.label = "genero")
```

## Ejemplo tratamiento 4. Interaction plot



## Modelos mixtos ANOVA con R. ezANOVA()

```
options( contrasts = c( "contr.sum", "contr.poly" ) )  
ezANOVA( data = tratRe, dv = faltaRac,  
          wid = id, between = genero,  
          within = mes, type = 3 )
```

# Modelo de regresión lineal

## Introducción al modelo de regresión

## Regresión lineal. Introducción

La base de datos `produccion.csv` contiene el tiempo (minutos) que dura un proceso de fabricación de un número de artículos.

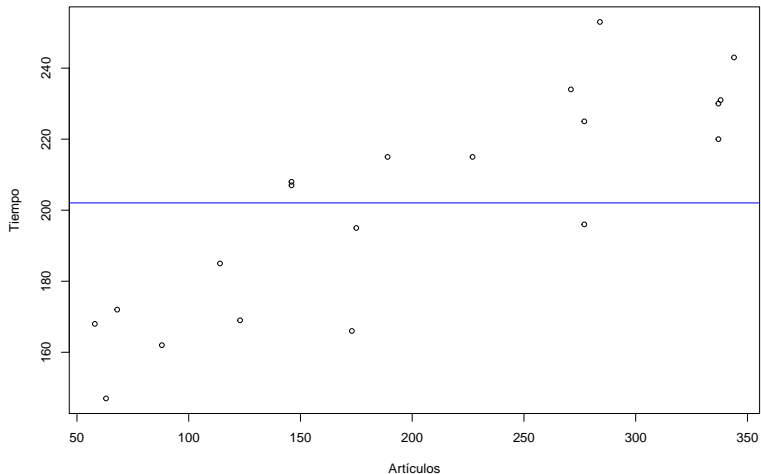
```
prod <- read.table( "files/produccion.csv", sep = ";",  
                    head = TRUE )  
head( prod, 4 )
```

##	caso	tiempo	articulos
## 1	1	195	175
## 2	2	215	189
## 3	3	243	344
## 4	4	162	88

```
# mean( prod$tiempo ) # 202.05
```



# Regresión lineal. Introducción



# Regresión lineal. Introducción

- **La regresión** es el conjunto de técnicas usadas para estudiar la relación entre variables.
- Estamos interesados en
  - conocer **el efecto** que una o varias variables pueden causar sobre otra
  - **predecir** en mayor o menor grado valores de una variable a partir de otra.
- Se trata de una técnica para **explorar** y **cuantificar** la relación de dependencia entre
  - una variable cuantitativa llamada *variable dependiente o respuesta* ( $Y$ )
  - una o más variables independientes llamadas *variables predictoras* ( $X_1, X_2, \dots, X_k$ ).

# Regresión lineal. Introducción

- Modelizaremos la relación lineal entre dos o más variables mediante una **ecuación lineal** de la forma

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_k$$

- Regresión lineal **simple**
  - relación entre dos variables
  - ecuación de una recta  $y = mx + n$ .
- Regresión lineal **múltiple**
  - relación entre tres o más variables
  - un plano o un hiperplano.

## Estimación recta de regresión

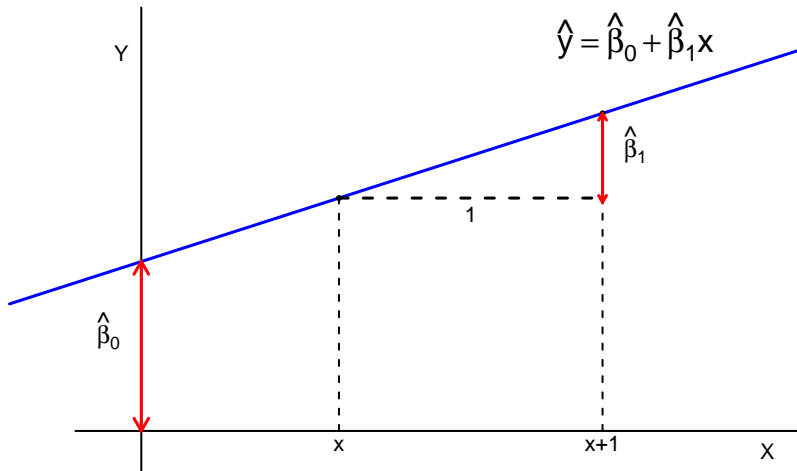
- Este ajuste consiste en estimar **los coeficientes de regresión**  $\beta_0$  y  $\beta_1$  para obtener la recta

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x,$$

donde  $\hat{y}$  es el valor ajustado por el modelo para el valor observado  $x$ .

- $\hat{\beta}_0$  es la **ordenada en el origen** (punto de corte con el eje  $Y$ ).
- $\hat{\beta}_1$  la **pendiente** de la recta del modelo de regresión.

## Ajuste de la recta de regresión



## Regresión lineal simple

# Introducción

- El modelo tiene la forma

$$Y = \beta_0 + \beta_1 X + e,$$

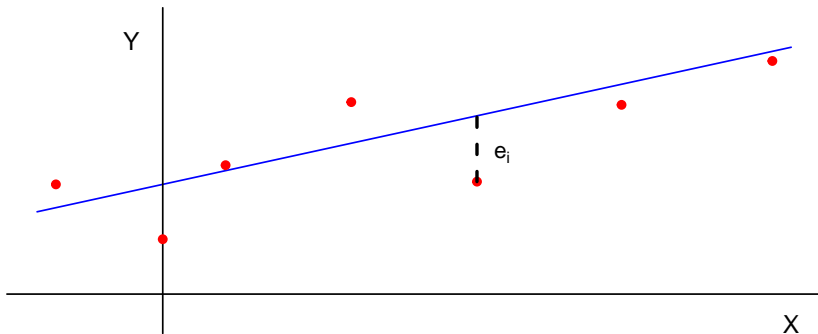
con  $\beta_0$  y  $\beta_1$  los **coeficientes de regresión**

- $\beta_0$  es el valor medio de la variable dependiente cuando la predictora es cero.
- $\beta_1$  es el efecto medio (positivo o negativo) sobre la variable dependiente al aumentar en una unidad el valor de la predictora  $X$ .

# Residuos

- Consideramos **los residuos**, las distancias verticales entre cada punto y la recta

$$e_i = y_i - \hat{y}_i$$





## Método de mínimos cuadrados

- Para estimar la ecuación de la recta de regresión podemos utilizar el **criterio de mínimos cuadrados**, por ser el de mayor aceptación.
- Al **ajustar** cualquier recta a un conjunto de datos existen **pequeñas diferencias** entre
  - los valores estimados por la recta
  - los valores reales observados.

## Método de mínimos cuadrados

- Si sumamos **diferencias positivas y negativas** estas tienden a cancelarse unas con otras.
- Elevamos al cuadrado las diferencias antes de sumarlas.
- Con el criterio de **mínimos cuadrados** calculamos  $\beta_0$  y  $\beta_1$  haciendo **mínima** la suma de los cuadrados de los residuos

$$SS_E = \sum_{i=1}^n e_i^2$$

- Existe una única recta que minimiza los residuos.

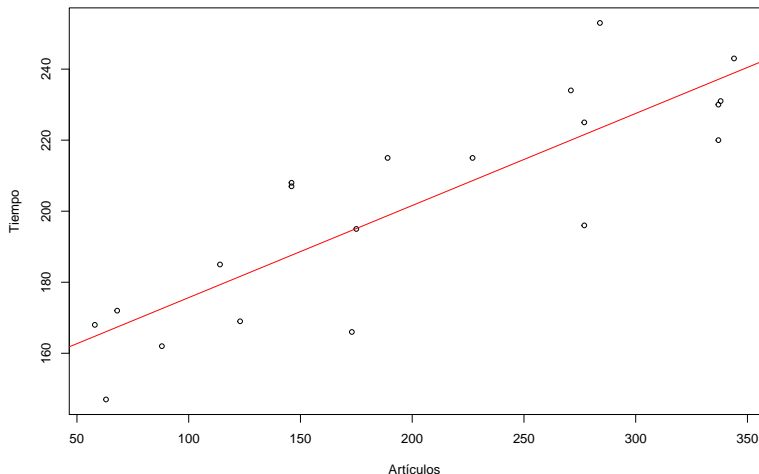
## Ejemplo producción. Ajuste del modelo

```
modelo <- lm( tiempo ~ articulos, data = prod)
modelo

##
## Call:
## lm(formula = tiempo ~ articulos, data = prod)
##
## Coefficients:
## (Intercept)      articulos
##    149.7477         0.2592
```

El modelo ajustado es  $tiempo = 150 + 0.26 \times articulos$

## Ejemplo producción. Recta regresión



## Bondad de ajuste del modelo

- La **variabilidad del modelo** se puede descomponer como  $SS_T = SS_M + SS_R$ .
- El **coeficiente de determinación**  $R^2 = \frac{SS_M}{SS_T}$ 
  - variabilidad total de la respuesta que es explicada por el modelo.
- El **estadístico F** se define como  $F = \frac{MS_M}{MS_R}$ 
  - contrasta si el modelo tiene significativa capacidad predictiva
  - si la  $SS_M$  es suficientemente grande con respecto al número de variables involucradas en el modelo.
- Si  $SS_M$  es grande entonces el modelo de regresión es muy diferente de la media y significa que se ha hecho una gran mejora a la hora de predecir la variable dependiente

## Ejemplo producción. Resumen del modelo

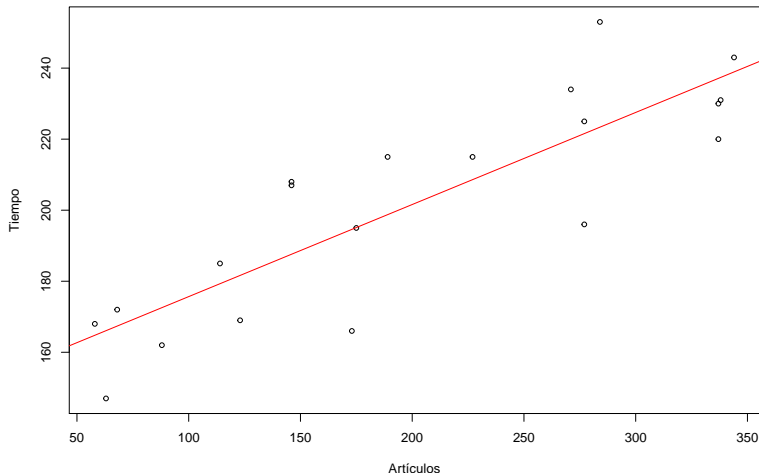
```
summary( modelo )
```

```
##  
## Call:  
## lm(formula = tiempo ~ articulos, data = prod)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -28.597 -11.079   3.329   8.302  29.627   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) 149.74770    8.32815   17.98 6.00e-13 ***  
## articulos    0.25924    0.03714    6.98 1.61e-06 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 16.25 on 18 degrees of freedom  
## Multiple R-squared:  0.7302, Adjusted R-squared:  0.7152   
## F-statistic: 48.72 on 1 and 18 DF,  p-value: 1.615e-06
```

## Ejemplo producción. Modelo ajustado

El modelo ajustado es  $tiempo = 150 + 0.26 \times articulos$

## Ejemplo producción. Recta regresión





## Diagnóstico del modelo

- No basta con ver gráficamente que es un modelo útil.
- Debemos **comprobar ciertos supuestos ‘matemáticos’** que nos hablan de la bondad y calidad del modelo. Las hipótesis son:
  - Linealidad, homocedasticidad e independencia (gráficamente)
  - Media cero, varianza constante, incorrelación y normalidad de **los residuos** (analíticamente).

## Ejemplo producción. Diagnóstico del modelo.

- Obtención de:
  - Valores ajustados: valores ajustados (valores de la variable respuesta) para las observaciones originales de la predictora.
  - Residuos: diferencia entre valor observado de la respuesta y valor ajustado por el modelo.
  - Estadísticos: residuos estudentizados del modelo ajustado.

```
prod$fitted.modelo <- fitted( modelo )  
prod$residuals.modelo <- residuals( modelo )  
prod$rstudent.modelo <- rstudent( modelo )
```

## Ejemplo producción. Normalidad de los residuos.

```
shapiro.test(prod$rstudent.modelo)
```

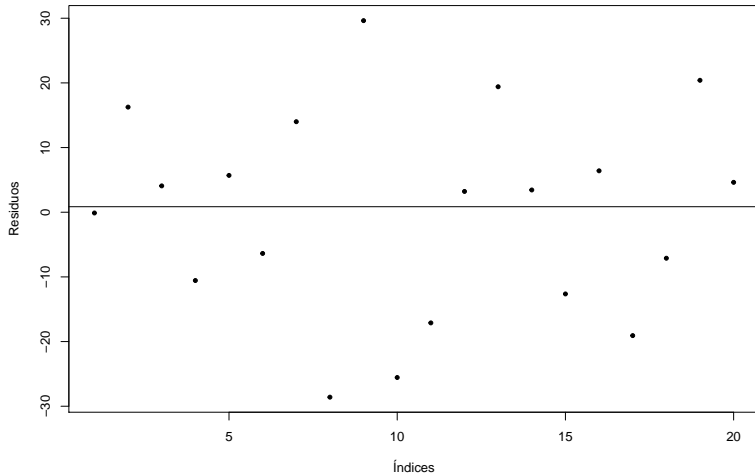
```
##  
## Shapiro-Wilk normality test  
##  
## data:  prod$rstudent.modelo  
## W = 0.97959, p-value = 0.9287
```

## Ejemplo producción. Homocedasticidad de los residuos.

```
library(lmtest)  
bptest(modelo)
```

```
##  
## studentized Breusch-Pagan test  
##  
## data:  modelo  
## BP = 0.10128, df = 1, p-value = 0.7503
```

## Ejemplo producción. Incorrelación de los residuos.



## Ejemplo producción. Incorrelación de los residuos.

- **Test de Durbin-Watson:** asume bajo la hipótesis nula que no existe correlación.

```
dwtest(articulos ~ tiempo, alternative = "two.sided", data = prod)
```

```
##  
## Durbin-Watson test  
##  
## data:  articulos ~ tiempo  
## DW = 2.3145, p-value = 0.4141  
## alternative hypothesis: true autocorrelation is not 0
```

## Referencias y bibliografía

## Referencias y bibliografía



## Referencias y bibliografía

Maurandi-López, A., Balsalobre R, C., & Río-Alonso, L. del. (2013). *Fundamentos estadísticos para investigación. introducción a r*. Bubok Publishing SL. Retrieved from <http://www.bubok.es/libros/223207/>  
Fundamentos-estadisticos-para-investigacionIntroduccion-a-R