

Extra Practice Questions - Week 2

Statistics and Econometrics

Question 1

Using the data set `ceosal1.RData` to answer the following questions. Consider an equation to explain salaries of CEOs in terms of annual firm sales, return on equity (*roe*, in percentage form), and return on the firm's stock (*ros*, in percentage form):

$$\log(\text{salary}) = \beta_0 + \beta_1 \log(\text{sales}) + \beta_2 \text{roe} + \beta_3 \text{ros} + u$$

1. In terms of the model parameters, state the null hypothesis that, after controlling for *sales* and *roe*, *ros* has no effect on CEO salary. State the alternative that better stock market performance increases a CEO's salary.
2. Estimate the model and report your results. By what percentage is *salary* predicted to increase if *ros* increases by 50 basis points (i.e., *ros* increases by 50)? Does *ros* have a practically large effect on *salary*?
3. Test the null hypothesis that *ros* has no effect on *salary* against the alternative that *ros* has a positive effect. Carry out the test at the 10% significance level (please show clearly the test statistic and the critical value used in your testing).
4. Would you include *ros* in a final model explaining CEO compensation in terms of firm performance? Explain.

Question 2

Use the data set `lawsch85.RData` to answer the following questions. Consider an equation to explain the median starting salary for new law school graduates

$$\log(\text{salary}) = \beta_0 + \beta_1 LSAT + \beta_2 GPA + \beta_3 \log(\text{libvol}) + \beta_4 \log(\text{cost}) + \beta_5 \text{rank} + u,$$

where *LSAT* is the median LSAT score for the graduating class, *GPA* is the median college GPA for the class, *libvol* is the number of volumes in the law school library, *cost* is the annual cost of attending law school, and *rank* is a law school ranking (with *rank* = 1 being the best).

1. Estimate the model. State and test the null hypothesis that the rank of law schools has no causal effect on median starting salary (please show clearly the test statistic and the critical value used in your testing).
2. Are features of students - namely, *LSAT* and *GPA* - individually or jointly significant for explaining *salary*?
3. Test whether the size of the class (*clsize*) or the size of the faculty (*faculty*) needs to be added to this equation: carry out a single test for joint significance of the two variables.

Question 3

Use the data in `discrim.RData` to answer this question. These are ZIP code-level data on prices for various items at fast-food restaurants, along with characteristics of the zip code population, in New Jersey and Pennsylvania. The idea is to see whether fast-food restaurants charge higher prices in areas with a larger concentration of African Americans.

1. Consider a model to explain the price of soda in its log form, $\log(\text{psoda})$, in terms of the proportion of the population that is African American and log of median income:

$$\log(\text{psoda}) = \beta_0 + \beta_1 \text{prpbldk} + \beta_2 \log(\text{income}) + u.$$

Estimate the model and report your results. If *prpblick* increases by .20, what is the estimated percentage change in *psoda*?

2. Compare the estimate from part 1 with the simple regression estimate from $\log(\textit{psoda})$ on *prpblick*. Is the discrimination effect larger or smaller when you control for income? Explain.
3. Now add the variable *prppov* to the regression in part 1 and report your results. What happens to $\hat{\beta}_{\textit{prpblick}}$? Is $\hat{\beta}_{\textit{prpblick}}$ statistically different from zero at the 5% level against a two-sided alternative? What about at the 1% level? (please show clearly the test statistic and the critical value used in your testing)
4. Find the correlation between $\log(\textit{income})$ and *prppov*. Is it roughly what you expected?
5. Now add the variable $\log(\textit{hseval})$ to the regression in part 3. Interpret its coefficient and report the two-sided p-value for $H_0 : \beta_{\log(\textit{hseval})} = 0$.
6. In the regression in part 5, what happens to the individual statistical significance of $\log(\textit{income})$ and *prppov*? Are these variables jointly significant? (Report p-value) What do you make of your answers?
7. Given the results of the previous regressions, which one would you report as most reliable in determining whether the racial makeup of a zip code influences local fast-food prices?