

Contest Quiz 6

Question Sheet

In this quiz we will cover the concepts of testing hypothesis and multiple linear regression.

NOTE: Use the full data set unless stated otherwise. Please round your results to two decimal places. Do not round any interim results.

EXAMPLE: If your unrounded solution is 0.13897439, drop all decimal places except the first three. This leaves you with 0.138. If the third decimal place is 5 or above (as is the case here), round up. This gives 0.14.

Question 1

Load the bank.txt data set into EViews or RExcel from the url:
<http://thiloklein.de/R/bank>

We are interested in describing the salaries in a US bank based on employees' education, beginning salary, gender, minority and category of job (custodial for jobcat=1, clerical for jobcat=2, managerial for jobcat=3). Run the regression:

$$\log(\text{salary})_i = \beta_0 + \beta_1 \text{educ}_i + \beta_3 \log(\text{salbegin})_i + \beta_4 \text{male}_i + \beta_5 \text{minority}_i + \beta_6 \text{clerical}_i + \beta_7 \text{managerial}_i + u_i$$

- I) What is the expected difference in salary between an individual with clerical job and an individual with custodial job?
(a) 0 (b) 9.39% (c) 12.16% (d) 20.13%
- II) What is the expected difference in salary between an individual with managerial job and an individual with custodial job?
(a) 0 (b) 9.39% (c) 12.16% (d) 20.13%
- III) Are there any significant differences in salary between clerical job and custodial jobs (assume a 5% significance level)?
(a) Yes (b) No
- IV) Are there any significant differences in salary between managerial job and custodial jobs (assume a 5% significance level)?
(a) Yes (b) No
- V) Are there any significant differences in salary between clerical job and managerial jobs (assume a 5% significance level)?
(a) Yes (b) No

VI) Is there a variation in the salary due to being simultaneously female and manager? (assume same model as above and add only a femalemanager variable)

(a) Yes (b) No

Now return to the original model

VII) Evaluate the F-statistic corresponding to the hypothesis that custodial and clerical jobs have the same discriminatory effect joint with the hypothesis that for every year more spent in education salary tends to be 5% more.

(a) 50.67 (b) 26.00 (c) 2.39 (d) 43.81

VIII) Evaluate the unadjusted R^2 of the model stated above

(a) 0.83 (b) 0.86 (c) 0.82 (d) 0.99

Question 2

We are interested in interpreting the regression coefficients for the following fitted equations. Assume that all coefficients are significant.

I) Given that the estimated regression equation is

$$earnings_i = -13.93 + 2.46schooling_i + e_i$$

which of the following statement is true? (assume that schooling is measured in years and earnings in thousand pounds)

(a) for every extra year spent in school, earnings increases on average by 2.46% (b) for every extra year spent in school, earnings increases on average by 2.46 thousands (c) there is no change in average earning for every increase in years of schooling (d) for every percent increase in schooling, earnings increases by 2.46 thousands.

II) Given that the estimated regression equation is

$$\log(earnings)_i = -13.93 + 2.46schooling_i + e_i$$

which of the following statement is true? (assume that schooling is measured in years and earnings in thousand pounds)

(a) for every extra year spent in school, earnings increases on average by 2.46% (b) for every extra year spent in school, earnings increases on average by 2.46 thousand (c) for every extra year spent in school, earnings increases on average by 246% (d) for every percent increase in schooling earnings increases on average by 2.46 thousands

III) Given that the estimated regression equation is

$$\log(foodexpenditure)_i = 3.16 + 0.48\log(totalexpenditure)_i + e_i$$

which of the following statement is true? (assume that food and total expenditure is measured in dollars)

(a) 0.48 dollars out of the marginal dollar are spend on food eaten at home (b) for every extra dollar spent in total, 48% is spent on average on food (c) 1 percent change in total expenditure will cause on average a 48% change in food expenditure (d) None of the above

Question 3

Use gasoline.txt (<http://thiloklein.de/R/gasoline>) and estimate the demand for gasoline as a function of GDP and the price of gasoline (include a constant in the regression). For this question only, round your answers in to 3 decimal places. Estimate the model:

$$\log(\text{gasoline})_i = \beta_0 + \beta_1 \log(\text{GDP}) + \beta_2 \log(\text{price})_i + u_i$$

- I) Evaluate the sum of squares residuals
(a) 0.060 (b) 0.032 (c) 0.975 (d) 0.047
- II) Consider the restriction of $\beta_1 = 1$. Evaluate the sum of squares residual for the restricted model.
(a) 0.060 (b) 0.066 (c) 0.067 (d) 0.068
- III) By using (I) and (II) evaluate the F-statistic for the hypothesis that our restriction in (II) is indeed valid. Please use the following formula:

$$F = \frac{RSS_{res.} - RSS_{unres.}}{RSS_{unres.}/(n - k)}$$

where n is the number of observations and k the number of parameters in the unrestricted model.

- (a) 5.658 (b) 5.681 (c) 5.147 (d) 5.800
- IV) Given that $F(1,58)$ at the 5% significance level is 4.007 what do we conclude about the hypothesis $H_0: \beta_1 = 1$ vs $H_A: \beta_1 \neq 1$?
(a) Reject the null at 5% significance level (b) Do not reject the null at 5% significance level

Question 4

Rerun the regression in Question 1.

- I) Evaluate the variance inflation factor (VIF) for the variable $\log(\text{salbegin})$
(a) 0.06 (b) 2.02 (c) 2.14 (d) 4.33
- II) Evaluate VIF for the variable educ
(a) 0.06 (b) 2.02 (c) 2.14 (d) 4.33
- III) Which variable has the largest VIF
(a) male (b) minority (c) clerical (d) $\log(\text{salbegin})$
- IV) In general, the VIF-statistic ranges between which values??
(a) $(-\infty, +\infty)$ (b) $(0, +\infty)$ (c) $(1, +\infty)$ (d) $(-1, +1)$