Regression Models for Data Science in R

Charlene Garridos

2023-03-08

Contents

1 Ordinary Least Squares

Ordinary least squares (OLS) is the *workhorse of statistics*. It gives a way of taking complicated outcomes and explaining behavior (such as trends) using linearity. The simplest application of OLS is fitting a line.

1.1 General least squares for linear equations

https://www.youtube.com/watch?v=LapyH7MG3Q4&list=PLpl-gQkQivXjqHAJd2t-J_One_fYE55tC&index=6

Fitting the best line: Let Y_i be the i^{th} child's height and X_i be the i^{th} (average over the pair of) parents' heights. Consider finding the best line

> Child's Height = β_0 + Parent's Height β_1.

Use least squares

$$\sum_{i=1}^{n} \{Y_i - (\beta_0 + \beta_1 X_i)\}^2$$

> **Note**: Minimizing this equation will minimize the sum of the squared distances between the fitted line at the parents' heights $\beta_i X_i$ and the observed child heights Y_i.

Result: The least squares of the line:

$$Y = \beta_0 + \beta_i X_i$$

through the data pairs X_i, Y_i with Y_i as the *outcome* obtains the line $Y = \hat{\beta}_0 + \hat{\beta}_1 X$ where:

$$\hat{\beta}_1 = Cor(Y, X)\frac{Sd(Y)}{Sd(X)} \text{ and } \hat{\beta}_0 = \bar{Y} = \hat{\beta}_1 \bar{X}$$

Elaborate:

- $\hat{\beta}_1$ has the units of Y/X, $\hat{\beta}_0$ has the units of Y.
- The line passes through the point $(\bar{X}, \bar{Y})$.
- The slope of the regression line with X as the outcome and Y as the predictor is $Cor(Y, X)Sd(X)/Sd(Y)$.The slope is the same one you would get if you centered the data, $(X_i - \bar{X}, Y_i - \bar{Y})$, and did regression through the origin.

Regression through the origin, assuming that $\beta_0 = 0$, yields the following solution to the *least squares criteria*:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}$$

> **Note:** If you normalized the data, $\{\frac{X_i - \bar{X}}{Sd(X)}, \frac{Y_i - \bar{Y}}{Sd(Y)}\}$, the slope is $Cor(Y, X)$.

1.2 Revisisting Galton's Data

https://www.youtube.com/watch?v=O7cDyrjWBBc&index=7&list=PLpl-gQkQivXjqHAJd2t-J_One_fYE55tC

1.3 Showing the OLS Result

Proof of why the ordinary least squares result works out to be the way that it is:

https://www.youtube.com/watch?v=COVQX8WZVA8&index=8&list=PLpl-gQkQivXjqHAJd2t-J_One_fYE55tC