1: Regression

(a) Reason:

State with reason whether the following statements are true, false, or uncertain. Be precise.

- The t test of significance discussed in this chapter requires that the sampling distributions of estimators β_0 and β_1 follow the normal distribution.
- Even though the disturbance term in the CLRM is not normally distributed, the OLS estimators are still unbiased.
- If there is no intercept in the regression model, the estimated $u_i (= \hat{u}_i)$ will not sum to zero.
- The p value and the size of a test statistic mean the same thing.
- In a regression model that contains the intercept, the sum of the residuals is always zero.
- If a null hypothesis is notrejected, it is true.
- The higher the value of σ^2 , the larger is the variance of β_1.
- The conditional and unconditional means of a random variable are the same things.
- In the two–variable PRF, if the slope coefficient β_1 is zero, the intercept β_0 is estimated by the sample mean $\bar{Y}$.
- The conditional variance, $var(Y_i|X_i) = \sigma^2$, and the unconditional variance of Y, $var(Y) = \sigma^2$, will be the same if X had no influence on Y.

(b) Reason:

Let ρ^2 represent the true population coefficient of determination. Suppose you want to test the hypothesis that $\rho^2 = 0$. Verbally explain how you would test this hypothesis.

(c) Solve:

What is known as the characteristic line of modern investment analysis is simply the regression line obtained from the following model:

$$r_{it} = \alpha_i + \beta_i r_{mt} + u_t$$

where:

- r_{it} = the rate of return on the i–th security in time t,
- r_{mt} = the rate of return on the market portfolio in time t,
- u_t = stochastic disturbance term.

In this model β_i is known as the beta coefficient of the i–th security, a measure of market (or systematic) risk of a security.

On the basis of 240 monthly rates of return for the period 1956–1976, Fogler and Ganapathy obtained the following characteristic line for IBM stock in relation to the market portfolio index developed at the University of Chicago:

$$\begin{array}{ll} \hat{r}_{it} = 0.7264 + 1.0598\, r_{mt} & r^2 = 0.4710 \\ \text{se} = (0.3001)\ (0.0728) & df = 238 \\ & F_{1,238} = 211.896 \end{array}$$

(a) A security whose beta coefficient is greater than one is said to be a volatile or aggressive security. Was IBM a volatile security in the time period under study?

(b) Is the intercept coefficient significantly different from zero? If it is, what is its practical meaning?

(d) Solve:

Consider the following regression output:

$$\begin{array}{lll} \hat{Y}_i = 0.2033 + 0.6560\, X_t & & \\ \text{se} = (0.0976)\ (0.1961) & & \\ r^2 = 0.397 & RSS = 0.0544 & ESS = 0.0358 \end{array}$$

where Y = labor force participation rate (LFPR) of women in 1972 and X = LFPR of women in 1968. The regression results were obtained from a sample of 19 cities in the United States.

(a) How do you interpret this regression?

(b) Test the hypothesis: $H_0 : \beta_1 = 1$ against $H_1 : \beta_1 > 1$. Which test do you use? And why? What are the underlying assumptions of the test(s) you use?

(c) Suppose that the LFPR in 1968 was 0.58 (or 58%). On the basis of the regression results given above, what is the mean LFPR in 1972? Establish a 95 percent confidence interval for the mean prediction.

(d) How would you test the hypothesis that the error term in the population regression is normally distributed? Show the necessary calculations.

(e) R commands: (with interpretation)

Grade point average (continued from the previous homework). The director of admissions of a small college selected 120 students at random from the new freshman class in a study to determine whether a student's grade point average (OPA) at the end of the freshman year (Y) can be predicted from the ACT test score (X). The results of the study follow. Assume that first–order regression model is appropriate.

Data are available in the homework page (the link is near this homework).

(a) Obtain a 95% interval estimate of the mean freshman OPA for students whose ACT test score is 28. Interpret your confidence interval.

(b) Mary Jones obtained a score of 28 on the entrance test. Predict her freshman OPA–using a 95% prediction interval. Interpret your prediction interval.

(c) Is the prediction interval in point (b) wider than the confidence interval in point (a)? Should it be? Explain the reason.

(d) Determine the boundary values of the 95% confidence band for the regression line when $X = 28$. Is your confidence band wider at this point than the confidence interval in point (a)? Should it be?

(e) Set up the ANOVA table.

(f) What is estimated by MSR in your ANOVA table? by MSE? Under what condition do MSR and MSE estimate the same quantity?

(g) Conduct an F test of whether or not $\beta_1 = 0$. Control the α risk at 01. State the alternatives, decision rule, and conclusion.

(h) What is the absolute magnitude of the reduction in the variation of Y when X is introduced into the regression model? What is the relative reduction? What is the name of the latter measure?

(i) Obtain r and attach the appropriate sign.

(j) Which measure, R^2 or r, has the more clear–cut operational interpretation? Explain.

(k) Plot an histogram of the residuals and comment it.

(l) Plot a Q–Q plot of the residuals and comment it.

(m) Compute a Jarque–Bera test of normality.

NOTE: Solve the exercise in the following two ways:

- exploiting the proper regression R functions
- using R as a calculator (without exploiting internal functions for regression, but using only the arithmetic operators and the mean function).