Aplicación 1.1. Curva de Engel para el gasto en alimentos

J. Ramajo

2020

## El modelo de regresión lineal simple (con dos variables)

En este primer ejemplo veremos cómo realizar con R una estimación por mínimos cuadrados ordinarios (MCO) de un modelo de regesión simple, con dos variables. Utilizaremos un documento de R Markdown (<http://rmarkdown.rstudio.com>) para ejecutar los comandos y reproducir los resultados.

library(readr)  
engel <- read\_delim("ENGEL\_ALIM\_1.csv", ";", escape\_double = FALSE, trim\_ws = TRUE)

## Parsed with column specification:  
## cols(  
## GALIM = col\_double(),  
## RENTA = col\_double()  
## )

library(stargazer)

##   
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

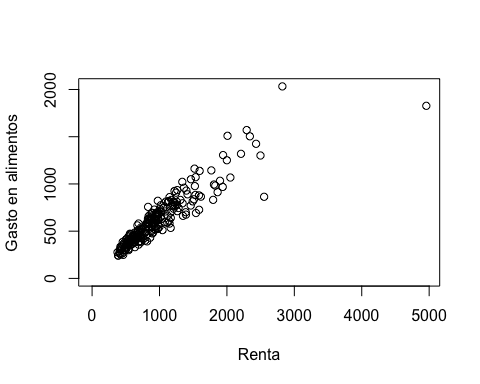
# Estadísticos descriptivos  
head(engel)

## # A tibble: 6 x 2  
## GALIM RENTA  
## <dbl> <dbl>  
## 1 256. 420.  
## 2 311. 541.  
## 3 486. 901.  
## 4 403. 639.  
## 5 496. 751.  
## 6 634. 946.

summary(engel)

## GALIM RENTA   
## Min. : 242.3 Min. : 377.1   
## 1st Qu.: 429.7 1st Qu.: 638.9   
## Median : 582.5 Median : 884.0   
## Mean : 624.2 Mean : 982.5   
## 3rd Qu.: 743.9 3rd Qu.:1164.0   
## Max. :2032.7 Max. :4957.8

# Diagrama de puntos (scatter plot) de las varialbes x e y.  
plot(engel$RENTA, engel$GALIM,  
ylim=c(0, max(engel$GALIM)),  
xlim=c(0, max(engel$RENTA)),  
xlab="Renta",  
ylab="Gasto en alimentos",  
type = "p")



# Estimación de la regresión lineal:

# GALIM\_i = beta\_1 + beta\_2 \* RENTA\_i + e\_i

# Estimación de un modelo por MCO y anclaje de los resultados en el objeto "ols"  
ols <- lm(GALIM ~ RENTA, data = engel)   
#  
# Resumen de resultados  
summary(ols)

##   
## Call:  
## lm(formula = GALIM ~ RENTA, data = engel)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -725.70 -60.24 -4.32 53.41 515.77   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 147.47539 15.95708 9.242 <2e-16 \*\*\*  
## RENTA 0.48518 0.01437 33.772 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 114.1 on 233 degrees of freedom  
## Multiple R-squared: 0.8304, Adjusted R-squared: 0.8296   
## F-statistic: 1141 on 1 and 233 DF, p-value: < 2.2e-16

ols$coefficients

## (Intercept) RENTA   
## 147.4753885 0.4851784

(b1 <- coef(ols)[[1]])

## [1] 147.4754

(b2 <- coef(ols)[[2]])

## [1] 0.4851784

vcov(ols)

## (Intercept) RENTA  
## (Intercept) 254.6283412 -0.2027754824  
## RENTA -0.2027755 0.0002063929

(varb1 <- vcov(ols)[1, 1])

## [1] 254.6283

(varb2 <- vcov(ols)[2, 2])

## [1] 0.0002063929

(covb1b2 <- vcov(ols)[1,2])

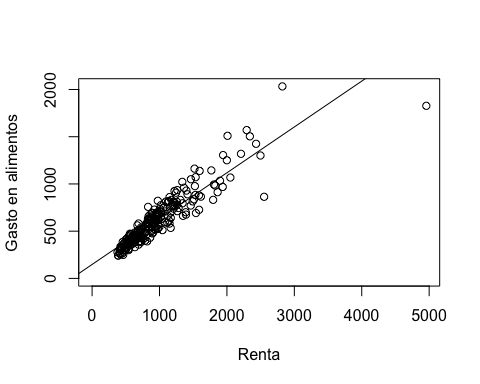
## [1] -0.2027755

#  
stargazer(ols, type = "text", title = "Resultados de la regresión")

##   
## Resultados de la regresión  
## ===============================================  
## Dependent variable:   
## ---------------------------  
## GALIM   
## -----------------------------------------------  
## RENTA 0.485\*\*\*   
## (0.014)   
##   
## Constant 147.475\*\*\*   
## (15.957)   
##   
## -----------------------------------------------  
## Observations 235   
## R2 0.830   
## Adjusted R2 0.830   
## Residual Std. Error 114.108 (df = 233)   
## F Statistic 1,140.534\*\*\* (df = 1; 233)   
## ===============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

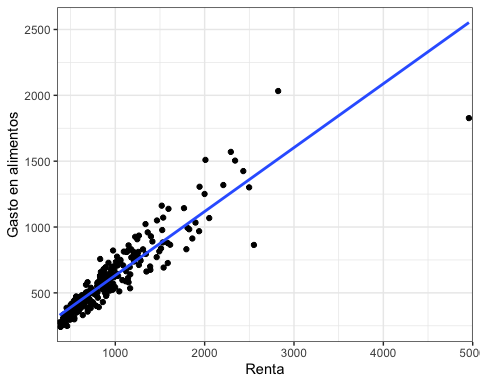
# Gráficas asociadas al modelo:

# Diagrama de puntos y línea de regresión  
# Método estándar  
b1 <- coef(ols)[[1]]  
b2 <- coef(ols)[[2]]  
#  
plot(engel$RENTA, engel$GALIM,  
ylim=c(0, max(engel$GALIM)),  
xlim=c(0, max(engel$RENTA)),  
xlab="Renta",  
ylab="Gasto en alimentos",  
type = "p")  
abline(b1,b2)  
# Gráfica de resultados con el método ggplot  
library(ggplot2)



ggplot(engel, aes(x = RENTA, y = GALIM)) + geom\_point() + geom\_smooth(method = "lm", se = FALSE) + scale\_x\_continuous(limits = c(350, 5000), expand = c(0, 0)) + theme\_bw() + labs(x = "Renta", y = "Gasto en alimentos")

## `geom\_smooth()` using formula 'y ~ x'



# Predicción con el modelo de regresión simple:

ols <- lm(GALIM ~ RENTA, data = engel)  
new\_RENTA <- data.frame(RENTA=c(400, 2000, 4500))  
pred\_GALIM <- predict(ols, new\_RENTA)  
names(pred\_GALIM) <-c("Renta = 400", "2000", "4500")  
pred\_GALIM

## Renta = 400 2000 4500   
## 341.5468 1117.8322 2330.7783