

Analisis de regresión con R

CNE/ISCIII

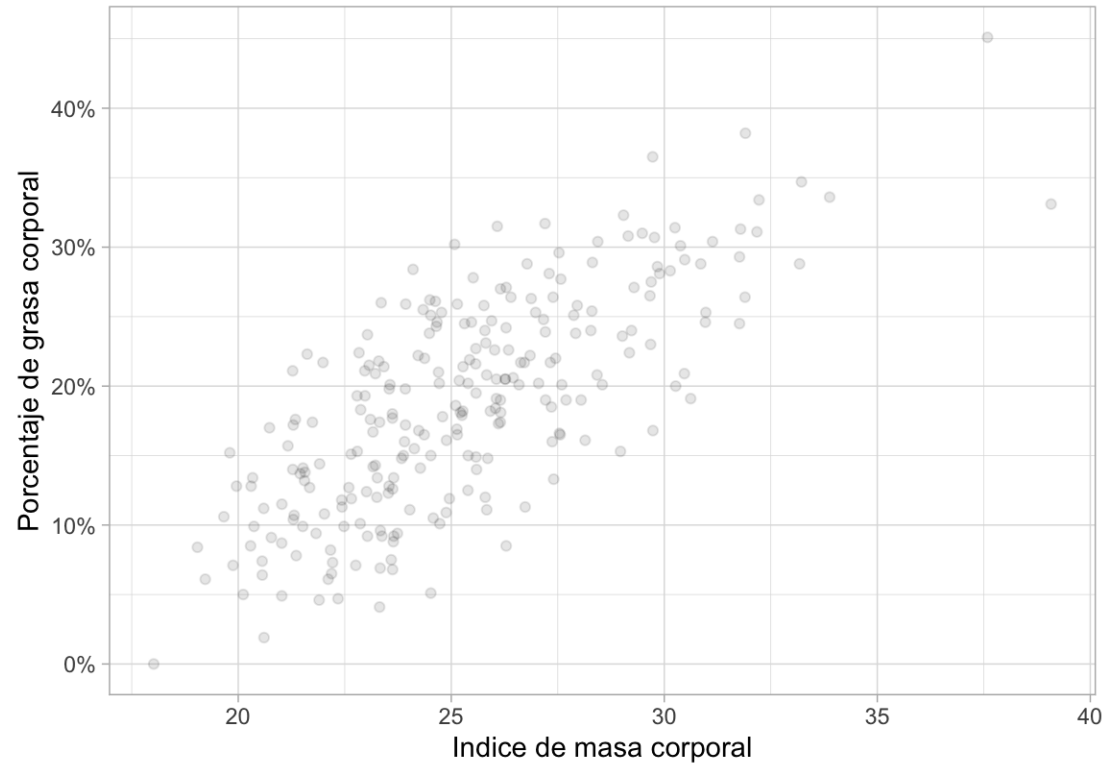
Estructura del curso

1. Correlación (`cor`)
2. Regresión lineal simple (`lm`)
3. Regresión múltiple
4. Regresión logística (`glm`)

Correlación

Asociación entre dos variables

IMC versus porcentaje de grasa corporal



Motivos para medir la asociación

En estudios epidemiológicos

Objetivos del estudio:

- Etiología (determinar factores de riesgos involucrados en una enfermedad)
- Intervención (impacto de una vacuna en la evolución de una epidemia)
- Predicción (diseñar pruebas diagnosticas para una enfermedad)

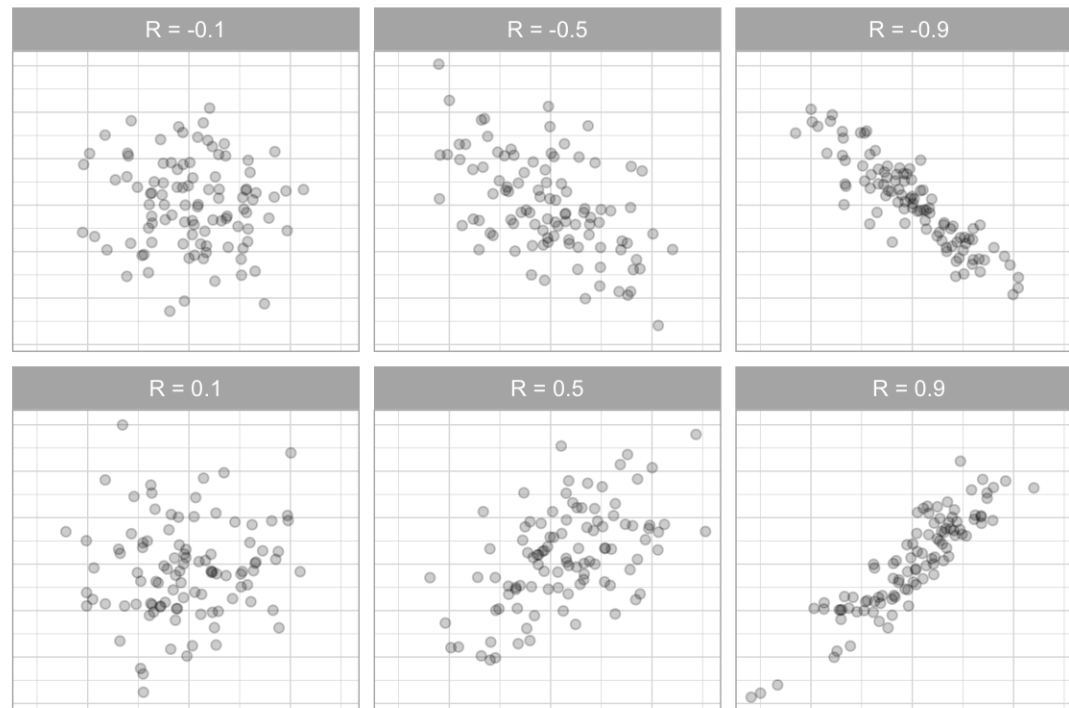
Diseño del estudio:

- Sesgo de confusión (asociación del café con el riesgo de cáncer de pulmón)
- Sesgo de selección (selección de los controles asociada con la exposición)
- Sesgo de medición (demencia asociada con el recuerdo de la exposición)

Coeficiente de correlación de Pearson (R)

Medida (simétrica e invariante por cambio de escala) de asociación lineal entre dos variables X e Y :

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad -1 \leq R \leq 1$$



Un ejemplo sencillo

```
starwars=read.csv("data/starwars.csv")
str(starwars)
```

```
## 'data.frame':    87 obs. of  3 variables:
##  $ name   : chr  "Luke Skywalker" "C-3PO" "R2-D2" "Darth Vader" ...
##  $ height: int  172 167 96 202 150 178 165 97 183 182 ...
##  $ mass   : num  77 75 32 136 49 120 75 32 84 77 ...
```

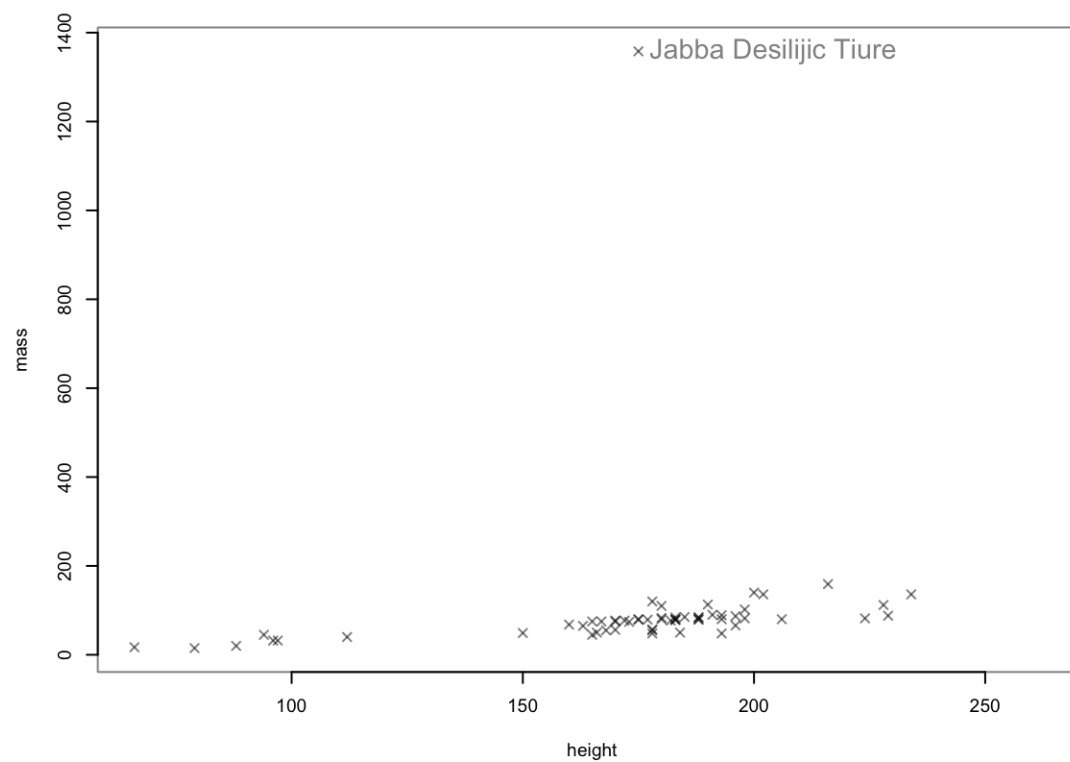
```
# Calculo de la correlación lineal entre peso y altura
cor(starwars$mass, starwars$height)
```

```
## [1] NA
```

```
cor(starwars$mass, starwars$height, use="pairwise.complete.obs")
```

```
## [1] 0.1338842
```

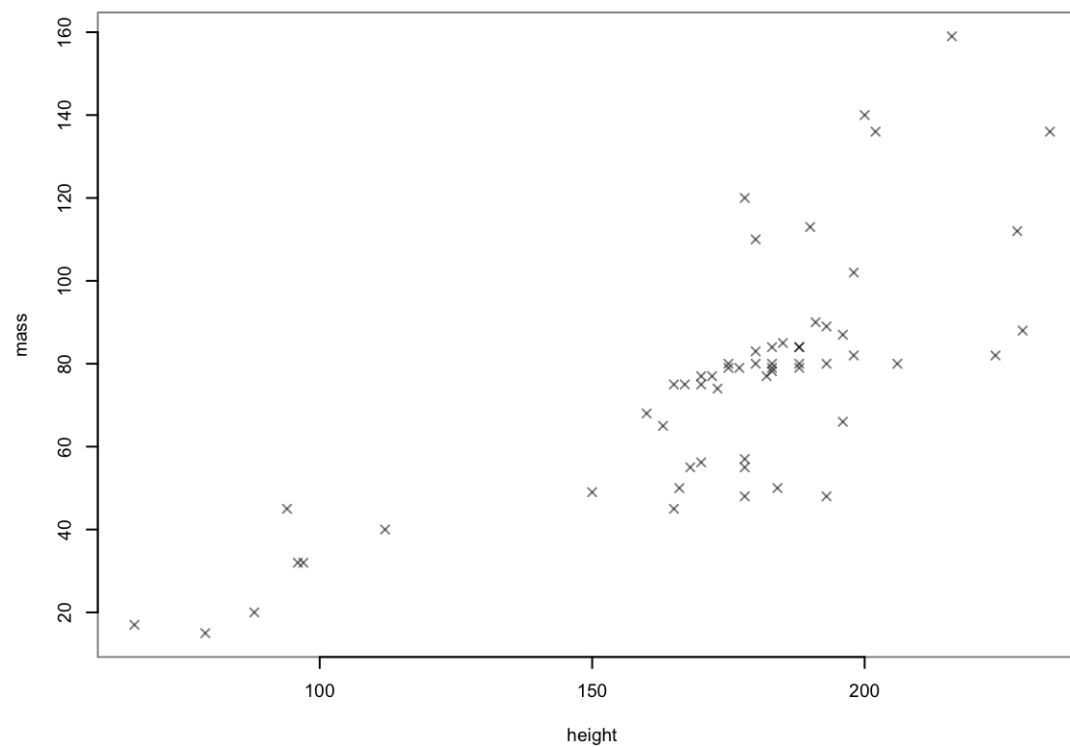
```
plot(mass~height, data=starwars)
outlier=subset(starwars, mass>1000)
text(outlier$height, outlier$mass, outlier$name, pos=4, cex=1.5)
```




```
sin.jabba=subset(starwars,mass<1000)  
cor(sin.jabba$mass, sin.jabba$height, use="pairwise.complete.obs")
```

```
## [1] 0.7612612
```

```
plot(mass~height, data=sin.jabba)
```



Asociaciones en una base de datos

```
neonatos <- read.csv("data/neonatos.csv")
```

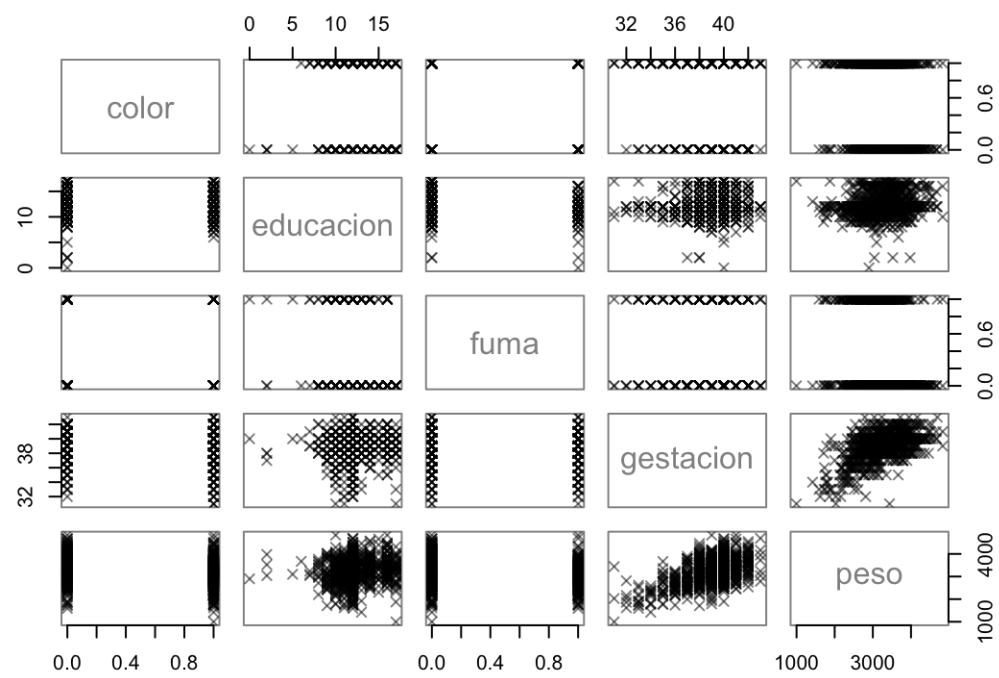
```
str(neonatos) #peso de neonatos según tiempo gestación, color, educación y habito tabaquico de la madre
```

```
## 'data.frame':    1092 obs. of  5 variables:
## $ color      : logi  FALSE FALSE FALSE FALSE FALSE TRUE ...
## $ educacion: int   0 2 2 2 5 6 7 7 7 8 ...
## $ fuma       : logi   TRUE FALSE TRUE FALSE TRUE FALSE ...
## $ gestacion: int  40 38 37 38 40 40 41 39 39 38 ...
## $ peso       : int  2898 3977 3040 3523 3100 3670 3097 3040 3239 2955 ...
```

```
cor(neonatos)
```

```
##           color  educacion      fuma  gestacion      peso
## color      1.00000000 -0.14504912  0.04957376 -0.17463028 -0.2566835
## educacion -0.14504912  1.00000000 -0.22140608  0.06556138  0.1249150
## fuma       0.04957376 -0.22140608  1.00000000 -0.12360134 -0.2202095
## gestacion -0.17463028  0.06556138 -0.12360134  1.00000000  0.5385007
## peso      -0.25668349  0.12491498 -0.22020949  0.53850072  1.0000000
```

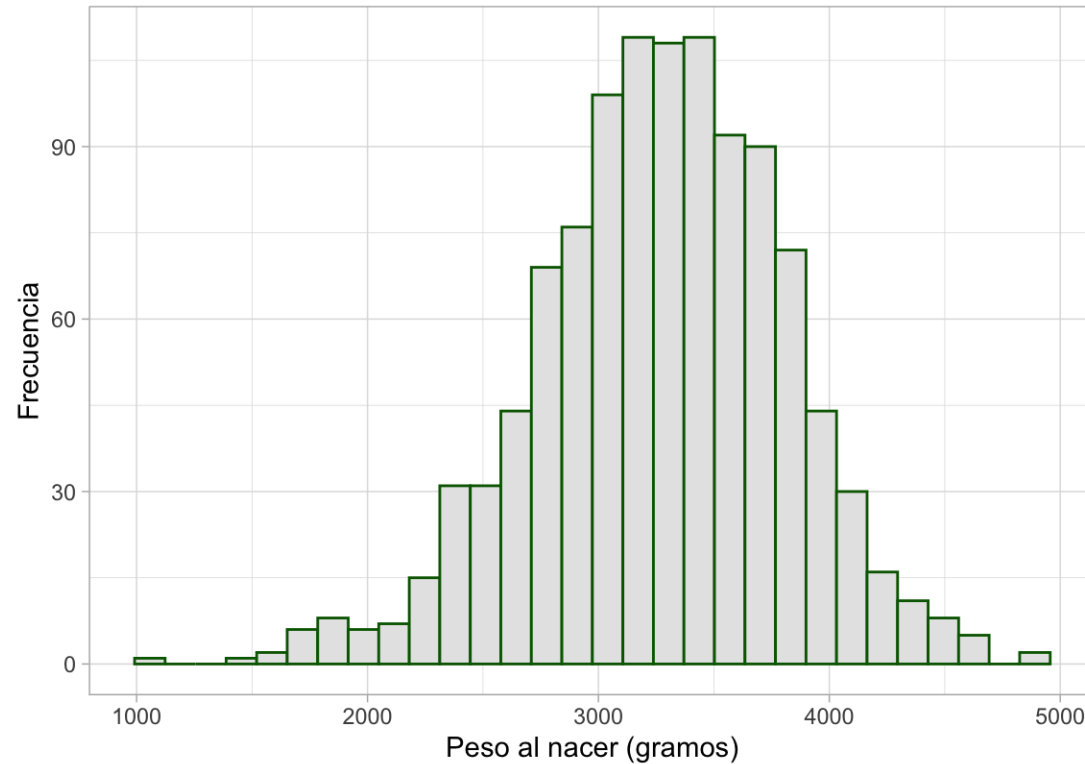
`pairs(neonatos)`



Regresión lineal simple

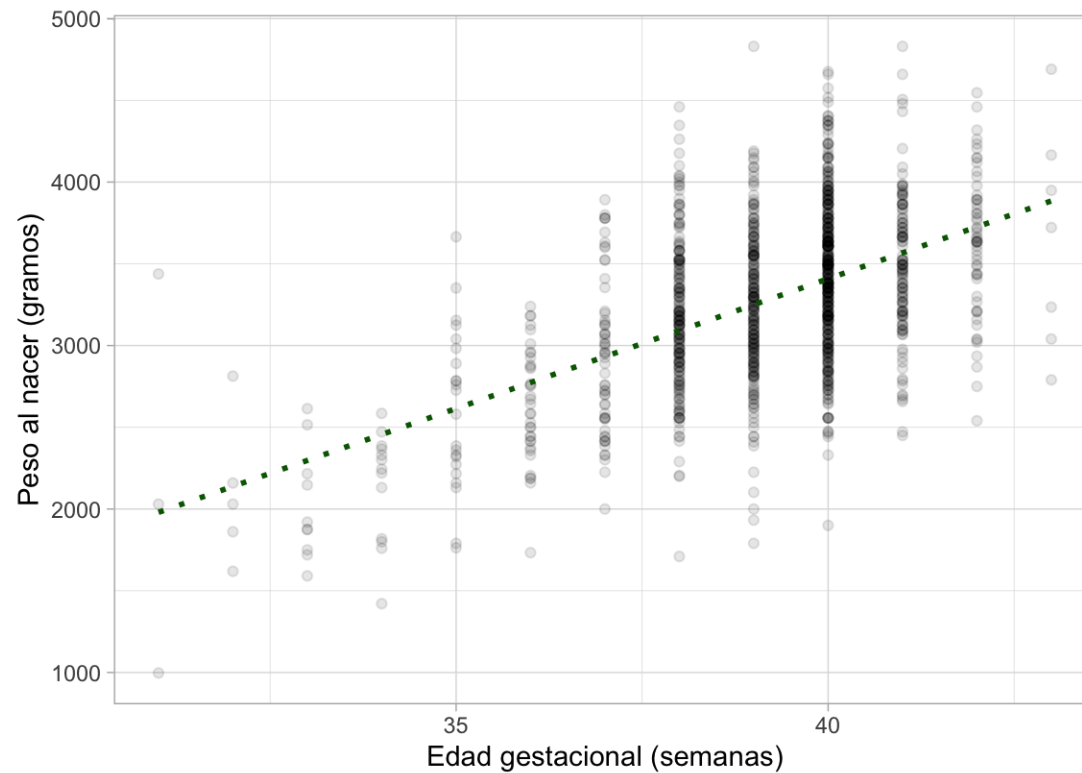
Controlar las variaciones

¿Porqué no nacemos todos con el mismo peso?



Controlar las variaciones

Efecto de la edad gestacional sobre el peso al nacer



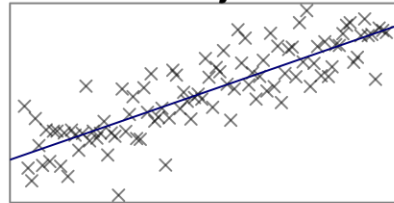
El modelo

Asumiendo un efecto constante

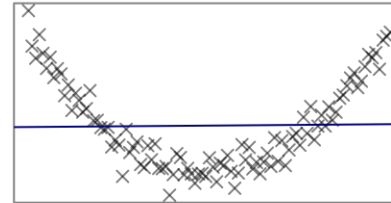
$$Y = \beta_0 + \beta_1 X + \varepsilon$$

donde el residuo ε tiene una **distribución normal** de **media cero** y desviación típica σ **constante**.

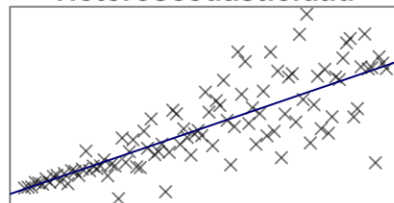
Buen ajuste



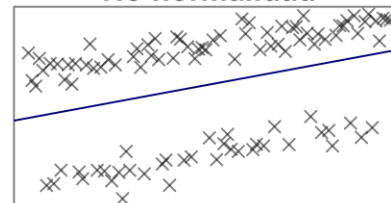
No linealidad



Heteroscedasticidad



No normalidad



Interpretación

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- β_0 : valor promedio de Y cuando $X = 0$.
- β_1 : cambio promedio en Y cuando X aumenta de una unidad.
- σ : variabilidad de Y no explicada por X .

Ejemplo: Efecto de la gestación sobre el peso al nacer

$$\text{Peso(gramos)} = \beta_0 + \beta_1 \text{Gestación(semanas)} + \varepsilon$$

- β_0 : no se puede aquí interpretar!
- β_1 : variación promedia del peso al aumentar la gestación de una semana.
- σ : variabilidad del peso entre neonatos con la misma edad gestacional.

función `lm`

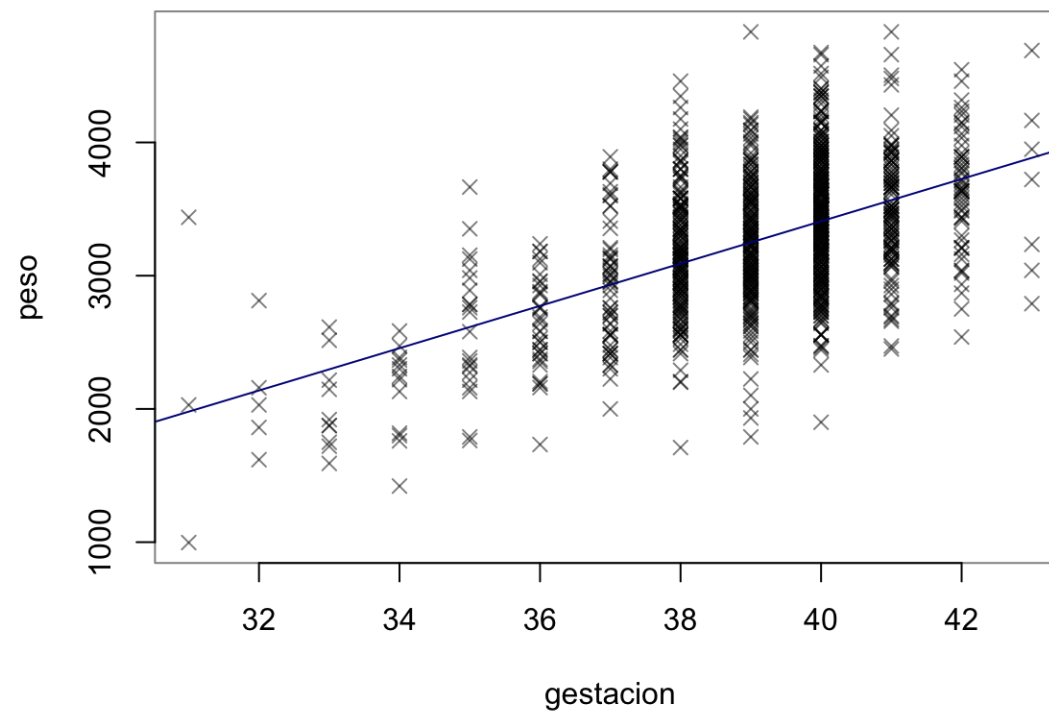
```
simple = lm(peso~gestacion, data=neonatos) # lm: linear model
summary(simple)
```

```
##
## Call:
## lm(formula = peso ~ gestacion, data = neonatos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1508.40  -302.41   -12.64    296.86   1580.35
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2941.381     294.642  -9.983  <2e-16 ***
## gestacion    158.744       7.524   21.099  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 454.1 on 1090 degrees of freedom
## Multiple R-squared:  0.29, Adjusted R-squared:  0.2893
## F-statistic: 445.2 on 1 and 1090 DF, p-value: < 2.2e-16
```

```
coef(simple) #ordenada en el origen y pendiente
```

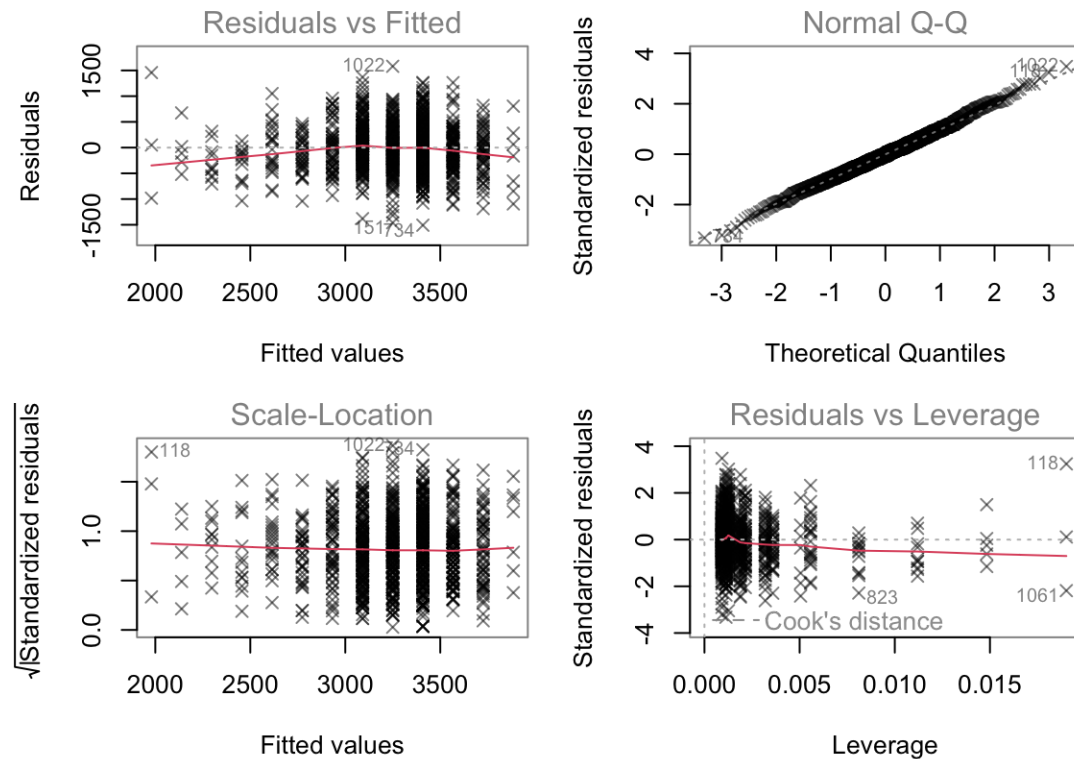
```
## (Intercept)  gestacion  
## -2941.3815    158.7445
```

```
plot(peso~gestacion, data=neonatos)  
abline(reg=simple,col="blue4") #recta de regresión
```



Chequeo de las asunciones

```
par(mfrow=c(2,2))  
plot(simple)
```



Regresión lineal simple versus test de Student

```
tabaco = lm(peso~fuma, data=neonatos)
summary(tabaco)$coef
```

```
##              Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) 3334.4928   18.20218 183.191961 0.000000e+00
## fumaTRUE    -279.1052   37.44767  -7.453206 1.851559e-13
```

```
t.test(peso~fuma, data=neonatos, var.equal=TRUE) # se asume misma varianza en los dos grupos
```

```
##
## Two Sample t-test
##
## data:  peso by fuma
## t = 7.4532, df = 1090, p-value = 1.852e-13
## alternative hypothesis: true difference in means between group FALSE and group TRUE is not equal to 0
## 95 percent confidence interval:
##  205.6275 352.5829
## sample estimates:
## mean in group FALSE  mean in group TRUE
##           3334.493           3055.388
```

Regresión múltiple

Regresión con más de un predictor

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

- β_0 : valor promedio de Y cuando $X_1 = X_2 = \dots = X_k = 0$.
- β_j : cambio promedio en Y cuando X_j aumenta de una unidad... y los demás predictores permanecen constantes.
- σ : variabilidad de Y no explicada por los predictores del modelo.

Ejemplo:

$$\text{Peso} = \beta_0 + \beta_1 \text{Gestación} + \beta_2 \text{Raza} + \varepsilon$$

- β_1 : variación del peso al aumentar la gestación de una semana entre neonatos de una misma raza.
- β_2 : diferencia de peso de neonatos de diferente raza y misma edad gestacional.
- σ : variabilidad del peso entre neonatos de misma raza y misma edad gestacional.

Ajuste

Efecto crudo versus efecto ajustado

```
crudo = lm(peso~color, data=neonatos)
summary(crudo)$coef
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3433.6178	24.55239	139.848604	0.000000e+00
colorTRUE	-280.7689	32.02120	-8.768218	6.856359e-18

```
ajustado = lm(peso~gestacion+color, data=neonatos)
summary(ajustado)$coef
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2495.6582	296.848595	-8.407176	1.301190e-16
gestacion	150.1083	7.496326	20.024244	3.417816e-76
colorTRUE	-183.5025	27.815521	-6.597125	6.528576e-11

¡Interpretar!

Intervalos de confianza

función confint

```
summary(ajustado)$coef
```

```
##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) -2495.6582  296.848595 -8.407176 1.301190e-16
## gestacion    150.1083    7.496326 20.024244 3.417816e-76
## colorTRUE    -183.5025   27.815521 -6.597125 6.528576e-11
```

```
confint(ajustado) #cálculo del intervalo de confianza de los coeficientes
```

```
##              2.5 %      97.5 %
## (Intercept) -3078.1182 -1913.1983
## gestacion    135.3994   164.8171
## colorTRUE    -238.0805  -128.9244
```


Predicción

función predict

```
prediccion = predict(ajustado) #por defecto, predicción de Y para valores de X observados
head(cbind(neonatos,prediccion))
```

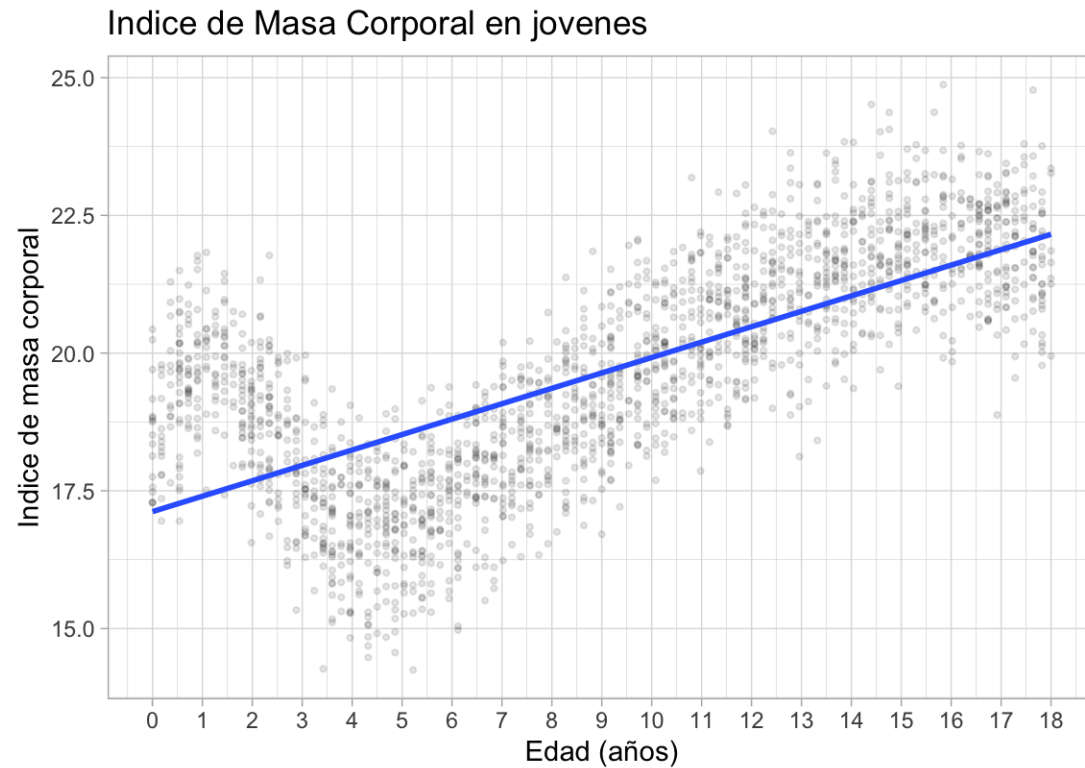
```
##   color educacion  fuma gestacion peso prediccion
## 1 FALSE         0  TRUE         40 2898   3508.672
## 2 FALSE         2 FALSE         38 3977   3208.455
## 3 FALSE         2  TRUE         37 3040   3058.347
## 4 FALSE         2 FALSE         38 3523   3208.455
## 5 FALSE         5  TRUE         40 3100   3508.672
## 6  TRUE         6 FALSE         40 3670   3325.169
```

```
# Predicción fuera de la muestra: peso medio de un neonato negro con 30 semanas de edad gestacional
nuevo = data.frame(gestacion=30, color= TRUE)
predict(ajustado, newdata = nuevo , interval="confidence")
```

```
##           fit          lwr          upr
## 1 1824.087 1689.391 1958.783
```

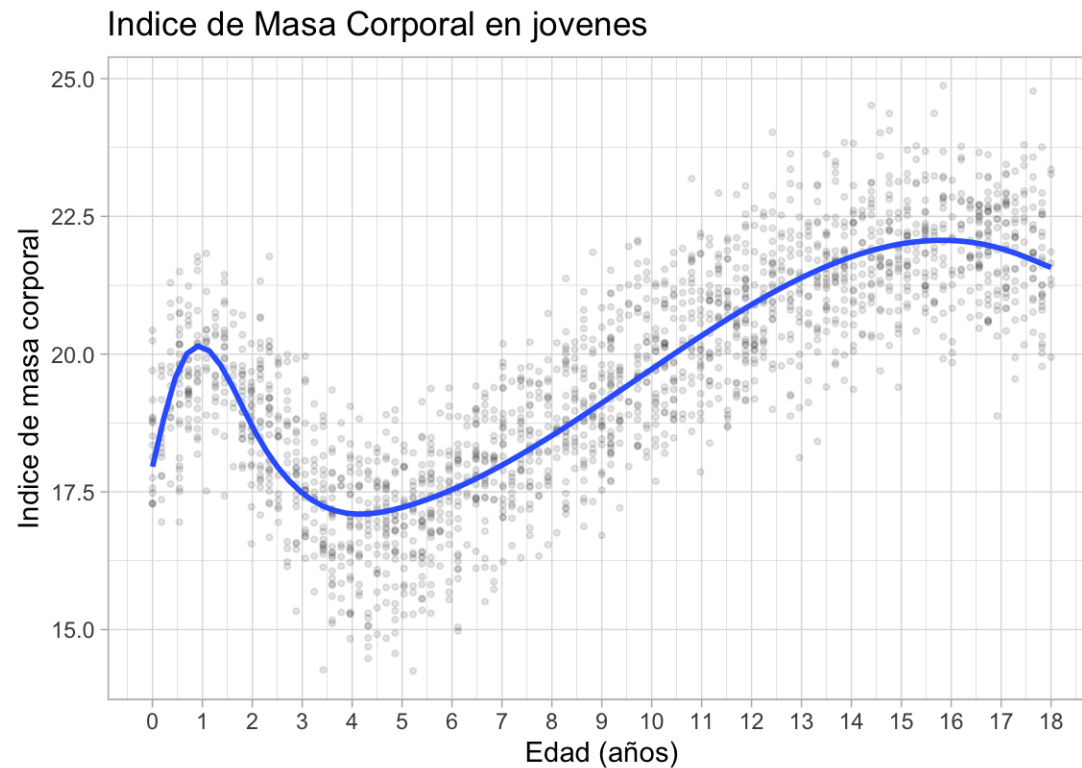
No linealidad

El modelo de regresión lineal simple no se ajusta bien a los datos si el efecto de X no es constante.



El modelo de regresión múltiple es muy flexible y permite estimar efectos no lineales.

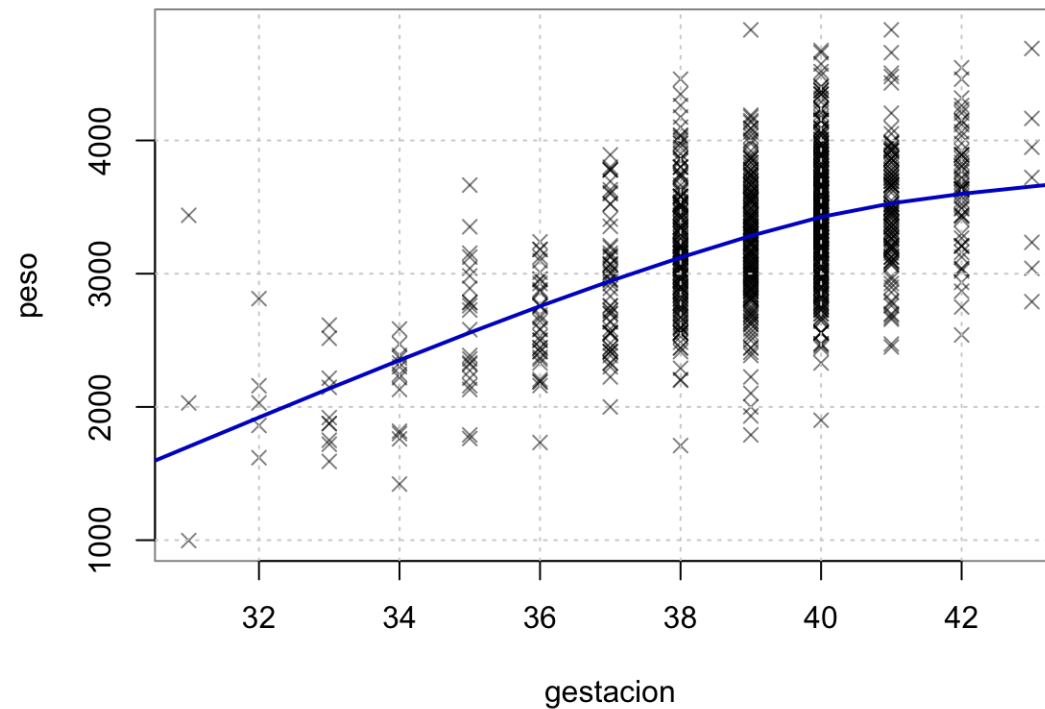
$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_k X^k + \varepsilon \quad (\text{regresión polinomial / splines})$$
$$= \beta_0 + \beta_1 (\text{X medio}) + \beta_2 (\text{X alto}) + \varepsilon \quad (\text{regresión politomica})$$



Regresión spline

```
# Estimación del efecto de la gestación
fit.spline = lm(peso~ns(gestacion,3), data=neonatos) # require(splines)
rango=data.frame(gestacion=30:45)
pred.spline= predict(fit.spline,rango)

plot(peso~gestacion, data=neonatos); grid()
lines(rango$gestacion, pred.spline, col="blue3", lwd=2)
```



Regresión politomica

```
categoriza<-function(x) cut(x, c(0,10,13,20),  
                           labels=c(".primaria", ".segundaria", ".superior"),  
                           include.lowest=TRUE)  
neonatos$educacion_cat=categoriza(neonatos$educacion)  
table(neonatos$educacion_cat)
```

```
##  
##      .primaria .segundaria  .superior  
##           166           689           237
```

```
summary( fit.politomica <- lm(peso~educacion_cat, data=neonatos))$coef # Estimación del efecto de la educación
```

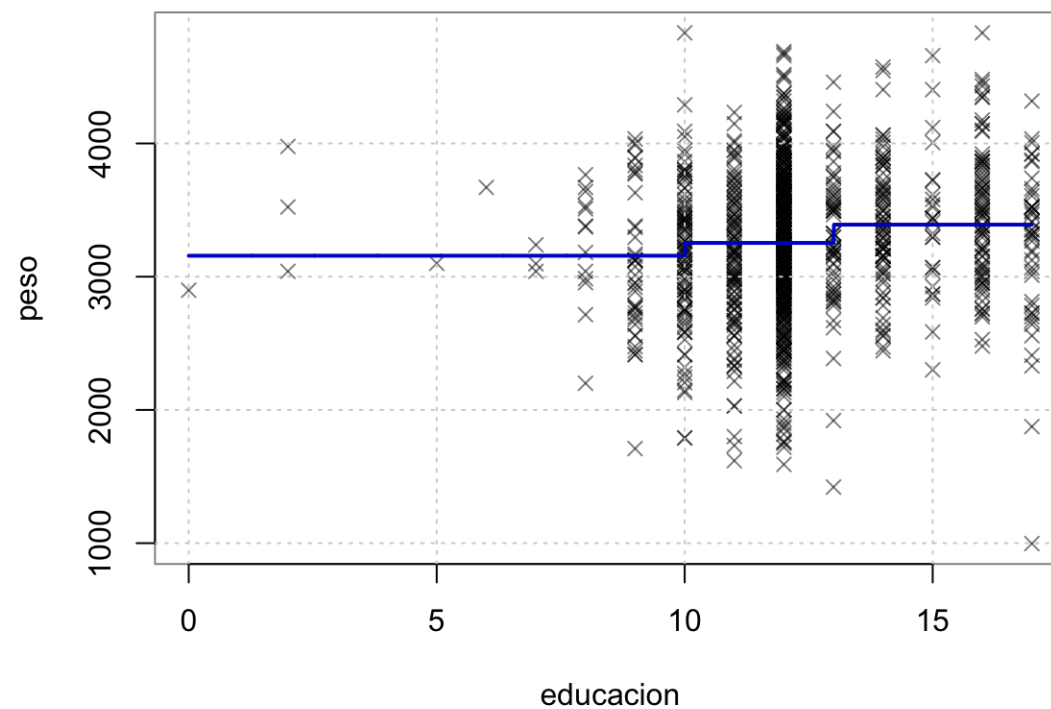
```
##              Estimate Std. Error  t value    Pr(>|t|)  
## (Intercept)    3157.32530    41.46437  76.145497 0.000000e+00  
## educacion_cat.segundaria    96.08108    46.19006   2.080125 3.774751e-02  
## educacion_cat.superior    233.15571    54.06964   4.312138 1.763431e-05
```

```

rango=data.frame(educacion=seq(0,17,.01))
rango$educacion_cat = categoriza(rango$educacion)
pred.politomica= predict(fit.politomica,rango)

plot(peso~educacion, data=neonatos); grid()
lines(rango$educacion,pred.politomica,col="blue3", lwd=2)

```



Selección de modelo

Descomposición de la variabilidad y Test de ajuste

$$\underbrace{\sum_i (y_i - \bar{y})^2}_{\text{Variabilidad Total (VT)}} = \underbrace{\sum_i (\hat{y}_i - \bar{y})^2}_{\text{Variabilidad Explicada (VE)}} + \underbrace{\sum_i (\hat{y}_i - y_i)^2}_{\text{Variabilidad No Explicada (VNE)}}$$

```
anova(simple)
```

```
## Analysis of Variance Table
##
## Response: peso
##           Df      Sum Sq Mean Sq F value    Pr(>F)
## gestacion   1  91790877 91790877  445.17 < 2.2e-16 ***
## Residuals 1090 224747921   206191
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(simple)$r.squared # R^2 = VE / VT proporción de variabilidad explicada por el modelo
```

```
## [1] 0.289983
```

```
anova(gestacion.tabaco <- lm(peso ~ gestacion + fuma, data = neonatos))
```

```
## Analysis of Variance Table
##
## Response: peso
##           Df      Sum Sq Mean Sq F value    Pr(>F)
## gestacion   1  91790877 91790877 460.309 < 2.2e-16 ***
## fuma         1   7588895  7588895  38.056 9.674e-10 ***
## Residuals 1089 217159026   199411
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(simple,gestacion.tabaco)
```

```
## Analysis of Variance Table
##
## Model 1: peso ~ gestacion
## Model 2: peso ~ gestacion + fuma
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1    1090 224747921
## 2    1089 217159026  1   7588895 38.056 9.674e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Evitando el sobre ajuste

Criterio de información de Akaike

AIC: Variabilidad residual penalizada por el número de predictores

$$\text{AIC} = n \log(\text{VNE}) + 2k + \text{Constante}$$

donde n es el tamaño muestral y k el número de predictores ($n > k + 1$).

```
for(k in 1:4) {  
  fit=lm(peso~ ns(gestacion,k), data=neonatos)  
  cat("AIC para spline con",k,"grados de libertad:",AIC(fit),"\n")  
}
```

```
## AIC para spline con 1 grados de libertad: 16465.28  
## AIC para spline con 2 grados de libertad: 16453.41  
## AIC para spline con 3 grados de libertad: 16454.91  
## AIC para spline con 4 grados de libertad: 16456.7
```

El modelo que minimiza el AIC es el más parsimonioso: optimiza compromiso entre bondad del ajuste y estabilidad del modelo.

Selección automática (¡sólo para fines predictivos!)

```
summary( todo <- lm(peso~., data=neonatos) ) # modelo completo
```

```
##
## Call:
## lm(formula = peso ~ ., data = neonatos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1452.35  -293.52   -0.63   287.07  1568.08
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2220.957     332.187   -6.686 3.67e-11 ***
## colorTRUE       -172.630      27.669   -6.239 6.30e-10 ***
## educacion        -5.257      15.146   -0.347  0.729
## fumaTRUE       -182.257      32.224   -5.656 1.98e-08 ***
## gestacion       144.300       7.435   19.408 < 2e-16 ***
## educacion_cat.segundaria  46.676      54.917    0.850  0.396
## educacion_cat.superior  111.949     102.127    1.096  0.273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 438.2 on 1085 degrees of freedom
## Multiple R-squared:  0.3419, Adjusted R-squared:  0.3383
## F-statistic: 93.95 on 6 and 1085 DF,  p-value: < 2.2e-16
```

Selección automática (función step)

```
mejor=step(todo, trace=0) # "mejor" modelo con fines de predicción
summary(mejor)
```

```
##
## Call:
## lm(formula = peso ~ color + educacion + fuma + gestacion, data = neonatos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1456.38  -296.42    1.04   294.37  1570.48
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2365.604    305.399  -7.746 2.17e-14 ***
## colorTRUE   -173.242     27.606  -6.276 5.02e-10 ***
## educacion      9.851      6.599   1.493  0.136
## fumaTRUE     -181.613     32.201  -5.640 2.17e-08 ***
## gestacion    144.635      7.423  19.486 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 438 on 1087 degrees of freedom
## Multiple R-squared:  0.3412, Adjusted R-squared:  0.3387
## F-statistic: 140.7 on 4 and 1087 DF,  p-value: < 2.2e-16
```

Modificación del efecto

Interacción (*)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 \times X_2 = \begin{cases} \beta_0 + \beta_1 X_1 & \text{si } X_2 = 0 \\ (\beta_0 + \beta_2) + (\beta_1 + \beta_{12}) X_1 & \text{si } X_2 = 1 \end{cases}$$

```
summary( lm(peso ~ color*educacion, data=neonatos) )$coef
```

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	3299.52876	131.49927	25.091612	5.380434e-110
## colorTRUE	-614.78502	191.37962	-3.212385	1.354991e-03
## educacion	10.61017	10.22406	1.037765	2.996101e-01
## colorTRUE:educacion	28.31774	15.34100	1.845886	6.518022e-02

Interpretación: el peso del neonato “aumenta” en promedio de 10.6g por cada año de estudio de la madre si es blanca y de $10.6 + 28.3 \simeq 39$ g si la madre es negra.

Selección automática de interacciones

```
interacciones=lm(peso~(.-educacion_cat)^2, data=neonatos) #todos los modelos con interacción  
summary(step(interacciones, trace=0))$coef #mejor modelo
```

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	2604.381140	1982.716064	1.3135422	1.892785e-01
## colorTRUE	-740.634023	169.735199	-4.3634675	1.402506e-05
## educacion	-374.277061	157.095582	-2.3824799	1.736825e-02
## fumaTRUE	-1461.741756	653.059835	-2.2382968	2.540430e-02
## gestacion	25.064585	50.427382	0.4970432	6.192595e-01
## colorTRUE:educacion	41.505578	13.304905	3.1195698	1.859015e-03
## colorTRUE:fumaTRUE	264.895025	67.404243	3.9299459	9.035207e-05
## educacion:gestacion	9.292652	3.994529	2.3263451	2.018395e-02
## fumaTRUE:gestacion	28.561911	16.509816	1.7299957	8.391594e-02

Regresión logística

Odds

Medida de ocurrencia de un evento o enfermedad

$$\text{Odds} = \frac{P(Y = 1)}{P(Y = 0)}, \quad P(Y = 1) = \frac{\text{Odds}}{1 + \text{Odds}}$$

```
neonatos40=subset(neonatos,gestacion == 40) # neonatos con una edad de gestación de 40 semanas
neonatos40$peg = (neonatos40$peso <= 2800) # niños pequeños para su edad gestacional (PEG)
table(neonatos40$peg)
```

```
##
## FALSE TRUE
##    345    29
```

$$\text{Odds} = \frac{P(\text{PEG} = 1)}{P(\text{PEG} = 0)} = \frac{29}{345}$$

Odds ratio (OR)

Medida de asociación para variables dicotómicas

Definición: Ratio entre el odds de enfermedad dada una exposición particular y el odds de dicha enfermedad en ausencia de esa exposición.

```
xtabs(~ fuma + peg, data=neonatos40)
```

```
##           peg
## fuma      FALSE TRUE
##  FALSE    276   17
##   TRUE     69   12
```

Efecto del hábito tabáquico de la madre sobre el riesgo de PEG:

$$OR = \frac{\text{Odds de PEG en fumadoras}}{\text{Odds de PEG en no fumadoras}} = \frac{12/69}{17/276} = 2.82$$

Interpretación: el riesgo de PEG en madres fumadoras es casi 3 veces mayor que en mujeres no fumadoras!

Modelo logístico

Controlando las variaciones del odds

$$\log(\text{Odds}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

- $\exp(\beta_0)$: valor promedio del odds cuando $X = 0$
- $\exp(\beta_j)$: cambio relativo en el odds cuando X_j aumenta de una unidad y las demás variables permanecen constantes

Ejemplo: Odds de PEG en función del hábito tabáquico de la madre:

$$\log(\text{Odds}) = \beta_0 + \beta_1 \text{fumadora}$$

- $\exp(\beta_0)$: Odds de PEG entre madres que no fuman
- $\exp(\beta_1)$: efecto (OR) del hábito tabáquico de la madre sobre el odds de PEG

función glm

```
(12/69)/(17/276) #Cálculo a mano del OR
```

```
## [1] 2.823529
```

```
or = glm(peg~fuma, data=neonatos40, family=binomial) # glm: generalized linear model
exp(cbind(coef(or), confint(or)))
```

```
##                2.5 %      97.5 %
## (Intercept) 0.0615942 0.03631059 0.09730624
## fumaTRUE    2.8235294 1.26184096 6.15040233
```

```
ors = glm(peg~fuma*color, data=neonatos40, family=binomial)
exp(cbind(coef(ors), confint(ors)))
```

```
##                2.5 %      97.5 %
## (Intercept)    0.04225353 0.01658271 0.08747496
## fumaTRUE        6.10752611 1.98784810 19.77160353
## colorTRUE       1.94278582 0.71832538 5.77573672
## fumaTRUE:colorTRUE 0.20995353 0.03725610 1.04443963
```

Análisis con datos agregados

Supervivencia en la tragedia del Titanic

```
require(data.table)
temp=data.table(Titanic)
str(temp)
```

```
## Classes 'data.table' and 'data.frame':  32 obs. of  5 variables:
##  $ Class      : chr  "1st" "2nd" "3rd" "Crew" ...
##  $ Sex        : chr  "Male" "Male" "Male" "Male" ...
##  $ Age        : chr  "Child" "Child" "Child" "Child" ...
##  $ Survived: chr  "No" "No" "No" "No" ...
##  $ N          : num  0 0 35 0 0 0 17 0 118 154 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```
titanic=dcast(temp, Class+Sex+Age~Survived, value.var="N")
head(titanic)
```

```
##      Class    Sex   Age  No Yes
## 1:    1st Female Adult    4 140
## 2:    1st Female Child    0   1
## 3:    1st   Male Adult 118  57
## 4:    1st   Male Child    0   5
## 5:    2nd Female Adult   13  80
## 6:    2nd Female Child    0  13
```

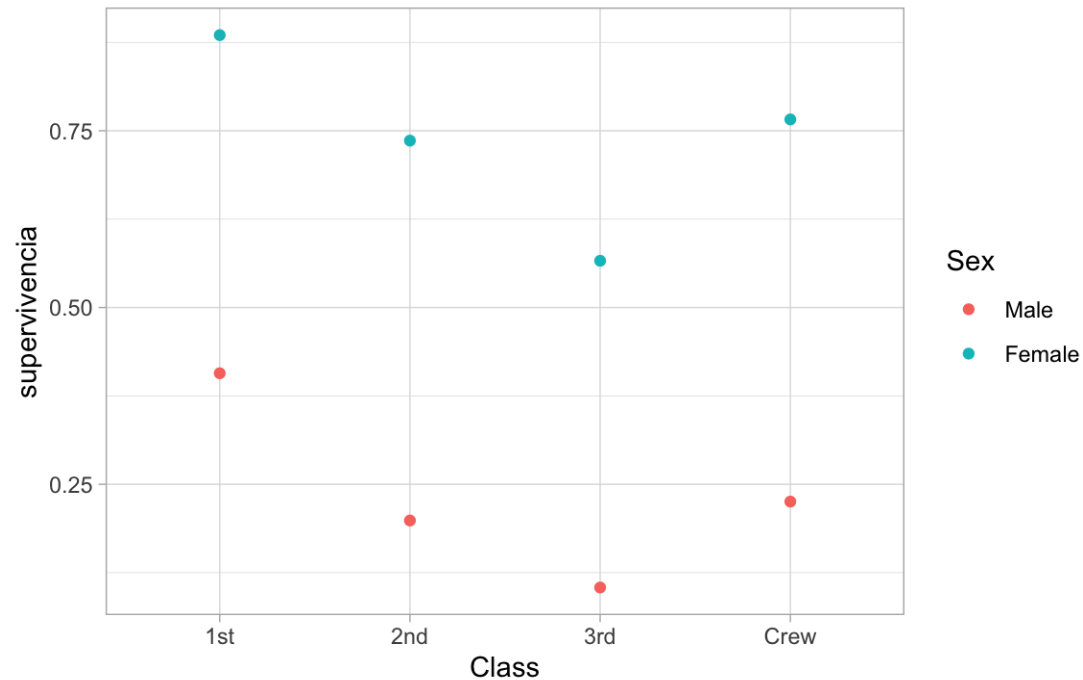
```
supervivencia = glm(cbind(Yes,No)~., data=titanic, family=binomial)
exp(cbind(coef(supervivencia), confint(supervivencia)))
```

```
##                                2.5 %      97.5 %
## (Intercept) 7.72017801 5.58826444 10.7961800
## Class2nd    0.36128255 0.24530635  0.5291859
## Class3rd    0.16901595 0.12032127  0.2358425
## ClassCrew   0.42414659 0.31152317  0.5774746
## SexMale     0.08891625 0.06724914  0.1166490
## AgeChild    2.89082630 1.79245896  4.6710188
```

Predicción del evento

Supervivencia en la tragedia del Titanic

```
nuevos=expand.grid(Class=c("1st","2nd","3rd","Crew"),  
                    Sex=c("Male","Female"), Age="Adult") #sólo adultos  
nuevos$proba=predict(supervivencia, nuevos, type="response")  
ggplot2::qplot(Class, proba, data=nuevos, color=Sex) #más facil con ggplot2
```



Publicación de resultados

Tablas de resultados de la regresión

Paquete `gtsummary`

```
gtsummary::tbl_regression(supervivencia, exp=TRUE)
```

Characteristic	OR ¹	95% CI ¹	p-value
Class			
1st	—	—	
2nd	0.36	0.25, 0.53	<0.001
3rd	0.17	0.12, 0.24	<0.001
Crew	0.42	0.31, 0.58	<0.001
Sex			
Female	—	—	
Male	0.09	0.07, 0.12	<0.001
Age			
Adult	—	—	
Child	2.89	1.79, 4.67	<0.001
¹ OR = Odds Ratio, CI = Confidence Interval			