Exam Three Answers

Econ 180: Quantitative Methods
Professor Robert J. Lemke
Spring 2006

Directions: A bluebook has been provided for your answers. In order to facilitate anonymous grading, put your ID number (not your name) on your bluebook. There are four questions to the exam for a total of 100 points, weighted as indicated.

1. (15 points) A sociology student is interested in knowing if student attitudes toward drinking vary by sex. The student randomly asks 87 female students and 102 male students if they plan on binge drinking sometime in the next 10 days. Of those asked, 17 female students and 27 male students claimed to have such plans. Test the null hypothesis that female and male students plan to binge drink sometime in the next 10 days at the same rate against the alternative hypothesis that rates differ by sex by calculating the *p*-value. What is the conclusion you would draw from the data?

   Answer. The difference in sample proportions is [27 ÷ 102] – [17 ÷ 87] = 0.0693. The sample standard deviation in the difference in means is

   $$s = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{0.001908 + 0.001807} = 0.06095\,.$$

   Thus, as the null hypothesis is that there is no difference in the proportion of binge drinkers across sexes, the test statistic is 0.0693 ÷ 0.06095 ≈ 1.14. Finally, the *p*-value = 2 * [0.5 – 0.3729] = 0.2542. Thus, the *p*-value = 25.42%, and therefore the data fail to reject the claim that the rate of binge drinking is difference across the sexes.

2. (15 points) An economics student is interested in knowing if the average wage of full-time male construction workers is different between eastern German states (the former East Germany) and western German states (the former West Germany). The student collects a random sample of the annual wage earned by full-time male construction workers in the east and the west. In the east, 1,219 workers are surveyed. Their average wage is €17535 (€ is the Euro symbol) with a standard deviation of €8,213. In the west, 1,856 workers are surveyed. Their average wage is €18,333 with a standard deviation of €12,838. Test the null hypothesis that average wages are the same against the alternative hypothesis that average wages are different at the 5 percent significance level.

   Answer. The difference in sample means is 18333 – 17535 = 798. The sample standard deviation in the difference in means is

   $$s = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(8213)^2}{1219} + \frac{(12838)^2}{1856}} = \sqrt{55335 + 88801} = 380.$$

   Thus, as the null hypothesis is that there is no difference in average wages across the two locations, the test statistic is 798 ÷ 380 = 2.10. Given this test statistic and as the alternative hypothesis is two-sided, we have that the critical value is $Z_{\alpha/2} = Z_{0.025} = 1.96$. Thus, as the test statistic exceeds the critical value, the data reject the hypothesis that there is no difference in average salary in favor of the alternative hypothesis that there is a difference in average salaries across the two locations.

3. (35 points) The Alumni Office is interested in knowing the relationship between income and giving back to the College by alumni. Thus, it estimates the following single-variable regression equation:

$$Giving = \beta_0 + \beta_1 \cdot Income + \varepsilon,$$

where *Giving* and *Income* are both reported in dollars.

a. Before estimating the parameters, what sign do you expect to find for β_1? Why?

   Answer. I would expect the sign on β_1 to be positive as I would expect richer alumni to give more money back to the College.

b. Suppose the estimated equation (with standard errors in parentheses) turns out to be:

$$\begin{array}{c} Giving = -12.2 + 0.0023 Income. \\ (23.1) \quad (0.0011) \end{array}$$

   Interpret the estimated coefficient on *Income*. What is the *t*-statistic associated with the coefficient on income? What is the *p*-value associated with the estimated coefficient on income? Is income a statistically significant predictor of giving back to the College?

   Answer. For every \$1 increase in income of an alum, the alum is expected to give \$0.0023 more to the College. For a better quantification, alums are predicted to give the College 10,000 x 0.0023 = \$23 more for every additional \$10,000 in income. The *t*-statistic associated with the coefficient on income is 0.0023 ÷ 0.0011 = 2.09. The *p*-value associated with the estimated coefficient on income, therefore, is 2 * [0.5 – 4817] = 0.0366 as *p*-values of estimated coefficients are always produced assuming a two-sided test. Finally, as the *p*-value on income is 3.66%, most researchers would conclude that income is a statistically significant predictor of giving back to the College.

c. One particular alum has an income of \$80,000 and gives \$500 back to the College. What is the error term associated with this alum?

   Answer. The error term = actual – predicted = 500 – (-12.2 + 0.0023*80000) = 500 – 171.8 = \$328.20.

d. What would have been the estimated regression equation (including standard errors) if the Alumni Office had regressed *Giving* on *IncomeK* where *IncomeK* = *Income* / 1000? How should one interpret the estimated coefficient on *IncomeK*?

   Answer. With this adjustment in income variable, the new estimated regression would be:

$$\begin{array}{c} Giving = -12.2 + 2.3 IncomeK. \\ (23.1) \; (1.1) \end{array}$$

   Notice that this new estimated regression has the same interpretation as before. In particular, a \$10,000 increase in income, which is an increase in *IncomeK* of 10 units, is predicted to be associated with an increased gift of \$23.

4. (35 points) McDonalds wants to investigate the determinants of revenue for its restaurants. To do this, McDonald's collects restaurant-level data on 540 restaurants. The codebook, descriptive statistics, and regression results are listed below. For the record, the US Census Bureau classifies all states as being either in the east, midwest, south, or west. There are no other regions, and no state is in more than one region.

   a. Interpret the coefficient estimate on Population.

      Answer. Every additional person residing in the city in which the restaurant is located is expected to increase weekly revenues by $1.343.

   b. Are the *p*-values listed in the final column for a one- or two-sided test? How do you know?

      Answer. The *p*-values for regression output are always given for two-sided tests. This can be checked by looking at the *t*-statistic on Franchise, which is 2.42. The area to the right of 2.42 for a normal(0,1) variable is 0.5 – 0.4922 = 0.0078. Doubling this, which is assuming a two-sided test, yields a *p*-value of 0.0156.

   c. Interpret the coefficient estimate on Franchise? Is being a franchise restaurant a statistically significant predictor of a restaurant's weekly revenue?

      Answer. A restaurant that is operated by a franchiser is expected to have weekly revenues that are $31,249 higher than a comparable restaurant operated by McDonald's directly (i.e., by not a franchiser).

   d. Interpret the coefficient estimate on East? Comment on statistical significance?

      Answer. A restaurant that is located in the east is expected to earn weekly revenues that are $44,329 higher than a comparable restaurant located in the midwest. The *p*-value of the estimated coefficient on East is essentially zero, and therefore the effect of being in the east is statistically different from zero but this is known to be true only when emphasizing that it is the effect of being in the east compared to being in the midwest. It is unclear from these results, for example, that being in the east is statistically different from being in the west.

   e. What is the *t*-statistic associated with the estimated coefficient on AvgWage?

      Answer. The *t*-statistic associated with the estimated coefficient on AvgWage = 23453 ÷ 18923 = 1.24.

   f. What is the *p*-value associated with the estimated coefficient on South?

      Answer. The *p*-value associated with the estimated coefficient on South = 2 * [0.5 – 0.4726] = 0.0548.

   g. What is the Adjusted R-squared?

      Answer. The adjusted R-squared = $1 - (1 - r^2)([n-1]/[n-k-1]) = 1 - (0.6707)(539/532) =$ 0.3205

Codebook:

Revenue:	Weekly revenue of the restaurant.
Population:	Population of the city in which the restaurant resides.
AvgWage:	Average hourly wage paid by the restaurant.
Interstate:	Dummy variable equals 1 if located in an exit off of an interstate highway; equals 0 if not.
Franchise:	Dummy variable equals 1 if the restaurant is run by a franchise holder (i.e., not by McDonald's corporation directly); equals 0 otherwise (i.e., is run by McDonald's corporation directly).
Mall:	Dummy variable equals 1 if the restaurant is located in a mall; equals 0 otherwise.
East:	Dummy variable equals 1 if located in the eastern region of the United States; = 0 otherwise.
Midwest:	Dummy variable equals 1 if located in the midwestern region of the United States; = 0 otherwise.
South:	Dummy variable equals 1 if located in the southern region of the United States; = 0 otherwise.
West:	Dummy variable equals 1 if located in the western region of the United States; = 0 otherwise.

Descriptive Statistics (*N* = 540):

Variable	Mean	St. Dev.	Min.	Max.
Revenue	$157,345	$22,458	$95,675	$234,003
Population	23,435	12,378	4,034	1,234,493
AvgWage	$7.36	$0.57	$6.29	$9.12
Franchise	0.846	0.1303	0	1
Mall	0.092	0.0835	0	1
East	0.278	0.2007	0	1
Midwest	0.211	0.1665	0	1
South	0.194	0.1564	0	1
West	0.317	0.2165	0	1

Regression Results: Dependent Variable = Revenue

	Coef. Est.	Stnd. Error	*t*-Statistic	*p*-value
Population	1.343	0.329	4.08	0.0000
AvgWage	23,453	18,923		0.2150
Franchise	31,249	12,913	2.42	0.0155
Mall	8,922	6,321	1.41	0.1585
East	44,329	10,231	4.33	0.0000
South	-12,344	6,435	1.92	
West	2,382	3,492	0.68	0.4965
Constant	78,423	23,322		
R-squared	0.3293			
Number of obs	540			