Econometría I

Pablo Astudillo-Estevéz Elías José Mantilla

2023-10-13

Introducción

La regresión es el ejercicio de ajustar rectas (o curvas) a un conjunto de datos. La regresión le permite a investigadores resumir como los valores típicos o predicciones de una variable dependiente cambian según varíen otras variables denotadas como independientes o predictores. Andrew Gelman en Regression And Other Stories es una gran referencia!

Election & the Economy

Un ejemplo de regresión aplicada ilustrado en el libro mencionado relaciona los porcentajes del voto popular que le corresponden al partido incumbente con los promedios de las tasas de crecimiento económica durante las administraciones de dichos incumbentes. A partir de esta misma base de datos se buscará ilustrar una rutina de trabajo en R.

```
# load libraries
library(tidyverse)

# load data
url <- paste0("https://raw.githubusercontent.com/avehtari/",
              "ROS-Examples/master/ElectionsEconomy/data/hibbs.dat")
hibbs <- read.table(url, sep=" ", header=TRUE) %>%
  as_tibble()
```

Análisis Exploratorio

```
ggplot(aes(vote), data = hibbs) +
  stat_ecdf(geom = "point")
```

```
corr <- cor(df$growth, df$vote)

ggplot(aes(growth, vote), data = hibbs) +
  geom_point() +
  annotate("text",
           x = 0.5,
           y = 60,
           label =
             paste0("r = ", round(corr, digits = 2)))
```

Estimación e Interpretación de Modelos de Regresión Lineal

```
fit.1 <- lm(vote ~ growth, data = hibbs)

arm::display(fit.1)
```

```
ggplot(aes(growth, vote), data = hibbs) +
  geom_point() +
  geom_abline(intercept =coef(fit.1)[1],
              slope = coef(fit.1)[2]) +
  annotate("text",
           x = 0.5,
           y = 60,
           label = str_glue(
             "y = {format(round(coef(fit.1)[1], digits = 2))} +",
             "{format(round(coef(fit.1)[2], digits = 2))} * x"
           ))
```

```
newdata <- tibble(growth = c(-1, 0, 1.5, 4))

tibble(
  .preds = predict(fit.1,
                   newdata = newdata),
  ) %>%
  cbind(newdata)
```

Evaluación y Diagnóstico

```
tibble(
  vote = hibbs$vote,
  .pred = predict(fit.1)) %>%
  ggplot(aes(vote, .pred)) +
  geom_point() +
  geom_abline()
```

```
tibble(
  vote = hibbs$vote,
  .pred = predict(fit.1)) %>%
  mutate(
```

```
    .resid = vote - .pred) %>%
  ggplot(aes(.pred, .resid)) +
  geom_point() +
  geom_hline(yintercept = 0)
```

Earnings

```
# read data
url <- paste0("https://raw.githubusercontent.com/avehtari/",
              "ROS-Examples/master/Earnings/data/earnings.csv")

earnings <-
  read_csv(url) %>%
  as_tibble()
```

Análisis Exploratorio

```
ggplot(aes(earn), data = earnings) +
  stat_ecdf(geom = "point") +
  scale_x_continuous(labels = scales::dollar)
```

```
ggplot(aes(earn), data = earnings) +
  stat_ecdf(geom = "point") +
  scale_x_log10(labels = scales::dollar)
```

```
ggplot(aes(earn, color = ethnicity), data = earnings) +
  stat_ecdf(geom = "point") +
  scale_x_log10(labels = scales::dollar)
```

Estimación e Interpretación de Modelos de Regresión Lineal

```
fit.2 <- lm(earnk ~ male, data = earnings)

arm::display(fit.2)
```

```r
ggplot(aes(male, earnk), data = earnings) +
  geom_jitter(width = 0.1) +
  geom_abline(intercept = coef(fit.2)[1],
              slope = coef(fit.2)[2]) +
  scale_y_continuous(limits = c(0, 200))
```

```r
fit.3 <- lm(earnk ~ male + height, data = earnings)

arm::display(fit.3)
```

```r
earnings <-
  earnings %>%
  mutate(c_height = height - mean(height))

fit.4 <- lm(earnk ~ male + c_height, data = earnings)

arm::display(fit.4)
```

```r
earnings <-
  earnings %>%
  mutate(
    white = if_else(ethnicity == "White", 1, 0),
    other = if_else(ethnicity == "Other", 1, 0),
    black = if_else(ethnicity == "Black", 1, 0),
    hispanic = if_else(ethnicity == "Hispanic", 1, 0))


fit.5 <- lm(earnk ~ c_height + male + white + other + hispanic, data =earnings)

arm::display(fit.5)
```