Regression Ch 2

Lydia Gibson

2/19/2022

Simple Linear Regression

2.1 Intro and Least Square Estimates

2.1.1 Simple Linear Regression Models

The regression of a random variable Y on a random variable X is

$$E(Y|X = x),$$

the expected value of Y when X takes the specific value x.

The regression of Y on X is linear if

$$E(Y|X = x) = \beta_0 + \beta_1 x$$

where the unknown parameters β_0 and β_1 determine the intercept and the slope of a specific straight line, respectively.

If the regression of Y on X is linear, then for $i = 1, 2, \ldots, n$

$$Y_i = E(Y|X = x) + e_i = \beta_0 + \beta_1 x + e_1$$

where e is the random error in Y_i, and is such that $E(e|X) = 0$

Estimating the population slope and intercept

The equation of the line which "best" fits our data, that is, choose b_0 and b_1 such that $\hat{y} = b_0 + b_1 x$ is as close as possible to to y_i.

We shall refer to $\hat{y}_i$ as the ith **predicted value** or the **fitted value** of y_i, the observed values of y.

Residuals

We wish to minimize the difference between the actual value of y (y_i) and the predicted value of y ($\hat{y}_i$). The difference is called the residual, $\hat{e}_i$, that is,

$$\hat{e}_i = y_i - \hat{y}_i$$

Least square line of best fit

A very popular method of choosing b_0 and b_1 is called the method of least squares, which minimizes the sum of the squared residuals (or residual sum of squares [RSS]).

$$RSS = \sum_{i=1}^{n} \hat{e}_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_1)^2.$$

For RSS to be a minimus with respect to b_0 and b_1, we require

$$\frac{\partial RSS}{\partial b_0} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i) = 0$$

and

$$\frac{\partial RSS}{\partial b_1} = -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) = 0$$

Rearranging terms in the last two equations gives

$$\sum_{i=1}^{n} y_i = b_0 n + b_1 \sum_{i=1}^{n} x_i$$

and

$$\sum_{i=1}^{n} x_i y_i = b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2$$

.

These last two equations are called the **normal equations**. Solving these equations for b_0 and b_1 gives the so-called **least squares estimates** of the intercept

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

and the slope

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_)(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{SXY}{SXX}$$

Estimating the variance of the random error term (pg 19)

2.2 Inference About the Slope and the Intercept

2.2.1 Assumptions in order to make Inferences

2.2.2 Slope of the Regression Line

2.2.3 Intercept of the Regression Line

2.3 Confidence Intervals for the Population Regression Line

2.4 Prediction Intervals for the Actual Value of Y

2.5 Analysis of Variance

2.6 Dummy Variable Regression

2.7 Derivations of Results

2.7.1 The Slope of the Regression Line

2.7.2 The Intercept of the Regression Line

2.7.3CI for the Population Regression Line

2.7.4 Predicition Intervals for the Actual Value of Y