Final exam

5/1/2013

Instructions: You may use a calculator, the normal probability table attached to the back of this exam, and one sheet of notes. Show your work. Each multiple choice question is worth 2 points, while each short answer question is worth 5 points. As there are 10 multiple choice questions and 16 short answer sub-questions, the exam is out of 100 points. Good luck!

Problem 1 You estimate the following regression model:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_1 * X_2 + \beta_4 * X_2^2 + \epsilon \tag{1}$$

You obtain the following estimates:

Coefficient	estimate
β_0	2
β_1	4
β_2	1
β_3	-1
β_4	-2

For a given observation, $X_2 = 5$. Based on your results, what is the marginal effect on Y of a one-unit increase in X_1 from $X_1 = 6$ to $X_1 = 7$.

a. decrease by 1 unit.

b. increase by 1 unit.

c. increase by 6 units.

d. decrease by 16 units.

e. none of the above

Problem 2 Referring to the regression results in problem 1, what is the effect of a one-unit increase in X_2 from 5 to 6, given that $X_1 = 7$?

a. decrease by 4 units.

b. increase by 9 units.

c. increase by 16 units.

d. decrease by 28 units.

e. none of the above

Problem 3 Suppose an interval estimate for the population mean was $[50, 58]$. The standard deviation is $\sigma = 3.2$, and a sample of size $n = 71$ was used. The mean of the sample was:

a. 54

b. 55

c. 56

d. 57

e. none of the above

Problem 4 Carl takes a random sample of size 49 from a population of interest and finds that $\bar{x} = 21$, $\sigma = 7$. What is the 95% confidence interval associated with Carl's data?

a. $[18.69, 23.31]$

b. $[19.36, 22.64]$

c. $[19.04, 22.96]$

d. $[20, 22]$

e. none of the above

Problem 5 David wishes to perform the following hypothesis test, with $\alpha = .1$:

$$\begin{aligned} H_0 &: \mu = 0 \\ H_A &: \mu \neq 0 \end{aligned} \tag{2}$$

David should reject H_0 if $\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ is greater than X or less than $-X$, where X equals:

a. 1.28

b. 1.64

c. 1.96

d. 2.31

e. none of the above

Problem 6 Edgar is interested in the following hypothesis test:

$$\begin{aligned} H_0 &: \mu = 8 \\ H_A &: \mu \neq 8 \end{aligned} \tag{3}$$

Edgar collects a random sample from the population of interest and finds that $\bar{x} = 8.5$, $\sigma = 4$, $n = 256$. The p-value of Edgar's test is:

a. .008

b. .016

c. .023

d. .046

e. none of the above

Problem 7 To estimate the mean of a normal population whose standard deviation is 6, with a bound on the error of estimation equal to 1.2 and confidence level of 99% requires a sample size of at least:

a. 166

b. 167

c. 13

d. None of these choices

Problem 8 The owner of a local nightclub has recently surveyed a random sample of n = 300 customers of the club. She would now like to determine whether or not the mean age of her customers is over 35. If so, she plans to alter the entertainment to appeal to an older crowd. If not, no entertainment changes will be made. Suppose she found that the sample mean was 35.5 years and the population standard deviation was 5 years. What is the p-value associated with the test statistic?

a. .9582

b. 1.73

c. .0418

d. .0836

Problem 9 If the probability of committing a Type I error for a given test is decreased, then for a fixed sample size n, the probability of committing a Type II error will:

a. decrease

b. increase

c. stay the same

d. There not enough information to tell.

Problem 10 Y is a normal random variable with mean -7 and standard deviation 3. What is the probability that a draw from Y is between -13 and -4?

ANSWER:----------

Problem 11 Alice is interested in how salaries for similar work differ across the public and private sectors. She obtains employment data on 30,000 workers randomly sampled from the entire country.

She uses the following regression model, where $SALARY$ is annual salary in dollars, and $PUBLIC$ equals 1 if an individual is employed in the public sector, and 0 otherwise:

$$SALARY = \beta_0 + \beta_1 * PUBLIC + \epsilon \tag{4}$$

She obtains the following results:

R^2 .091

coefficient	estimate	p-value
Intercept	45,000	.00012
β_1	12,000	.03549

a. Interpret Alice's estimate of β_1.

b. Alice suggests that her results prove that public sector wages are excessive, and that governments could lower the salaries they offer and attract the same workers. Do you agree? Explain.

(continued on next page)

Bob thinks that public sector employees may have different individual characteristics than private sector employees. Using Alice's data, Bob estimates the following regression:

$$SALARY = \beta_0 + \beta_1 * PUBLIC + \beta_2 * EDUCATION + \beta_3 * AGE + \beta_4 * AGE^2 + \epsilon \quad (5)$$

where $EDUCATION$ is years of education and AGE is age of the employee. He obtains the following results:

R^2 .63

coefficient	estimate	p-value
β_0	$25,000$	.00005
β_1	$4,000$	.00312
β_2	$1,200$	.01579
β_3	300	.02697
β_4	-3	.00049

c. Bob's estimate of β_1 is lower than Alice's. Explain in simple, intuitive terms why this might be the case.

d. Based on a comparison of Bob's and Alice's results, do you think that public sector employees are more educated, or less educated than private sector employees, on average?

(continued on next page)

e. Do Bob's results predict that a 40-year old public-sector employee or a 60-year old public sector employee will make more, all else equal?

f. At approximately what age do Bob's results predict that SALARY will peak?

Problem 12 The following table is taken from Klick and Tabarrok, "Using terror alert levels to estimate the effect of police on crime". (1) and (2) denote two separate regressions run by the authors.

	(1)	(2)
High alert	-7.316^{*}	-6.046^{*}
	(2.877)	(2.537)
Log(midday ridership)		17.341^{**}
		(5.309)
R^2	.14	.17

NOTE: The dependent variable is the daily total number of crimes in Washington D.C. during the period March 12,2002–July 30,2003. Both regressions contain day-of-the-week fixed effects. The number of observations is 506. Standard errors are in parentheses.

* Significantly different from zero at the 5 percent level.

** Significantly different from zero at the 1 percent level.

a. In specification (1), interpret the coefficient estimate in plain English. Is it significant? What does it tell us?

b. Why did the authors choose midday ridership (on the DC subway system) as a control variable in specification (2)?

(continued on next page)

The following appears as column 1 of Table 4 in the same paper:

	Coefficient (standard error)
High Alert * District 1	-2.621^{**}
	(.044)
High Alert * Other Districts	$-.571$
	(.455)
Log(midday ridership)	2.477^{*}
	(.364)
Constant	-11.058^{**}
	(4.211)

NOTE: $R^2 = .28$. The dependent variable is the daily total number of crimes by district. The regression contains day-of-the-week fixed effects. Standard errors are in parentheses. The number of observations is 3,542.

* Significantly different from zero at the 5 percent level.

** Significantly different from zero at the 1 percent level.

c. Interpret the coefficient to High Alert * District 1 in plain English. Is it significant? What does it tell us about the effect of police in District 1?

d. Explain why the authors chose to single out District 1. What did they conclude about the magnitude of the elasticity of crime with respect to police in District 1?

Problem 13 Answer the following 2 questions about working with aggregated/disaggregated data.

a. Dean Blackwell notices that, relative to 1993, each Gatton major (economics, marketing, accounting, analytics, management, and finance) has a higher 4-year graduation rate. The majors offered by Gatton have not changed in the last 20 years. True/false/uncertain: Gatton college has a higher 4-year graduation rate in 2013 than in 1993. Explain.

b. Dean Walz of the engineering college, notices that relative to 1993, the college has a higher overall 4-year graduation rate. The majors offered by the engineering college have not changed in the last 20 years. True/false/uncertain: every major in the engineering college has a higher 4-year graduation rate in 2013 than that same major did in 1993. Explain.

Problem 14 Irina, a UK faculty member on the UK board of trustees, routinely sends out emails to all UK faculty complaining about various aspects of university governance. An example:

> Yes, the majority of board members seem hardly ever outside of the 18th floor of POT and the admin building. When they're at UK, all the happy meals and happy spin (tons of it) are brought to them. Unless the news on the true state of affairs travels to the penthouse, they are more or less clueless. Faculty trustees do try to explain what is really taking place on the ground. But, it's hard to penetrate the happy talk that fills the penthouse, and faculty trustees can be dismissed as a cranky minority.

Recently, Irina distributed a petition with inflammatory language directed towards university leaders to all UK faculty, and asked all of us to vote on whether or not we supported the petition. The results:

> 16 people voted against the petition, and 217 people voted in support of it. Therefore, there is overwhelming faculty sentiment in favor of the petition, which cannot and must not be dismissed.

There are about 2,000 UK faculty members. Explain, in plain English, what statistical error Irina is making. Your answer should be comprehensible to a 6-year old.

Problem 15 A study of health outcomes in the country of Elbonia finds that 85% of cancer patients are daily tea drinkers, while only 15% consume red meat daily. The study's authors suggest that tea may be worse for one's health than red meat. Do you agree with the authors, and if not, what additional evidence would you like to see to evaluate their claim? Explain.

Problem 16 A education study finds that Kentucky children who travel abroad at least once prior to turning 18 have higher high school graduation rates than children who have not been abroad by the time they turn 18. The study concludes that traveling abroad "instills a sense of wonder and respect for the greater world, and so incentivizes students to work even harder to earn their place in it." What additional evidence would you like to see to evaluate the study's claim, and why?

Problem 17 An economics student suspects that the return to education[1] is higher the wealthier his or her family is. Propose a regression model that would test this hypothesis (you can assume the existence of any data that you like). Explain what result would cause you to conclude that the hypothesis is true, and what result would cause you to conclude the opposite. Be clear.

[1] "return" = marginal benefit of one additional year of education, measured in labor market salary.