Regression Analysis TA

Chapter 1 – Linear Regression with One Predictor Variable

Chapter 2 – Inferences in Regression and Correlation Analysis

Po-Wei Chen

Institute of Statistics, National Chengchi University

March 13, 2023

1 Method of Least Squares

Simple Linear Regression Model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,\ \epsilon_i \overset{ind}{\sim} (0, \sigma^2)$$

The method of least squares considers the deviation of Y_i from its expected value:

$$Y_i - (\beta_0 + \beta_1 X_i)$$

The method of least squares requires that we consider the sum of the n squared deviations. This criterion is denoted by Q:

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

According to the method of least squares, the estimators of β_0 and β_1 are those values b_0 and b_1 respectively, that minimize the criterion Q for the given sample observations $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$.

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}, \qquad b_0 = \bar{Y} - b_1 \bar{X}$$

2 Method of Maximum Likelihood

Normal Error Regression Model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,\ \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

$$\hat{\beta}_{1MLE} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}, \qquad \hat{\beta}_{0MLE} = \bar{Y} - \hat{\beta}_{1MLE} \bar{X}, \qquad \hat{\sigma}^2_{MLE} = \frac{\sum (Y_i - \hat{Y}_i^2)}{n}$$

3 Problems

1.20. **Copier maintenance.** The Tri-City Office Equipment Corporation sells an imported copier on a franchise basis and performs preventive maintenance and repair service on this copier. The data below have been collected from 45 recent calls on users to perform routine preventive maintenance service; for each call, X is the number of copiers serviced and Y is the total number of minutes spent by the service person. Assume that first -order regression model (1.1) is appropriate.

i:	1	2	3	⋯	43	44	45
X_i	2	4	3	⋯	2	4	5
Y_i	20	60	46	⋯	27	61	77

(a) Obtain the estimated regression function.

(b) Plot the estimated regression function and the data. How well does the estimated regression function fit the data?

(c) Interpret b_0 in your estimated regression function. Does bo provide any relevant information here? Explain.

(d) Obtain a point estimate of the mean service time when $X = 5$ copiers are serviced.

2.5. Refer to **Copier maintenance** Problem 1.20.

(a) Estimate the change in the mean service time when the number of copiers serviced increases by one. Use a 90 percent confidence interval. Interpret your confidence interval.

(b) Conduct a t test to determine whether or not there is a linear association between X and Y here; control the a risk at .10. State the alternatives, decision rule, and conclusion. What is the P-value of your test?

(c) Are your results in parts (a) and (b) consistent? Explain.

(d) The manufacturer has suggested that the mean required time should not increase by more than 14 minutes for each additional copier that is serviced on a service call. Conduct a test to decide whether this standard is being satisfied by Tri-City. Control the risk of a Type I error at .05. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

4 Exercises

1.41. Refer to the regression model $Y_i = \beta_1 X_i + \epsilon_i, i = 1, \cdots, n$, in Exercise 1.29.

(a) Find the least squares estimator of β_1.

(b) Assume that the error terms ϵ_i are independent $\mathcal{N}(0, \sigma^2)$ and that σ^2 is known. State the likelihood function for the n sample observations on Y and obtain the maximum likelihood estimator of β_1. Is it the same as the least squares estimator?

(c) Show that the maximum likelihood estimator of β_1 is unbiased.

Unbiasedness

Let $X_1, X_2, ..., X_n$ denote a sample on a random variable X with pdf $f(x;\theta),\ \theta \in \Omega$. Let $T = T(X_1, X_2, ..., X_n)$ be a statistic. We say that T is an unbiased estimator of θ if $E(T) = \theta$.