Applied Statistics and Econometrics (Econ 8740)
Prof. Pierre Nguimkeu
GSU Department of Economics, Fall 2017

Practice Questions for Midterm

Question 1.

A researcher plans to study the effect of police on crime using data from a random sample of US cities. She plans to regress the city's crime rate on the (per capita) size of the city's police force.

(a) Explain why this regression is likely to suffer from omitted variable bias.

(b) Based on your answer in part (a), determine whether the regression will likely over- or underestimate the effect of police on the crime rate.

Question 2.

Suppose we are interested in determining the linear relationship between time spent studying (in hours) and an index of exam performance for students. A sample of three students was selected and the following summary statistics were recorded:

$$\sum x = 3 \qquad \sum x^2 = 3.5 \qquad \sum y = 6 \qquad \sum y^2 = 14 \qquad \sum xy = 6.5$$

(a) Define in the context of the problem what x and y are.

(b) What is the strength of the linear association between the time spent studying and exam performance?

(c) Find the sample regression line. Interpret the least squares coefficients.

(d) Compute the coefficient of determination and explain its meaning.

(e) Consider the hypotheses $H_0 : \beta_1 = 0$ and $H_1 : \beta_1 > 0$. If the standard error of $\hat{\beta}_1$ is 1.3, what is the smallest significance level for which the null hypothesis can be rejected?

Question 3. [Midterm, Fall 2013]

Consider the multiple linear regression with three independent variables. This model satisfies classical linear model assumptions MLR.1 through MLR.6.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

You would like to test the null hypothesis $H_0 : \beta_1 - 2\beta_2 = 1$ against alternative $H_1 : \beta_1 - 2\beta_2 \neq 1$.

(a) Let $\hat{\beta}_1$ and $\hat{\beta}_2$ denote the OLS estimators of β_1 and β_2. Find $\text{Var}[\hat{\beta}_1 - 2\hat{\beta}_2]$ in terms of variances of $\hat{\beta}_1$ and $\hat{\beta}_2$ and the covariance between them. Write the t-statistic for H_0. (9points)

(b) Define $\theta_1 = \beta_1 - 2\beta_2$. Write a regression involving θ_1, $\hat{\beta}_1$ and $\hat{\beta}_2$ that allows you to directly obtain $\hat{\theta}_1$, and its standard error. (7 points)

(c) Assuming that the hypothesis H_0 holds, write down the restricted model to be estimated. Explain how you would test H_0 using an F-statistic? Be sure to write out the formula for the F-statistic, and highlight what your unrestricted model is. What is the value of q and k from the F-statistic? (9 points)

Question 4. [Midterm, Fall 2013]

The following model describes the log monthly salary (variable lsalary) of a sample of lawyers in terms of their performance on the LSAT (variable LSAT; LSAT is the exam that individuals take to get into law school) and the log of the annual cost of attending law school (variable lcost). Denote the results from this regression with a hat (i.e. $\hat{\beta}$).

```
reg lsalary LSAT GPA
```

Source	SS	df	MS
Model	6.81992403	2	3.40996201
Residual	4.03690832	139	.029042506
Total	10.8568324	141	.076998811

Number of obs =	142
F(2, 139) =	117.41
Prob > F =	0.0000
R-squared =	[redacted]
Adj R-squared =	[redacted]
Root MSE =	.17042

lsalary	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
LSAT	.0267257	.0048707	5.49	0.000	[redacted]	[redacted]
GPA	.5289411	.1133635	4.67	0.000	.3048013	.7530809
_cons	4.55601	.5333218	8.54	0.000	3.501538	5.610482

(a) Interpret the coefficient estimate on lcost. How much is salary predicted to increase when an individuals LSAT score increases by 10 points? Calculate the R^2 for the model and interpret it. (12 points)

(b) Suppose that we regress LSAT on lcost and obtain the regression output below. Using this information, calculate the estimated coefficient on lcost that you would obtain from the simple linear regression: $lsalary = \delta_0 + \delta_1 lcost + u$.
Does this indicate that $\hat{\delta}$ is upward or downward biased? (6 points)

```
reg LSAT GPA
```

Source	SS	df	MS
Model	1925.4132	1	1925.4132
Residual	1280.73445	147	8.71247925
Total	3206.14765	148	21.6631598

Number of obs =	149
F(1, 147) =	220.99
Prob > F =	0.0000
R-squared =	0.6005
Adj R-squared =	0.5978
Root MSE =	2.9517

LSAT	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
GPA	18.05456	1.214495	14.87	0.000	15.65443	20.45468
_cons	98.72598	4.018458	24.57	0.000	90.78457	106.6674

(c) Return to the regression model in part (a). Test whether the coefficient estimate on LSAT is statistically equal to 0.05. Use a 2-sided test, with $\alpha = 5\%$. Specify the null, alternative, critical value, and degree of freedom. Make sure to show your work. (7 points)

Question 5. [Midterm, Fall 2015]

You estimate a model of tradeoff between time spent sleeping (*sleep*) and working (*totwrk*):

$$\begin{array}{rcl} \widehat{sleep} & = & 3,638.25 - 0.148totwrk - 11.13educ + 2.20age \\ & & (112.28) \quad (0.017) \qquad (5.88) \qquad (1.45) \qquad n = 706 \quad R^2 = 0.113. \end{array}$$

(a) Is *educ* or *age* individually significant at the 5% level? Justify your answer.(2 points)

(b) Dropping *educ* and *age* from the equation gives

$$\begin{array}{rcl} \widehat{sleep} & = & 3,586 - 0.151totwrk \\ & & (37.82) \quad (0.017) \qquad n = 706 \quad R^2 = 0.103. \end{array}$$

Are *educ* and *age* jointly significant in the original equation at 5%? Justify. (2 points)

(c) Does including *educ* and *age* in the model greatly affect the estimated tradeoff between sleeping and working? Would you keep them in your final results report? (2 points)

(d) Suppose the sleep equation contains heteroskedasticity. What does this mean about the tests computed in parts (a) and (b)? (1 point)

Homework 1

Homework 2

Chapter 2: Problem 7

Chapter 3: Problem 5, 7

Chapter 4: Problems 4, 7