Midterm II - Answers

Tomas Dvorak

Econ 253 Spring 2001

You have 75 minutes to answer the following 5 questions. Each question is worth 20 points, so you should plan to spend an equal amount of time on each one. You may use a calculator and a single-sided sheet of paper with notes and formulas. Any collaboration is considered a violation of the honor code, as is discussing the exam with students in other sections who have not yet taken the exam. Good luck!

1. The Good Food Company is interested in the health and fitness of its employees. The company recently sponsored an exercise session before the start of each shift and wishes to see if the program resulted in weight loss for its employees. The company surveyed 200 of its 100,000 employees before and after the program and found the following results:

before	after
$n_1 = 200$	$n_2 = 200$
$\overline{x}_1 = 180$	$\overline{x}_1 = 175$
$s_1 = 50$	$s_1 = 30$

Answer the following questions:

(a) Did the program have a statistically significant effect on the mean weight of employees?

Answer: Let's assume that *before* the program, employee weight was distributed $X_1 \sim N(\mu_1, \sigma_1^2)$, and after the program $X_2 \sim N(\mu_2, \sigma_2^2)$. To see if the program had any effect we test the following hypothesis:

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

This is the difference in means test. The test statistic $Z = \frac{180-175}{\sqrt{50^2/200+30^2/200}} = 1.21$ is distributed standard normal. The critical region at the 5% level of significance is $(-\infty, -1.96)$ and $(1.96, \infty)$. The test statistic falls in the acceptance region, thus we can not reject H_0 and we conclude that the program had no statistical effect on the mean weight of employees.

(b) Did the program significantly decrease the differences in weight among employees?

Answer:

$H_0 : \sigma_1 = \sigma_2$

$H_1 : \sigma_1 > \sigma_2$

This is a test for a comparison of two population variances. The test statistic $F = \frac{s_1}{s_2} = \frac{50}{30} = 1.66$ is distributed $F_{199,199}$. The critical region for this one-tailed test at the 5% level of significance is $(1.26, \infty)$. The test statistic falls in the critical region and we reject the null hypothesis and conclude that the program decreased the differences in weight among workers in the company.

2. Suppose a pharmaceutical company claims to have found a cure for cancer. When the FDA tests the drug, its null hypothesis is that the drug is ineffective in curing the disease. Answer the following questions:

(a) What is the type I error in this example?

Answer: A type I error occurs when we reject the null hypothesis when it is true. In this case, the type I error would occur if an ineffective drug is accepted as effective.

(b) What is the type II error?

Answer: A type II error occurs when we accept the null hypothesis when it is false. In this case it will occur if we reject a drug that is effective in curing cancer.

(c) How would you weigh the trade off between type I and type II errors in this case? How would you set the significance level?

Answer: This question requires judgment. Clearly, in this case committing a type II error is very serious as it deprives patients of available treatment. Committing a type I error means that some patients take a drug that is ineffective. In practice, we would maximize the probability of accepting the drug, i.e. we would set the significance level (probability of type I error) at a very high level (α very high).

3. For the simple linear model $Y_t = B_0 + B_1 X_t + u_t$, show that under certain assumptions, OLS estimates give unbiased predictions. That is, show that $E(\widehat{Y}_t) = E(Y_t)$. Be sure to state the assumptions essential to your answer.

Answer: First we find $E(Y_t) = E(B_0 + B_1X_t + u_t) = B_0 + B_1X_t$ by assumption 1, that x_t is nonstochastic and assumption 2 that $E(u_t) = 0$. Next we find $E(\widehat{Y}_t) = E(b_0 + b_1X_t) = E(b_0) + E(b_1X_t) = B_0 + B_1X_t$, by the fact that OLS estimators b_0 and b_1 are unbiased. Hence $E(\widehat{Y}_t) = E(Y_t)$ and OLS estimates do give us unbiased predictions.

4. The Office of the Registrar at a college took a random sample of 500 students and obtained their college GPA's ($COLGPA$), high school GPA's ($HSGPA$) and their combined SAT scores (SAT). The following model was estimated:

$$COLGPA = B_1 + B_2HSGPA + B_3SAT + u \tag{1}$$

The estimated slope coefficients and their standard errors are given as follows:

Coefficient	Standard Error
b_2=0.398	0.061
b_3=0.0007	0.0003
$ESS = 288$	
$RSS = 1089$	

Answer the following questions:

(a) Test the hypothesis that the regression coefficient $B_2 = 0$ at the 5% level of significance. Interpret your result and the estimated coefficient.

Answer: We know that under the null hypothesis $(H_0 : B_2 = 0))$ the test statistic $t = \frac{b_2}{\sigma_{b_2}} = 6.5$ is distributed t_{500-3}. The critical region is $(-\infty, -2)$ and $(2, \infty)$. Hence, we reject the null hypothesis that high school GPA has no influence on college GPA. The interpretation of the estimated coefficient is the following: holding SAT scores constant, a point increase in high school GPA would increase expected college GPA by 0.398 points.

(b) Test the hypothesis that the regression coefficient $B_3 = 0$ at the 5% level of significance. Interpret your result and the estimated coefficient.

Answer: In this case the test statistic is $t = 2.3$ which is in the critical region. We reject the hypothesis that SAT scores have no effect on college GPA. The interpretation of the estimated coefficient is the following: holding high school GPA constant, a point increase in SAT scores increases expected college GPA by 0.007 points.

(c) Calculate R^2.

Answer: $R^2 = \frac{ESS}{TSS} = \frac{288}{288+1089} = 20.9$

(d) Test the joint hypothesis that $HSGPA$ and SAT do not explain college GPA.

Answer: We know that under the null hypothesis, test statistic $F = \frac{ESS/k-1}{RSS/n-k} = \frac{288/2}{1089/497} = 65.7$ is distributed F with 2 and 497 degrees of freedom. The 5% critical region is $(3, \infty)$, thus we reject the null hypothesis that high school GPA and SAT scores do not explain college GPAs.

(e) Suppose a student decided to study hard to improve her SAT scores by 100 points. Can you tell how much would this increase her expected college GPA? (Hint: Think before jumping into the mathematics.)

Answer: If a student increased her SAT scores by 100 *and* her high school GPA stayed constant, her expected college GPA would increase by 0.7 points. However, it is likely that studying hard would also improve her high school GPA, making the impact of studying greater than 0.7.

5. Consider the following regression model of demand for bus travel. The data used in this study includes the following information for 40 cities across the U.S. in 1988:

- BUS = Demand for urban tranportation by bus in thousands of passenger hours
- FARE = Bus fare in dollars
- INCOME = Average income PER CAPITA
- POP = Population of city in thousands (Range 167 - 7323.3)

The following regression was estimated:

$$ln(BUS) = B_1 + B_2 ln(FARE) + B_3 ln(INCOME) + B_4 POP + u \qquad (2)$$

```
 regress LNBUS  LNFARE LNINCOME POP
  Source |       SS       df       MS                  Number of obs =      40
---------+------------------------------               F(  3,    36) =    8.27
   Model |  21.3257748     3   7.1085916               Prob > F      =  0.0003
Residual |  30.9306049    36  .859183469
R-squared     =  0.4081
---------+------------------------------               Adj R-squared =  0.3588
   Total |  52.2563797    39  1.33990717               Root MSE      =  .92692

------------------------------------------------------------------------------
   LNBUS |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  LNFARE |  -.0000508    .480407      0.000   1.000      -.9743614    .9742597
LNINCOME |  -3.272351   1.248092     -2.622   0.013
-5.803599   -.7411026
     POP |   .0006137   .0001263      4.860   0.000       .0003576    .0008698
   _cons |   38.26334   12.09056      3.165   0.003       13.74255    62.78414
------------------------------------------------------------------------------
```

Using the above computer output, answer the following questions:

(a) Which coefficients appear as significant determinants of the demand for bus travel? Are the signs of estimated coefficients consistent with your intuition? Explain.

Answer: Only income and population appear significant in the regression. Perhaps surprisingly, fare did not significantly influence the demand for bus travel. The sign on fare is what I would expect: the higher the fare, the lower the demand. The higher the income, the lower the demand. This is consistent with bus travel being an inferior good. Finally, the higher the population, the greater the demand for bus travel.

(b) Is the regression jointly significant at 1% level of significance?

Answer: Yes, from the p-value of the F-test we see that the regression is jointly significant.

(c) What is the interpretation of the coefficient on income?

Answer: Since both the dependent and independent variables are in logarithms, the coefficient says that if income increases by 1% the demand for bus travel will decrease by about 3%.

(d) Test the hypothesis that the demand for bus travel is unitary elastic (in absolute value) with respect to income, against the alternative that the demand is elastic. (Note: demand is considered elastic if the elasticity (in absolute value) is greater than 1 and inelastic if it is less than one.)

Answer: $H_0 : |B_3| = 1, H_1 : |B_3| > 1$ We know that under the null hypothesis $t = \frac{|b_3|-1}{\widehat{\sigma_{b_3}}} = \frac{2.27}{1.24} = 1.83$ is distributed t_{36}. The 5% critical region for this one-tailed test is $(1.684, \infty)$. Thus, we accept the hypothesis that demand for bus travel is elastic with respect to income.

(e) What is the marginal effect of an increase in population on demand for bus travel? Explain your answer.

Answer: The marginal effect of an increase in population $\frac{\Delta BUS}{\Delta POP} = B_3 BUS$. Hence, the marginal effect of an increase in population depends on the level of bus travel. The higher the existing demand for bus travel, the greater the effect of an increase in population.