Clase 4.0
Análisis

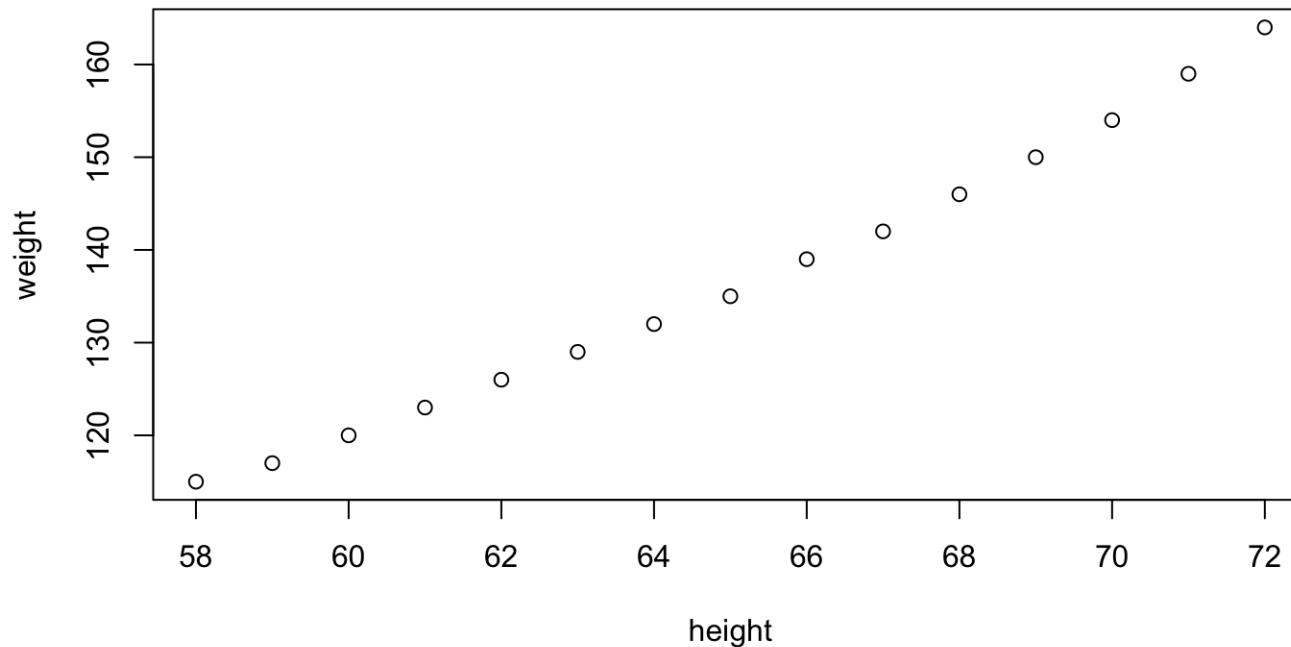Marcos Rosetti y Luis Pacheco-Cobos

Estadística y Manejo de Datos con R (EMDR) — Virtual

Modelos

Modelos

- Regresión sencilla, una relación de una variable predictora y una variable de respuesta.

```
plot(women)
```



```
lm(weight ~ height, data = women)
```

```
##
## Call:
## lm(formula = weight ~ height, data = women)
```
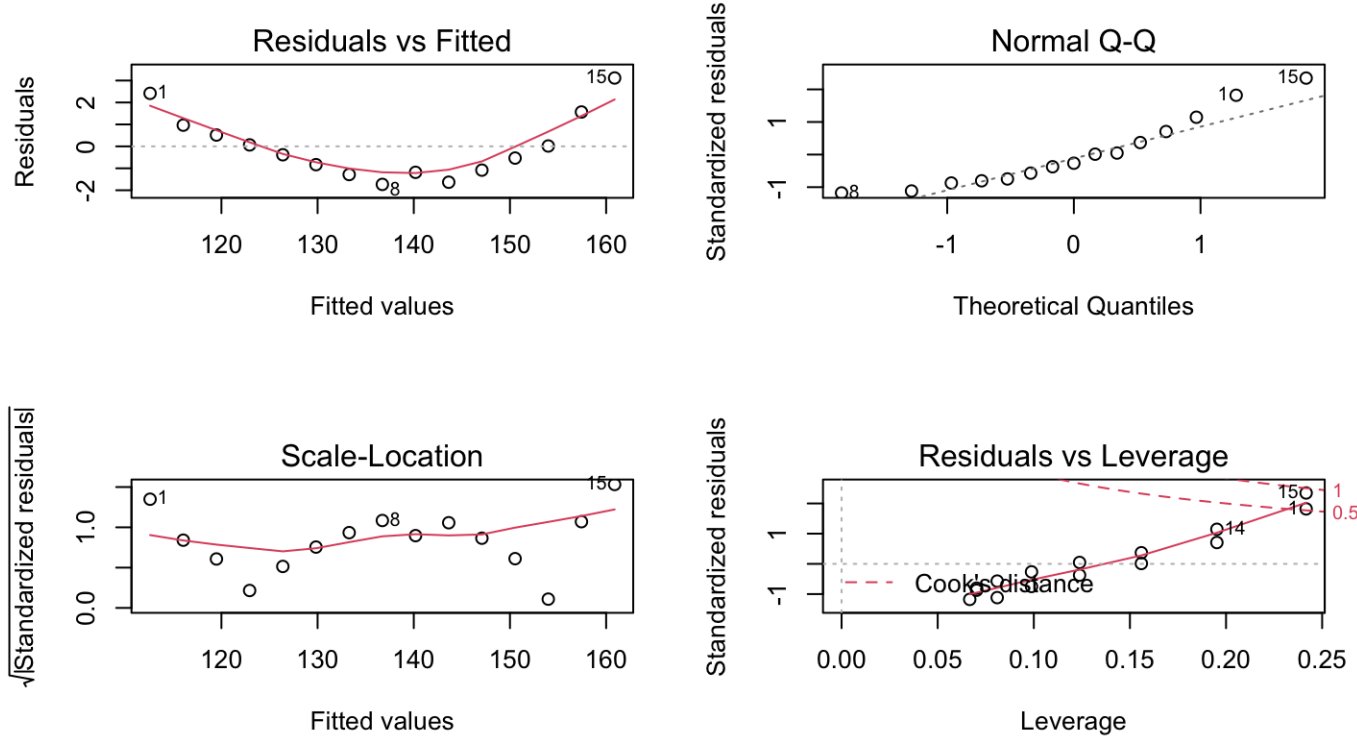
Modelos

- Detalles de la regresión.

```r
m1 <- lm(weight ~ height, data = women)
summary(m1)
```

```
##
## Call:
## lm(formula = weight ~ height, data = women)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7333 -1.1333 -0.3833  0.7417  3.1167
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -87.51667    5.93694  -14.74 1.71e-09 ***
## height        3.45000    0.09114   37.85 1.09e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.525 on 13 degrees of freedom
## Multiple R-squared:  0.991,  Adjusted R-squared:  0.9903
## F-statistic:  1433 on 1 and 13 DF,  p-value: 1.091e-14
```

Modelos

- Residuales ¿son homoscedásticos?

```
par(mfrow = c(2, 2))
plot(m1)
```

Modelos

Modelos

- Regresión múltiple, la relación de más de una variable predictora y una variable de respuesta.

```r
movies <- read.table("movies.csv", sep = ",", header = T)
head(movies)
```

```
##   Ganancias CostoProd CostoPromo LibrosVendidos
## 1      85.1       8.5   5.100000            4.7
## 2     106.3      12.9   5.800000            8.8
## 3      50.2       5.2   2.100000           15.1
## 4     130.6      10.7   8.399999           12.2
## 5      54.8       3.1   2.900000           10.6
## 6      30.3       3.5   1.200000            3.5
```

```
# Notoa: solicita a los facilitadores el conjunto de datos "movies.csv"
```

Modelos

- 1 variable predictora.

```r
lm.mov1 <- lm(Ganancias ~ CostoProd, data = movies)
summary(lm.mov1)
```

```
##
## Call:
## lm(formula = Ganancias ~ CostoProd, data = movies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.136  -9.029  -3.689   3.208  29.723
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   15.513     11.603   1.337 0.217989
## CostoProd      7.978      1.223   6.522 0.000184 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.26 on 8 degrees of freedom
## Multiple R-squared:  0.8417, Adjusted R-squared:  0.8219
## F-statistic: 42.54 on 1 and 8 DF,  p-value: 0.0001838
```

Modelos

- 2 variables predictoras.

```
lm.mov2 <- lm(Ganancias ~ CostoProd + CostoPromo, data = movies)
summary(lm.mov2)
```

```
##
## Call:
## lm(formula = Ganancias ~ CostoProd + CostoPromo, data = movies)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.4168  -2.5696   0.8052   2.1200  11.0463
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11.848      6.765   1.751  0.12334
## CostoProd       4.228      1.153   3.667  0.00800 **
## CostoPromo      7.436      1.806   4.117  0.00448 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.241 on 7 degrees of freedom
## Multiple R-squared:  0.9537, Adjusted R-squared:  0.9405
## F-statistic: 72.14 on 2 and 7 DF,  p-value: 2.131e-05
```

Modelos

- 3 variables predictoras.

```r
lm.mov3 <- lm(Ganancias ~ CostoProd + CostoPromo + LibrosVendidos, data = movies)
summary(lm.mov3)
```

```
##
## Call:
## lm(formula = Ganancias ~ CostoProd + CostoPromo + LibrosVendidos,
##     data = movies)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.4384  -3.1695   0.8499   3.5134   9.6207
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.6760     6.7602   1.135   0.2995
## CostoProd        3.6616     1.1178   3.276   0.0169 *
## CostoPromo       7.6211     1.6573   4.598   0.0037 **
## LibrosVendidos   0.8285     0.5394   1.536   0.1754
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.541 on 6 degrees of freedom
## Multiple R-squared:  0.9668, Adjusted R-squared:  0.9502
## F-statistic: 58.22 on 3 and 6 DF,  p-value: 7.913e-05
```

Modelos

- Comparando el modelo 1 contra modelo 2.

```
anova(lm.mov1, lm.mov2)
```

```
## Analysis of Variance Table
##
## Model 1: Ganancias ~ CostoProd
## Model 2: Ganancias ~ CostoProd + CostoPromo
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1      8 1626.27
## 2      7  475.37  1    1150.9 16.947 0.004478 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Modelos

- Comparando el modelo 2 contra modelo 3.

```
anova(lm.mov2, lm.mov3)
```

```
## Analysis of Variance Table
##
## Model 1: Ganancias ~ CostoProd + CostoPromo
## Model 2: Ganancias ~ CostoProd + CostoPromo + LibrosVendidos
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1      7 475.37
## 2      6 341.20  1    134.17 2.3594 0.1754
```

Licencia CC BY