

# Introducción a la estadística

## Bases indispensables y uso de

Olivier Devineau

`olivier.devineau@fcdarwin.org.ec`

Fundación Charles Darwin

Taller interno, 27–30 abril 2010

# Correlación y regresión

# Dos categorías de tests estadísticos

Introducción

Correlación

Modelo lineal

Regresión  
lineal

Otros tipos de  
regresión

Criticas a los  
modelos

Tests de comparación : 1 variable,  $\geq 2$  poblaciones

Tests de relación :  $\geq 2$  variables, 1 población

## $\geq 2$ variables es común en biología

Introducción

Correlación

Modelo lineal

Regresión  
lineal

Otros tipos de  
regresión

Criticas a los  
modelos

### 2 variables para el mismo individuo

- Presión sanguínea  $X_1$ , peso  $X_2$
- Abundancia de una especie de planta  $X_1$ , nivel del pH en el suelo  $X_2$ , temperatura  $X_3$
- Datos **b**ivariados o **multi**variados

⇒ ¿Cuál es la relación entre las variables?

# Relación entre $\geq 2$ variables

## La estadística correlacional

### Varios tipos de relación

- No conexión
- Relación | *handout* :  $1 > 0$  /  $< 0$ , causal / no
- Conexión funcional  $\rightarrow$  predicción

### Objetivo de la estadística correlacional

- Determinar validez y fuerza de la relación entre las variables
- Determinar la dirección de la relación

# Estadística correlacional

**Correlación:** ¿Cómo 2 variables varían juntas?

**Regresión:** Relación entre 1 variable dependiente y  $\geq 1$  variable independiente

**Análisis multivariados:** Relación entre  $\geq 2$  variables independientes / dependientes / ambos

# Noción de correlación

## Ejemplo

- 1 población: 2 variables continuas
  - Presión sanguínea  $X_1$ , peso  $X_2$
  - Cada muestra  $i$  : 1 valor por cada variable:  $x_{i_1}$  y  $x_{i_2}$
- 
- ¿La presión sanguínea y el peso son correlativas?

# Noción de correlación (2)

## Definición

**Correlación** se define en terminos de:

- Varianza de  $X_1$ :  $var(X_1)$
- Varianza de  $X_2$ :  $var(X_2)$
- ¿Como  $X_1$  y  $X_2$  varían juntas? Covarianza:  $cov(X_1, X_2)$

⇒ Coeficiente de correlación

$$r = \frac{cov(X_1, X_2)}{\sqrt{var(X_1) \cdot var(X_2)}}$$



# Noción de correlación (2)

## Definición

**Correlación** se define en terminos de:

- Varianza de  $X_1$ :  $var(X_1)$
- Varianza de  $X_2$ :  $var(X_2)$
- ¿Como  $X_1$  y  $X_2$  varían juntas? Covarianza:  $cov(X_1, X_2)$

⇒ Coeficiente de correlación

$$r = \frac{cov(X_1, X_2)}{\sqrt{var(X_1) \cdot var(X_2)}}$$

# Noción de correlación (2)

## Definición

**Correlación** se define en terminos de:

- Varianza de  $X_1$ :  $var(X_1)$
- Varianza de  $X_2$ :  $var(X_2)$
- ¿Como  $X_1$  y  $X_2$  varían juntas? Covarianza:  $cov(X_1, X_2)$

⇒ Coeficiente de correlación

$$r = \frac{cov(X_1, X_2)}{\sqrt{var(X_1) \cdot var(X_2)}}$$

# Noción de correlación (2)

## Definición

**Correlación** se define en terminos de:

- Varianza de  $X_1$ :  $var(X_1)$
- Varianza de  $X_2$ :  $var(X_2)$
- ¿Como  $X_1$  y  $X_2$  varían juntas? Covarianza:  $cov(X_1, X_2)$

⇒ Coeficiente de correlación

$$r = \frac{cov(X_1, X_2)}{\sqrt{var(X_1) \cdot var(X_2)}}$$

# El coeficiente de correlación $r$

Correlación de Pearson (paramétrica)

- No unidad
- $r \in [-1, 1]$
- Magnitud: fuerza de la relación
- Signo: dirección de la relación
- Muestra:  $r$ , Población:  $\rho$

# ¿Qué test para chequear la correlación?

$X_1$ : Presión sanguínea y  $X_2$ : peso

- ¿Hipótesis nula?

# ¿Qué test para chequear la correlación?

$X_1$ : Presión sanguínea y  $X_2$ : peso

- ¿Hipótesis nula?
- No hay una relación lineal entre la presión sanguínea y el peso

# ¿Qué test para chequear la correlación?

$X_1$ : Presión sanguínea y  $X_2$ : peso

- ¿Hipótesis nula?
- No hay una relación lineal entre la presión sanguínea y el peso
- $H_0 : \rho = 0$

# ¿Qué test para chequear la correlación?

$X_1$ : Presión sanguínea y  $X_2$ : peso

- ¿Hipótesis nula?
- No hay una relación lineal entre la presión sanguínea y el peso
- $H_0 : \rho = 0$
- Cuando  $H_0$  es verdadera,  $r \rightsquigarrow \mathcal{N}(\mu, \sigma)$



# ¿Qué test para chequear la correlación?

$X_1$ : Presión sanguínea y  $X_2$ : peso

- ¿Hipótesis nula?
- No hay una relación lineal entre la presión sanguínea y el peso
- $H_0 : \rho = 0$
- Cuando  $H_0$  es verdadera,  $r \rightsquigarrow \mathcal{N}(\mu, \sigma)$   
 $\Rightarrow$  uso de test  $t$  de Student

# Correlación no paramétrica

## Introducción

## Correlación

Noción de  
correlación

Coefficiente de  
correlación

Test

Observaciones

## Modelo lineal

## Regresión lineal

## Otros tipos de regresión

## Criticas a los modelos

- ¿Qué hacer cuando los requisitos no se cumplen?

⇒ Coeficiente de correlación de rango

- de Spearman:  $\rho$
- de Kendall:  $\tau$

- ¡Más conservadores!

# Correlación no paramétrica

- ¿Qué hacer cuando los requisitos no se cumplen?
- ⇒ Coeficiente de correlación de rango
- de Spearman:  $\rho$
  - de Kendall:  $\tau$
- ¡Más conservadores!

# Correlación no paramétrica

## Introducción

## Correlación

Noción de  
correlación

Coefficiente de  
correlación

Test

Observaciones

## Modelo lineal

## Regresión lineal

Otros tipos de  
regresión

Criticas a los  
modelos

- ¿Qué hacer cuando los requisitos no se cumplen?
- ⇒ Coeficiente de correlación de rango
- de Spearman:  $\rho$
  - de Kendall:  $\tau$
- ¡Más conservadores!

# Correlación no paramétrica

## Introducción

## Correlación

Noción de  
correlación

Coefficiente de  
correlación

Test

Observaciones

## Modelo lineal

## Regresión lineal

Otros tipos de  
regresión

Criticas a los  
modelos

- ¿Qué hacer cuando los requisitos no se cumplen?
- ⇒ Coeficiente de correlación de rango
- de Spearman:  $\rho$
  - de Kendall:  $\tau$
- ¡Más conservadores!

# Correlación no paramétrica

## Introducción

## Correlación

Noción de  
correlación

Coefficiente de  
correlación

Test

Observaciones

## Modelo lineal

## Regresión lineal

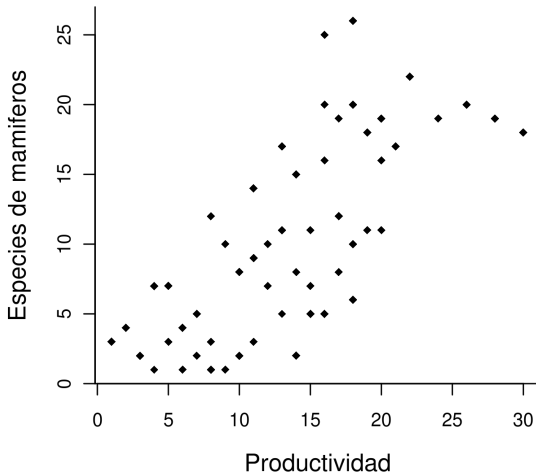
Otros tipos de  
regresión

Criticas a los  
modelos

- ¿Qué hacer cuando los requisitos no se cumplen?
- ⇒ Coeficiente de correlación de rango
- de Spearman:  $\rho$
  - de Kendall:  $\tau$
- ¡Más conservadores!

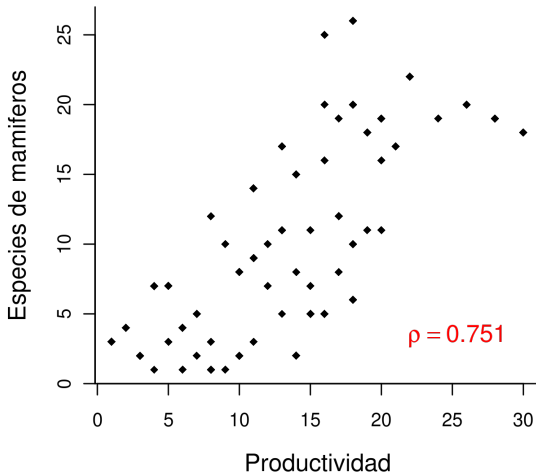
# La correlación depende de la escala

¡Las cosas no son siempre como parecen!



# La correlación depende de la escala

¡Las cosas no son siempre como parecen!

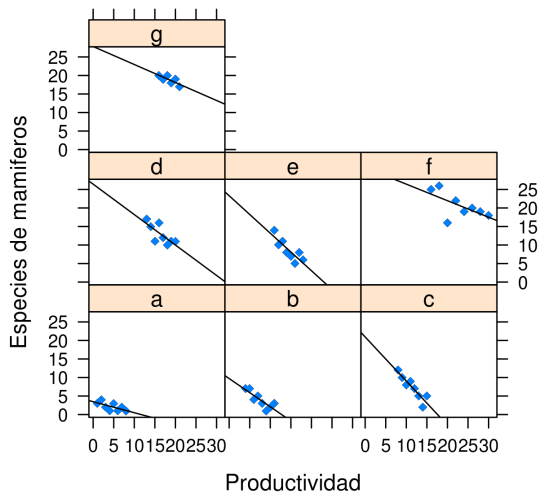




# La correlación depende de la escala

¡Las cosas no son siempre como parecen!

- Introducción
- Correlación
  - Noción de correlación
  - Coefficiente de correlación
  - Test
  - Observaciones
- Modelo lineal
- Regresión lineal
- Otros tipos de regresión
- Criticas a los modelos



# Modelo lineal: concepto general

- Se puede identificar:

# Modelo lineal: concepto general

- Se puede identificar:
  - 1 variable **respuesta** / **dependiente**  $Y$

# Modelo lineal: concepto general

- Se puede identificar:
  - 1 variable **respuesta** / **dependiente**  $Y$
  - $\geq 1$  variable **explicativa** / **predictiva** / **independiente** / **covariable**  $X_1, X_2, \dots$

# Modelo lineal: concepto general

- Se puede identificar:
  - 1 variable **respuesta** / **dependiente**  $Y$
  - $\geq 1$  variable **explicativa** / **predictiva** / **independiente** / **covariable**  $X_1, X_2, \dots$
- Cada unidad de muestra:  $y_i, x_{1_i}, x_{2_i} \dots$

# Modelo lineal: concepto general

- Se puede identificar:
  - 1 variable **respuesta** / **dependiente**  $Y$
  - $\geq 1$  variable **explicativa** / **predictiva** / **independiente** / **covariable**  $X_1, X_2, \dots$
- Cada unidad de muestra:  $y_i, x_{1i}, x_{2i} \dots$
- Explicar el patrón de  $Y$  con  $X$

# Modelo lineal

## Forma general de los modelos estadísticos

- *Variable dependiente = modelo + error*

# Modelo lineal

## Forma general de los modelos estadísticos

- *Variable dependiente = modelo + error*
- Modelo: covariables y **parámetros**



# Modelo lineal

## Forma general de los modelos estadísticos

- *Variable dependiente = modelo + error*
- Modelo: covariables y parámetros
- Covariables: continuas / categoricas / ambos

# Modelo lineal

## Forma general de los modelos estadísticos

- *Variable dependiente = modelo + error*
- Modelo: covariables y parámetros
- Covariables: continuas / categoricas / ambos
- Error: parte de la variable dependiente **que no esta explicada** por el modelo

# Modelo lineal

## Forma general de los modelos estadísticos

- *Variable dependiente = modelo + error*
- Modelo: covariables y parámetros
- Covariables: continuas / categoricas / ambos
- Error: parte de la variable dependiente que no esta explicada por el modelo
- Se supone una **distribución** para el componente del error, y de ahí para la variable dependiente  $Y$

# ¿Qué significa lineal?

- Relación de línea recta entre 2 variables

# ¿Qué significa lineal?

- Relación de línea recta entre 2 variables
- Combinación lineal de parámetros

# ¿Qué significa lineal?

- Relación de línea recta entre 2 variables
- Combinación lineal de parámetros
- No exponente, no multiplicación por otro parámetro

## ¿Qué significa lineal?

- Relación de línea recta entre 2 variables
- Combinación lineal de parámetros
- No exponente, no multiplicación por otro parámetro
- $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

# Análisis de regresión lineal

## Contexto

- Usar **datos** de una **muestra** para **estimar** valores de **parámetros** y sus errores estándar



# Análisis de regresión lineal

## Contexto

- Usar datos de una muestra para estimar valores de parámetros y sus errores estándar
- ¿Cuándo se usa?

# Análisis de regresión lineal

## Contexto

- Usar datos de una muestra para estimar valores de parámetros y sus errores estándar
- ¿Cuándo se usa?
- Variables explicativa y dependiente son **continuas**

# Análisis de regresión lineal

## Contexto

- Usar datos de una muestra para estimar valores de parámetros y sus errores estándar
- ¿Cuándo se usa?
- Variables explicativa y dependiente son continuas
- Altura, peso, volumen, temperatura ...

# Análisis de regresión lineal

## Contexto

- Usar datos de una muestra para estimar valores de parámetros y sus errores estándar
- ¿Cuándo se usa?
- Variables explicativa y dependiente son continuas
- Altura, peso, volumen, temperatura ...
- Nube de puntos → regresión lineal

# Análisis de regresión lineal

## Objetivos

- Describir la relación lineal entre  $Y$  y  $X$
- Determinar cuánto de la variación en  $Y$  se explica por la relación lineal con  $X$  y cuánto de esta variación no se puede explicar
- Predecir nuevos valores de  $Y$  a partir de valores de  $X$

# Análisis de regresión lineal

## Objetivos

- Describir la relación lineal entre  $Y$  y  $X$
- Determinar cuánto de la variación en  $Y$  se explica por la relación lineal con  $X$  y cuánto de esta variación no se puede explicar
- Predecir nuevos valores de  $Y$  a partir de valores de  $X$

# Análisis de regresión lineal

## Objetivos

Introducción

Correlación

Modelo lineal

Regresión  
lineal

**Regresión**

Estimación

Evaluación del  
ajuste

Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos

- Describir la relación lineal entre  $Y$  y  $X$
- Determinar cuánto de la variación en  $Y$  se explica por la relación lineal con  $X$  y cuánto de esta variación no se puede explicar
- Predecir nuevos valores de  $Y$  a partir de valores de  $X$

# Análisis de regresión lineal

## Varios tipos de regresión

- Regresión lineal: lo más simple y frecuente
- Regresión polinomial: chequear si una relación es no lineal
- Regresión no lineal
- Regresión no paramétrica: si no hay forma funcional



# Principio de la regresión lineal

Introducción

Correlación

Modelo lineal

Regresión  
lineal

**Regresión**

Estimación

Evaluación del  
ajuste

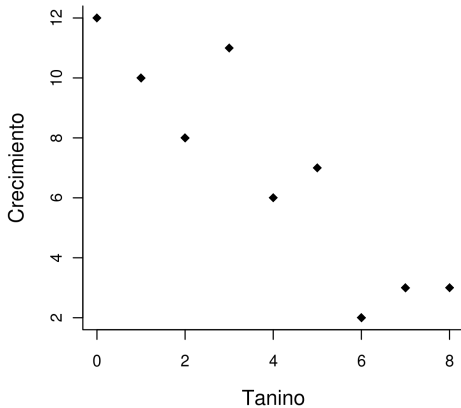
Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos

• Datos



# Principio de la regresión lineal

Introducción

Correlación

Modelo lineal

Regresión  
lineal

**Regresión**

Estimación

Evaluación del  
ajuste

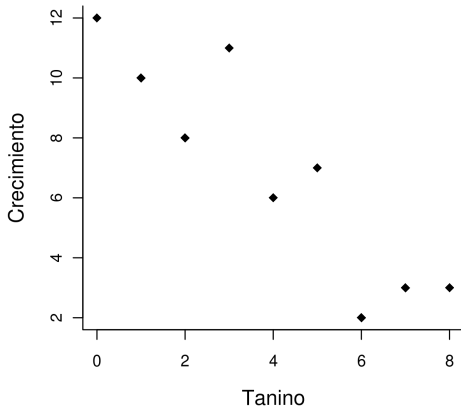
Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos

- Datos
- Modelo:  $y = a + bx$



# Principio de la regresión lineal

Introducción

Correlación

Modelo lineal

Regresión  
lineal

Regresión

Estimación

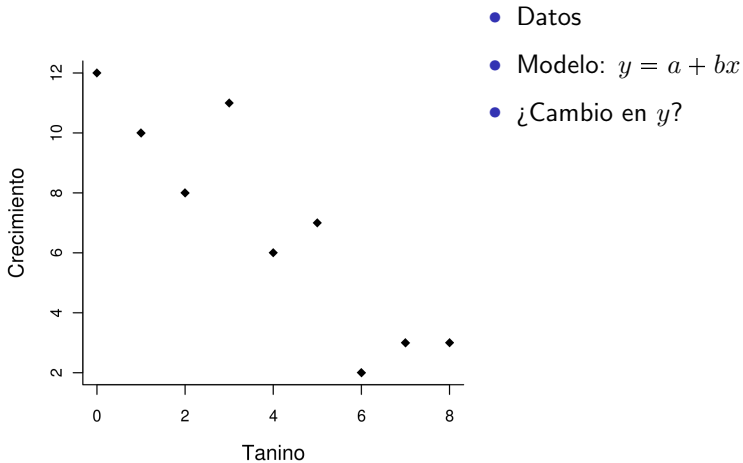
Evaluación del  
ajuste

Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos



# Principio de la regresión lineal

Introducción

Correlación

Modelo lineal

Regresión  
lineal

Regresión

Estimación

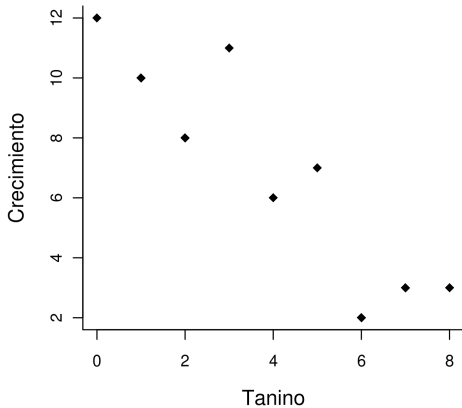
Evaluación del  
ajuste

Comparación de  
modelos

Condiciones

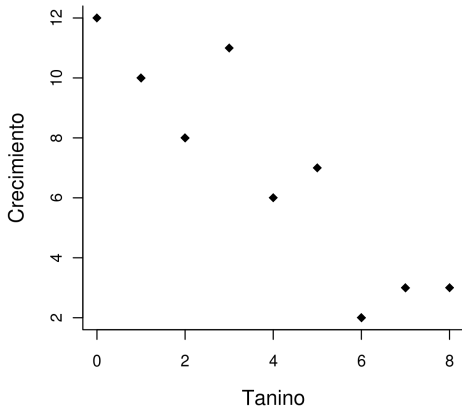
Otros tipos de  
regresión

Criticas a los  
modelos



- Datos
- Modelo:  $y = a + bx$
- ¿Cambio en  $y$ ?  
 $\delta y = -10$

# Principio de la regresión lineal



- Datos
- Modelo:  $y = a + bx$
- ¿Cambio en  $y$ ?  
 $\delta y = -10$
- ¿Cambio en  $x$ ?

# Principio de la regresión lineal

Introducción

Correlación

Modelo lineal

Regresión  
lineal

Regresión

Estimación

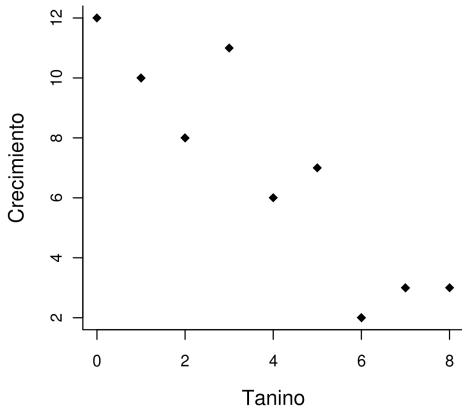
Evaluación del  
ajuste

Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos



- Datos
- Modelo:  $y = a + bx$
- ¿Cambio en  $y$ ?  
 $\delta y = -10$
- ¿Cambio en  $x$ ?  
 $\delta x = +8$

# Principio de la regresión lineal

Introducción

Correlación

Modelo lineal

Regresión  
lineal

Regresión

Estimación

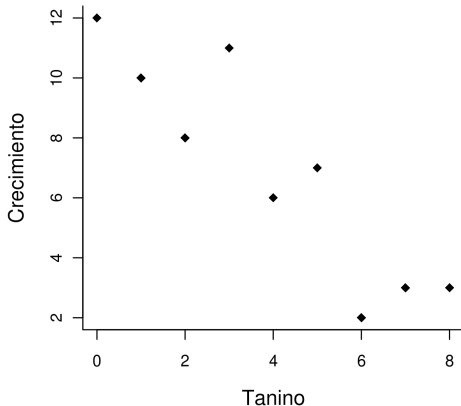
Evaluación del  
ajuste

Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos



- Datos
- Modelo:  $y = a + bx$
- ¿Cambio en  $y$ ?  
 $\delta y = -10$
- ¿Cambio en  $x$ ?  
 $\delta x = +8$
- Pendiente  
 $b = \delta y / \delta x = -1.25$

# Principio de la regresión lineal

Introducción

Correlación

Modelo lineal

Regresión  
lineal

Regresión

Estimación

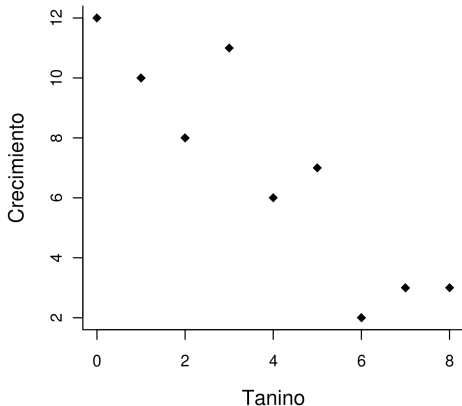
Evaluación del  
ajuste

Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos



- Datos
- Modelo:  $y = a + bx$
- ¿Cambio en  $y$ ?  
 $\delta y = -10$
- ¿Cambio en  $x$ ?  
 $\delta x = +8$
- Pendiente  
 $b = \delta y / \delta x = -1.25$
- ¿Ordenada al origen?



# Principio de la regresión lineal

Introducción

Correlación

Modelo lineal

Regresión  
lineal

Regresión

Estimación

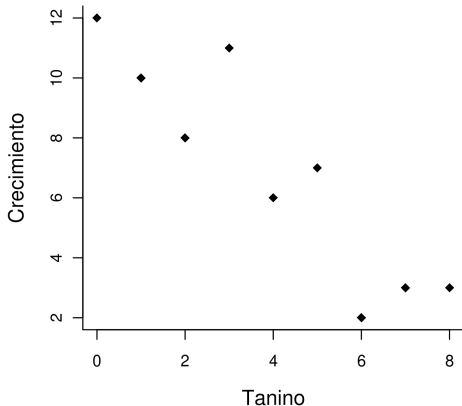
Evaluación del  
ajuste

Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos



- Datos
- Modelo:  $y = a + bx$
- ¿Cambio en  $y$ ?  
 $\delta y = -10$
- ¿Cambio en  $x$ ?  
 $\delta x = +8$
- Pendiente  
 $b = \delta y / \delta x = -1.25$
- ¿Ordenada al origen?  
 $a = 12$

# Principio de la regresión lineal

Introducción

Correlación

Modelo lineal

Regresión  
lineal

Regresión

Estimación

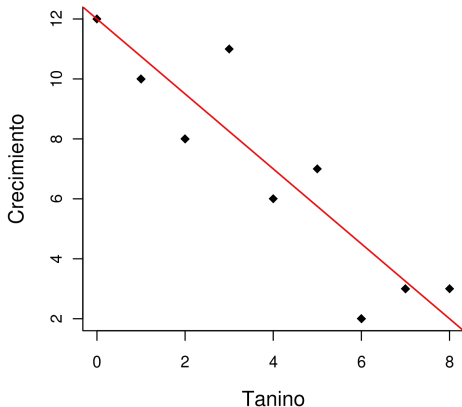
Evaluación del  
ajuste

Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos



- Datos
- Modelo:  $y = a + bx$
- ¿Cambio en  $y$ ?  
 $\delta y = -10$
- ¿Cambio en  $x$ ?  
 $\delta x = +8$
- Pendiente  
 $b = \delta y / \delta x = -1.25$
- ¿Ordenada al origen?  
 $a = 12$
- $y = 12 - 1.25x$

# Principio de la regresión lineal (2)

Introducción

Correlación

Modelo lineal

Regresión  
lineal

**Regresión**

Estimación

Evaluación del  
ajuste

Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos

- Ajustar un modelo a los datos
- Estimar los parámetros del modelo
- Probar varios valores de parámetros hasta encontrar el mejor modelo
- Máxima verosimilitud (Maximum Likelihood ML)
- Mínimos cuadrados (Ordinary Least Square OLS)

# Principio de la regresión lineal (2)

Introducción

Correlación

Modelo lineal

Regresión  
lineal

**Regresión**

Estimación

Evaluación del  
ajuste

Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos

- Ajustar un modelo a los datos
- Estimar los parámetros del modelo
- Probar varios valores de parámetros hasta encontrar el mejor modelo
- Máxima verosimilitud (Maximum Likelihood ML)
- Mínimos cuadrados (Ordinary Least Square OLS)

# Principio de la regresión lineal (2)

Introducción

Correlación

Modelo lineal

Regresión  
lineal

**Regresión**

Estimación

Evaluación del  
ajuste

Comparación de  
modelos  
Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos

- Ajustar un modelo a los datos
- Estimar los parámetros del modelo
- Probar varios valores de parámetros hasta encontrar el mejor modelo
- Máxima verosimilitud (Maximum Likelihood ML)
- Mínimos cuadrados (Ordinary Least Square OLS)

# Principio de la regresión lineal (2)

Introducción

Correlación

Modelo lineal

Regresión  
lineal

**Regresión**

Estimación

Evaluación del  
ajuste

Comparación de  
modelos  
Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos

- Ajustar un modelo a los datos
- Estimar los parámetros del modelo
- Probar varios valores de parámetros hasta encontrar el mejor modelo
- Máxima verosimilitud (Maximum Likelihood ML)
- Mínimos cuadrados (Ordinary Least Square OLS)

# Principio de la regresión lineal (2)

Introducción

Correlación

Modelo lineal

Regresión  
lineal

**Regresión**

Estimación

Evaluación del  
ajuste

Comparación de  
modelos  
Condiciones

Otros tipos de  
regresión

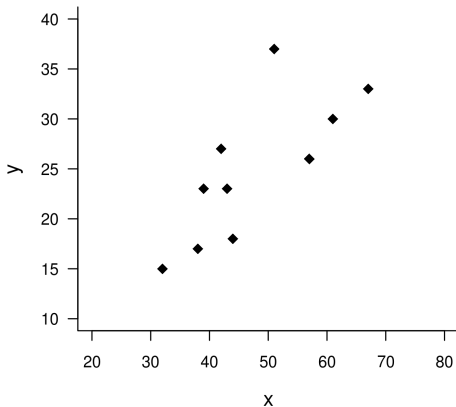
Criticas a los  
modelos

- Ajustar un modelo a los datos
- Estimar los parámetros del modelo
- Probar varios valores de parámetros hasta encontrar el mejor modelo
- Máxima verosimilitud (Maximum Likelihood ML)
- Mínimos cuadrados (Ordinary Least Square OLS)

# Cuadrados mínimos: principio

OLS: Ordinary Least Squares

• Datos



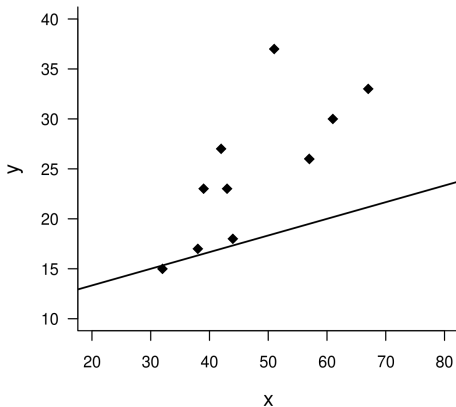


# Cuadrados mínimos: principio

OLS: Ordinary Least Squares

- Datos

- Modelo  $y = 10 + 1/6x$

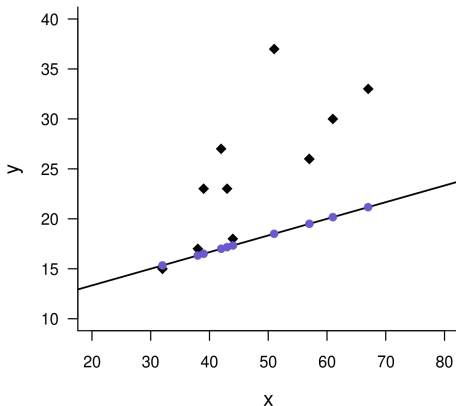


# Cuadrados mínimos: principio

OLS: Ordinary Least Squares

• Datos

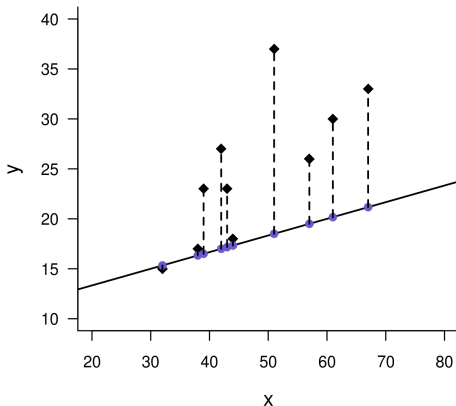
• Modelo  $y = 10 + 1/6x$



# Cuadrados mínimos: principio

OLS: Ordinary Least Squares

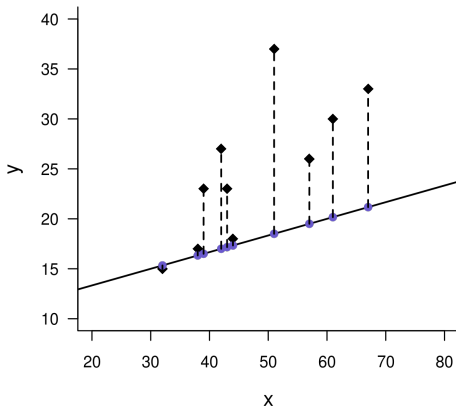
- Datos
- Modelo  $y = 10 + 1/6x$
- Residual  $e_i = y_i - \hat{y}_i$



# Cuadrados mínimos: principio

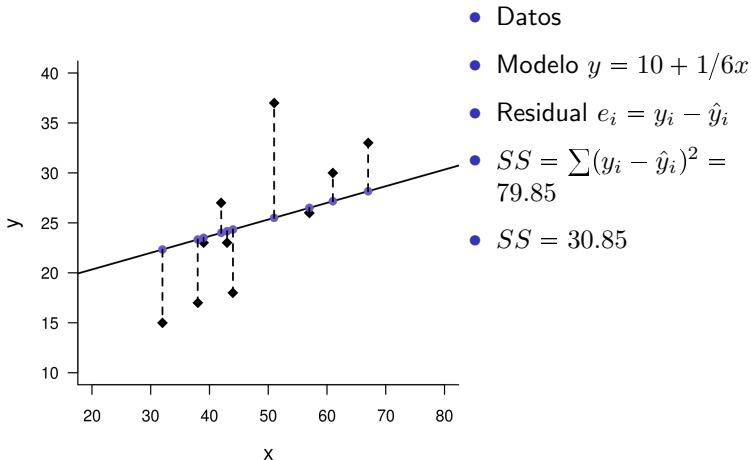
OLS: Ordinary Least Squares

- Datos
- Modelo  $y = 10 + 1/6x$
- Residual  $e_i = y_i - \hat{y}_i$
- $SS = \sum (y_i - \hat{y}_i)^2 = 79.85$



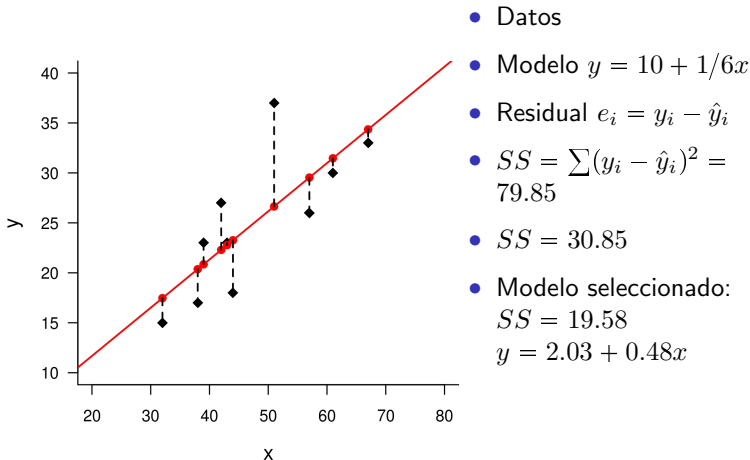
# Cuadrados mínimos: principio

OLS: Ordinary Least Squares



# Cuadrados mínimos: principio

OLS: Ordinary Least Squares



# Hipótesis nula en regresión

- ¿Cuál sería  $H_0$ ?

# Hipótesis nula en regresión

- ¿Cuál sería  $H_0$ ?
- No hay una relación lineal entre las variables



# Hipótesis nula en regresión

- ¿Cuál sería  $H_0$ ?
- No hay una relación lineal entre las variables
- Pendiente  $b = 0$

# Hipótesis nula en regresión

- ¿Cuál sería  $H_0$ ?
- No hay una relación lineal entre las variables
- Pendiente  $b = 0$   
→ Test de Fisher:  $F$

# Hipótesis nula en regresión

- ¿Cuál sería  $H_0$ ?
- No hay una relación lineal entre las variables
- Pendiente  $b = 0$ 
  - Test de Fisher:  $F$
  - Test de Student:  $t$

# Varianza explicada

$r^2$ : coeficiente de determinación

- Variación de  $Y$  explicada por la relación con  $X$ 
  - $(\text{coeficiente de correlación})^2$
  - $r^2 \in [0, 1]$
  - ¿Como se mejora el ajuste del modelo con pendiente comparado a un modelo sin pendiente?
  - $r^2$  inadecuado para comparar modelos con números de parámetros diferentes

# Varianza explicada

$r^2$ : coeficiente de determinación

- Variación de  $Y$  explicada por la relación con  $X$
- $(\text{coeficiente de correlación})^2$
- $r^2 \in [0, 1]$
- ¿Como se mejora el ajuste del modelo con pendiente comparado a un modelo sin pendiente?
- $r^2$  inadecuado para comparar modelos con números de parámetros diferentes

# Varianza explicada

$r^2$ : coeficiente de determinación

- Variación de  $Y$  explicada por la relación con  $X$
- $(\text{coeficiente de correlación})^2$
- $r^2 \in [0, 1]$
- ¿Como se mejora el ajuste del modelo con pendiente comparado a un modelo sin pendiente?
- $r^2$  inadecuado para comparar modelos con números de parámetros diferentes

# Varianza explicada

$r^2$ : coeficiente de determinación

- Variación de  $Y$  explicada por la relación con  $X$
- $(\text{coeficiente de correlación})^2$
- $r^2 \in [0, 1]$
- ¿Como se mejora el ajuste del modelo con pendiente comparado a un modelo sin pendiente?
- $r^2$  inadecuado para comparar modelos con números de parámetros diferentes

# Varianza explicada

$r^2$ : coeficiente de determinación

- Variación de  $Y$  explicada por la relación con  $X$
- $(\text{coeficiente de correlación})^2$
- $r^2 \in [0, 1]$
- ¿Como se mejora el ajuste del modelo con pendiente comparado a un modelo sin pendiente?
- $r^2$  inadecuado para comparar modelos con números de parámetros diferentes



# Comparar varios modelos

Introducción

Correlación

Modelo lineal

Regresión  
lineal

Regresión  
Estimación  
Evaluación del  
ajuste  
**Comparación de  
modelos**  
Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos

- Evaluar varias hipótesis → varios modelos
- $H_0$ : modelo simple,  $H_1$ : modelo más complejo
- Hay que comparar los modelos

# Comparar modelos de regresión

Introducción

Correlación

Modelo lineal

Regresión  
lineal

Regresión

Estimación

Evaluación del  
ajuste

Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos

## Minimos cuadrados (OLS)

- Ajuste: proporción de varianza explicada
- No-ajuste: proporción de varianza residual

⇒ Análisis de varianza

# Comparar modelos de regresión

Introducción

Correlación

Modelo lineal

Regresión  
lineal

Regresión

Estimación

Evaluación del  
ajuste

Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos

## Minimos cuadrados (OLS)

- Ajuste: proporción de varianza explicada
- No-ajuste: proporción de varianza residual

⇒ Análisis de varianza

# Comparar modelos de regresión

Introducción

Correlación

Modelo lineal

Regresión  
lineal

Regresión

Estimación

Evaluación del  
ajuste

Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos

## Minimos cuadrados (OLS)

- Ajuste: proporción de varianza explicada
- No-ajuste: proporción de varianza residual

⇒ **Análisis de varianza**

# Comparar modelos de regresión

Introducción

Correlación

Modelo lineal

Regresión  
lineal

Regresión

Estimación

Evaluación del  
ajuste

Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos

## Minimos cuadrados (OLS)

- Ajuste: proporción de varianza explicada
- No-ajuste: proporción de varianza residual

⇒ Análisis de varianza

## Máxima verosimilitud (ML)

- Ajuste: tamaño de la verosimilitud

⇒ Prueba de la razón de verosimilitud (Likelihood Ratio Test o AIC)

# Comparar modelos de regresión

Introducción

Correlación

Modelo lineal

Regresión  
lineal

Regresión

Estimación

Evaluación del  
ajuste

Comparación de  
modelos

Condiciones

Otros tipos de  
regresión

Criticas a los  
modelos

## Minimos cuadrados (OLS)

- Ajuste: proporción de varianza explicada
- No-ajuste: proporción de varianza residual

⇒ Análisis de varianza

## Máxima verosimilitud (ML)

- Ajuste: tamaño de la verosimilitud

⇒ Prueba de la razón de verosimilitud (**Likelihood Ratio Test**  
o **AIC**)

## Comparar modelos de regresión (2)

Siempre la misma lógica

- Medir el ajuste de cada modelo
- Comparar los ajustes de diferente modelos para examinar hipótesis sobre los parámetros

### Ejemplo: presión sanguínea y peso

- Modelo 1:  $P = \beta_0 + \varepsilon$
- Modelo 2:  $P = \beta_0 + \beta_1 * peso + \varepsilon$
- Comparar  $M_1$  y  $M_2$  es equivalente a evaluar  $H_0 : \beta_1 = 0$

# Condiciones del análisis de regresión (1)

- Involucran de los términos de errores ( $\varepsilon_i$ )
- De la variable dependiente  $Y$
- Importantes para intervalos de confianza
- Importantes para tests de hipótesis con distribución  $t$  o  $F$
- Residuales importantes para chequear condiciones



# Condiciones del análisis de regresión (2)

- **Normalidad:**  $\varepsilon$  tiene una distribución normal
- **Homogeneidad de la varianza:**  $\varepsilon$  tiene la misma varianza por cada  $x_i$ :  $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_i^2 = \dots = \sigma_\varepsilon^2$
- **Independencia:**  $\varepsilon$  son independientes: Los valores de  $Y$  para cualquier  $x_i$  no influyen los valores de  $Y$  para otra  $x_i$

# Condiciones del análisis de regresión (2)

- Normalidad:  $\varepsilon$  tiene una distribución normal
- Homogeneidad de la varianza:  $\varepsilon$  tiene la misma varianza por cada  $x_i$ :  $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_i^2 = \dots = \sigma_\varepsilon^2$
- Independencia:  $\varepsilon$  son independientes: Los valores de  $Y$  para cualquier  $x_i$  no influyen los valores de  $Y$  para otra  $x_i$

# Condiciones del análisis de regresión (2)

- Normalidad:  $\varepsilon$  tiene una distribución normal
- Homogeneidad de la varianza:  $\varepsilon$  tiene la misma varianza por cada  $x_i$ :  $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_i^2 = \dots = \sigma_\varepsilon^2$
- Independencia:  $\varepsilon$  son independientes: Los valores de  $Y$  para cualquier  $x_i$  no influyen los valores de  $Y$  para otra  $x_i$

# Homogeneidad de la varianza

Introducción

Correlación

Modelo lineal

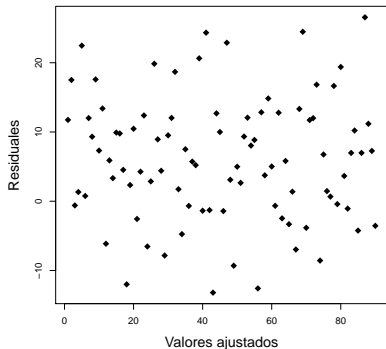
Regresión  
lineal

Regresión  
Estimación  
Evaluación del  
ajuste  
Comparación de  
modelos  
**Condiciones**

Otros tipos de  
regresión

Criticas a los  
modelos

- No tendencia



# Homogeneidad de la varianza

Introducción

Correlación

Modelo lineal

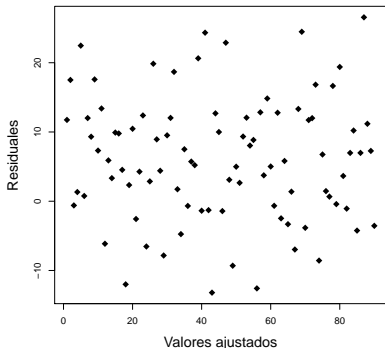
Regresión  
lineal

Regresión  
Estimación  
Evaluación del  
ajuste  
Comparación de  
modelos  
Condiciones

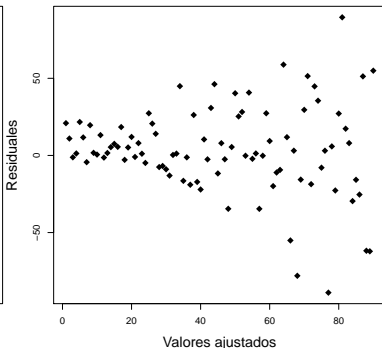
Otros tipos de  
regresión

Criticas a los  
modelos

- No tendencia



- Heteroscedasticidad



# Homogeneidad de la varianza

Introducción

Correlación

Modelo lineal

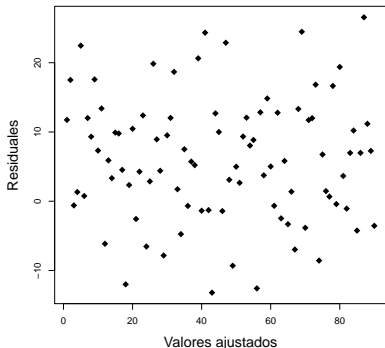
Regresión  
lineal

Regresión  
Estimación  
Evaluación del  
ajuste  
Comparación de  
modelos  
Condiciones

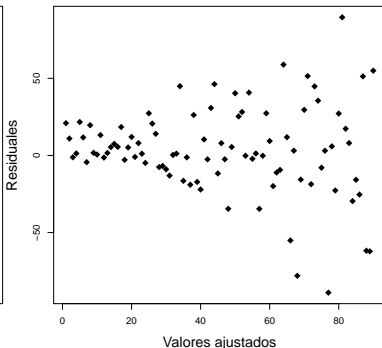
Otros tipos de  
regresión

Criticas a los  
modelos

- No tendencia



- Heteroscedasticidad



- Test de Levene, test de Bartlett

# Normalidad de los residuales

Introducción

Correlación

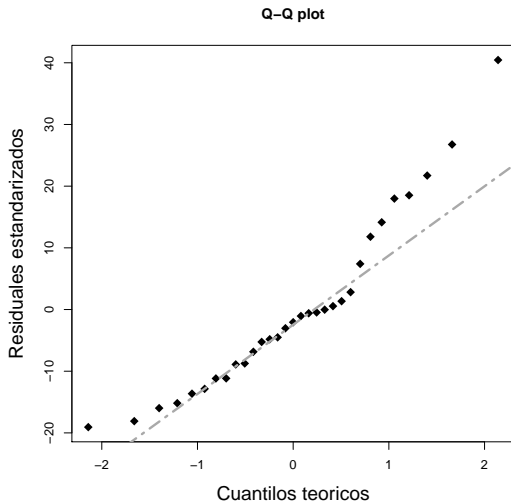
Modelo lineal

Regresión  
lineal

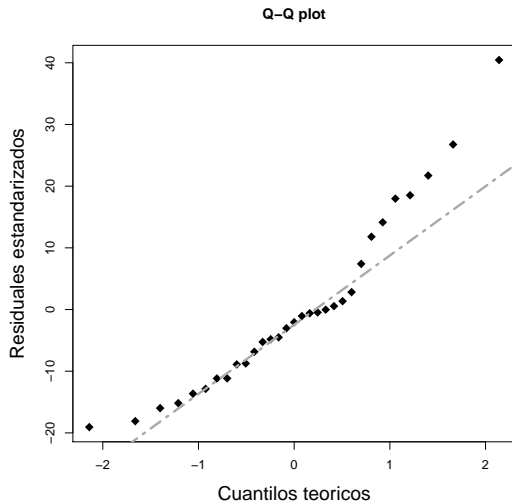
Regresión  
Estimación  
Evaluación del  
ajuste  
Comparación de  
modelos  
**Condiciones**

Otros tipos de  
regresión

Criticas a los  
modelos



# Normalidad de los residuales



- Test de Shapiro-Wilk



# ¿Qué hacer si las condiciones no cumplen?

- Residuales no son independientes:

# ¿Qué hacer si las condiciones no cumplen?

- Residuales no son independientes:
  - Modelos con efectos aleatorios (random effect models)

# ¿Qué hacer si las condiciones no cumplen?

- Residuales no son independientes:
  - Modelos con efectos aleatorios (random effect models)
- Residuales no son normales:

# ¿Qué hacer si las condiciones no cumplen?

- Residuales no son independientes:
  - Modelos con efectos aleatorios (random effect models)
- Residuales no son normales:
  - Alternativa no paramétrica

# ¿Qué hacer si las condiciones no cumplen?

- Residuales no son independientes:
  - Modelos con efectos aleatorios (random effect models)
- Residuales no son normales:
  - Alternativa no paramétrica
  - Transformación de los datos *log*, *sqrt*, *exp* ...

# ¿Qué hacer si las condiciones no cumplen?

- Residuales no son independientes:
  - Modelos con efectos aleatorios (random effect models)
- Residuales no son normales:
  - Alternativa no paramétrica
  - Transformación de los datos *log*, *sqrt*, *exp* ...
  - Modelo lineal generalizado (Generalized Linear Model GLM)

# ¿Qué hacer si las condiciones no cumplen?

- Residuales no son independientes:
  - Modelos con efectos aleatorios (random effect models)
- Residuales no son normales:
  - Alternativa no paramétrica
  - Transformación de los datos *log*, *sqrt*, *exp* ...
  - Modelo lineal generalizado (Generalized Linear Model GLM)
- Heterogeneidad de la varianza:

# ¿Qué hacer si las condiciones no cumplen?

- Residuales no son independientes:
  - Modelos con efectos aleatorios (random effect models)
- Residuales no son normales:
  - Alternativa no paramétrica
  - Transformación de los datos *log*, *sqrt*, *exp* ...
  - Modelo lineal generalizado (Generalized Linear Model GLM)
- Heterogeneidad de la varianza:
  - GLM

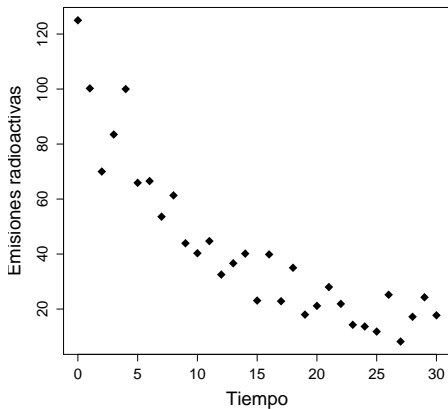


# Si el modelo es inadecuado, se puede. . .

- Transformar variable dependiente
- Transformar  $\geq 1$  variable explicativa
- Probar otras variables explicativas
- Usar una estructura de error diferente (GLM)
- Usar alternativa no paramétrica (smoothing)
- Usar pesos diferentes por diferentes valores de  $y$

# Regresión polinomial

Ejemplo: Desintegración radioactiva

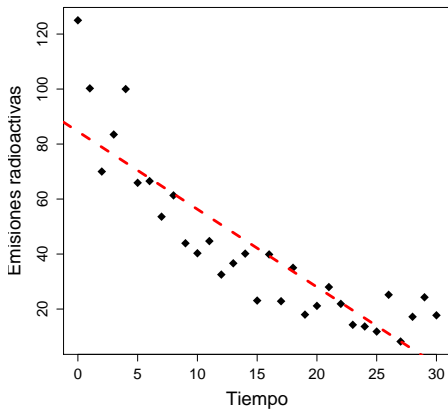


# Regresión polinomial

## Ejemplo: Desintegración radioactiva

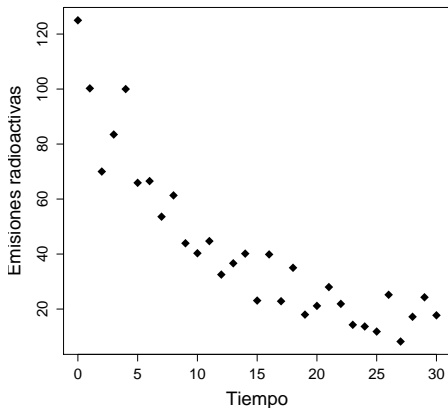
- Regresión lineal:

$$y = ax + b$$



# Regresión polinomial

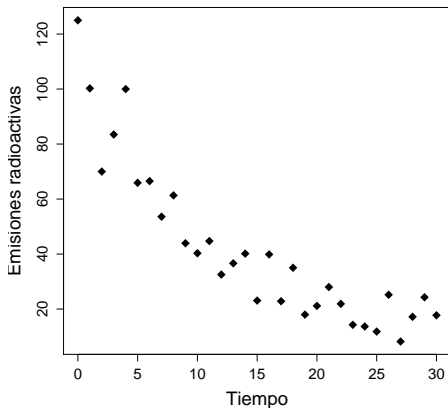
## Ejemplo: Desintegración radioactiva



- Regresión lineal:  
 $y = ax + b$
- Regresión polinómica

# Regresión polinomial

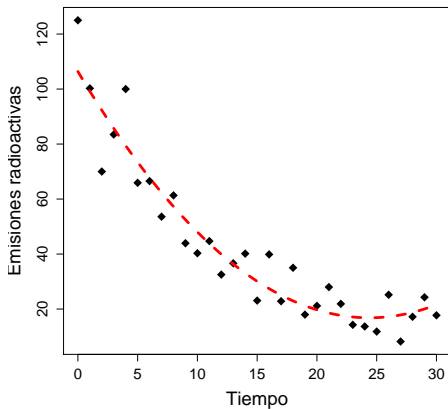
## Ejemplo: Desintegración radioactiva



- Regresión lineal:  
 $y = ax + b$
- Regresión polinómica
- $X_2 = X^2$

# Regresión polinomial

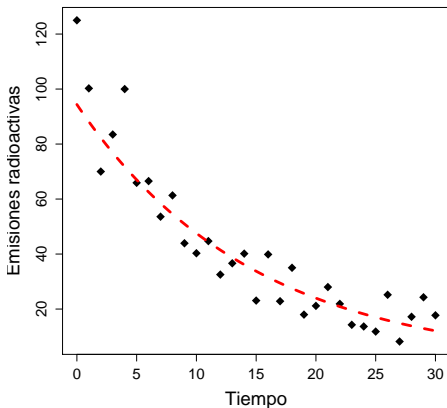
## Ejemplo: Desintegración radioactiva



- Regresión lineal:  
 $y = ax + b$
- Regresión polinómica
- $X_2 = X^2$
- $y = ax^2 + bx + c$

# Regresión polinomial

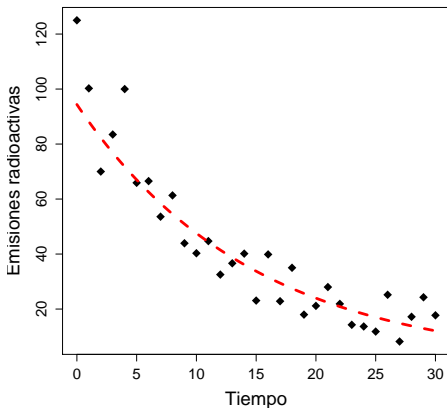
## Ejemplo: Desintegración radioactiva



- Regresión lineal:  
 $y = ax + b$
- Regresión polinómica
- $X_2 = X^2$
- $y = ax^2 + bx + c$
- $y = ae^{-bx}$

# Regresión polinomial

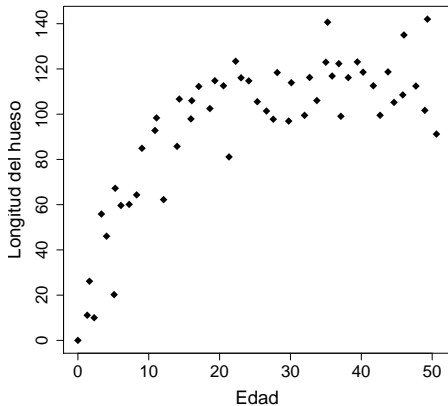
## Ejemplo: Desintegración radioactiva



- Regresión lineal:  
 $y = ax + b$
- Regresión polinómica
- $X_2 = X^2$
- $y = ax^2 + bx + c$
- $y = ae^{-bx}$
- ¡Descripción, no explicación!

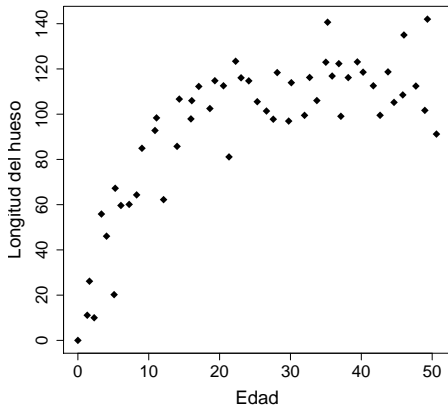


# Regresión no lineal y GAM

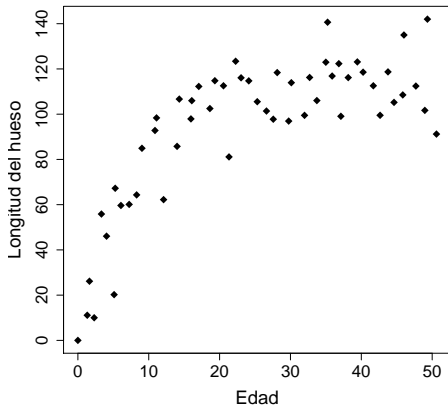



# Regresión no lineal y GAM

• : `nls()`

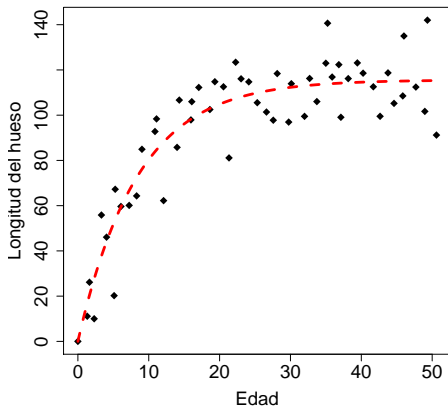



# Regresión no lineal y GAM



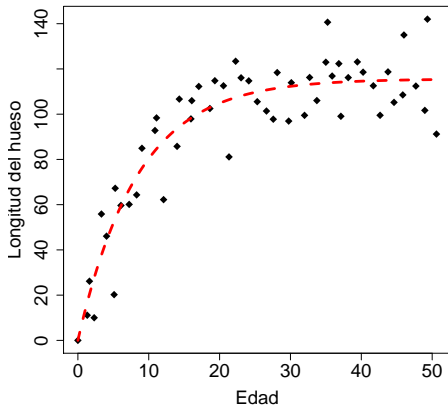
- : `nls()`
- Teoría:  
$$y = a - be^{-cx}$$


# Regresión no lineal y GAM



- : `nls()`
- Teoría:  
$$y = a - be^{-cx}$$

# Regresión no lineal y GAM



- : `nls()`
- Teoría:  
$$y = a - be^{-cx}$$
- No información:  
Modelos Aditivos  
Generalizados  
(Generalized Additive  
Models GAM)

# Recordatorio de vocabulario

- Normalidad de los errores:
  - Modelos lineales
- Normalidad + var. descriptivas continuas/categorías:
  - Modelos lineales generales
- Errores no normales y/o varianza no homogénea:
  - Modelos lineales generalizados (GLM)

# Modelos lineales generalizados (2)

Varianza no constante / residuales no normales

⇒ Se puede especificar la distribución de los errores

- Proporciones (regresión logística) → Binomial
- Conteos (modelo log-lineal) → Poisson
- Variable dependiente binaria (vivo/muerto) → Binomial
- Tiempo hasta muerte (varianza aumenta) → Exponencial

# Modelos lineales generalizados (2)

Varianza no constante / residuales no normales

⇒ Se puede especificar la distribución de los errores

- Proporciones (regresión logística) → Binomial
- Conteos (modelo log-lineal) → Poisson
- Variable dependiente binaria (vivo/muerto) → Binomial
- Tiempo hasta muerte (varianza aumenta) → Exponencial



## (No) enamorarse de su modelo . . .

- Todos los modelos son incorrectos
- Algunos modelos son mejores que otros
- El modelo correcto nunca se puede conocer con certeza
- Cuanto mas simple el modelo mejor