BUEC 333, Test 2

July 8, 2015, 14:30-17:20

Questions

1. **Chapter 4.** The following questions are about regression analysis.

   (a) In a regression equation, what does the error term capture? **Other factors, not captured by X, that influence Y.**

   (b) What is the difference between an error term and a residual? **The error term is a population quantity. It is the distance from (Yi,Xi) to the population regression line. The residual is a sample quantity. It is the distance from (Yi,Xi) to the sample regression line.**

   (c) Using the population regression equation *or* the estimated regression equation, write down *precisely* what the OLS estimator minimizes. Your answer should include "error terms" or "residuals", as well as a precise mathematical expression for what is being minimized. **It minimizes the sum of squares of residuals, $\sum_i (Y_i - b_0 - b_1 X_i)^2$.**

   (d) Is the OLS estimator $\hat{\beta}_1$ a random variable? Explain! **Yes, it is a function of the random sample $(Y_i, X_i), i = 1, \cdots, n$.**

   (e) Least squares Assumption 1 is: $E(u_i | X_i) = 0$. What is the interpretation of β_1 if Assumption 1 holds? Be precise. **It is the ceteris paribus change in Yi if Xi changes by one unit, if nothing else changes (ceteris paribus).**

   (f) Does β_1 have an interpretation if Assumption 1 does not hold? Elaborate. **Make sure to include the word prediction. The regression is still useful for prediction, and β_1 can be said to have an interpretation in that regard. Whether you said YES or NO depends on your explanation.**

   (g) A linear regression yields $\hat{\beta}_1 = 0$. Show that $R^2 = 0$. **Note that ESS=$\hat{\beta}_1^2 \sum_i \left(X_i - \bar{X}\right)^2$. The result follows immediately.**

2. **Chapter 4.** Consider the linear regression model, with equation $Y_i = \beta_0 + \beta_1 X_i + u_i$. You are given the following random sample of size $n = 3$:

X_i	Y_i	i
10	6	1
9	3	2
11	3	3

   (a) For this sample, Compute $\hat{\beta}_1$, the OLS estimator for the slope β_1. **0**

   (b) For this sample, Compute $\hat{\beta}_0$, the OLS estimator for the intercept β_0. **Ybar - beta1-hat * Xbar = 4-0*Xbar=4**

(c) Compute $\sum_{i=1}^{3} \hat{u}_i$. Compute $\sum_{i=1}^{3} X_i \hat{u}_i$. Is this what you would expect? Why? **0,0. This is an algebraic fact of the OLS estimator.**

(d) Compute the R^2. Interpret the number you obtain. **0! See 1(g). It means that 0% of the variation in Y is explained by X.**

3. **Chapter 4.** A regression of weekly average earnings (AWE, in \$) on age (measured in years) using a random sample of college-educated full-time workers aged 25-65 yields the following:

$$\widehat{AWE} = 696.7 + 9.6 \times Age,\; R^2 = 0.023,\; SER = 624.1$$

(a) Explain what the coefficient values 696.7 and 9.6 mean. AWE **is expected to increase by** \$9.6 **for each additional year of age. 696.7 is the expected** AWE **for someone whose** $Age = 0$**; otherwise it has no meaningful interpretation.**

(b) What are the units of measurement for the SER? What are the units of measurement for the R^2? SER **is measured in dollars per week.** R^2 **is unit free.**

(c) What does this regression predict will be the earnings for a 30-year old worker? For a 40 year old worker? **For** 30**-year-old:** $696.7 + 9.6 \times 30 = \$984.7$**. For** 40**-year-old:** $696.7 + 9.6 \times 40 = \$1080.7$**.**

(d) The average age in this sample is 41.6 years. What is the average value of AWE in the sample? $696.7 + 9.6 \times 41.6 = 1096.06$

4. **Chapter 5.** Consider the R output from the code that we ran in class in Figure [see exam], where we investigated the effect of class-size on student performance.

(a) Use the output to construct a 95% confidence interval for β_1. Only go as far as you can without a calculator. **-2.28 +- 1.96*0.48.**

(b) Interpret the interval you obtained under (a). **If I were to construct such intervals based on many, many different random samples, I would capture the true value of the regression coefficient in 95% of the cases. ALTernatively, use the terms "coverage probability".**

(c) Use the reported **p-value** to test the null hypothesis $H_0 : \beta_1 = 0$. Would you reject the null hypothesis at the 5% level? **Yes, reported p-value by R is 2.8*e-06.**

(d) What is the interpretation of $\hat{\beta}_1$? Make sure that your answer is specific to this example. You must use the word "class size". **For a given school, if we change class size (str) by one, and do not change anything else, we expect student performance to go up by beta1-hat.**

(e) Do you think that Assumption 1 holds (see question 1e)? Why/why not? **This is an open question, but I am looking for an answer that involves omitted variables. There are other factors that influence class-size that are correlated with str. See chapter 6 for a reasonable answer.**

5. **Chapter 14.** Suppose that Y_t follows the stationary AR(1) model

$$Y_t = 2.5 + \beta_1 Y_{t-1} + u_t$$

where u_t is i.i.d. with $E(u_i) = 0$ and $\text{var}(u_i) = \sigma^2$.

(a) Express $E(Y_t)$ in terms of β_1. **2.5/(1-b1)**

(b) What is $\text{var}(Y_t)$? **sigma2 / (1-b1^2).**

(c) Assume that $\beta_1 = 1$ and $\sigma^2 = 0$. Assume that $Y_1 = 2.5$. Can you compute Y_2? Once you have Y_2, can you compute Y_3? What about Y_{10}? **Yes. Y1=2.5, Y2=5, Y3=7.5, Y10=25.**

(d) Forget the additional assumptions you made under (c). Look at your answers under (a) and (b). What happens to them when $\beta_1 = 1$? Is the sequence Y_t still stationary when $\beta_1 = 1$? **Expectation and variance go to infinity. At beta1=1, they are undefined. The series is no longer stationary, as the mean and variance increases with time.**

6. **Chapter 14.** Consider the time series $(Y_1, Y_2, Y_3) = (-1, 0, 1)$.

(a) Compute the first sample autocovariance. **1/6.**

(b) Compute $\hat{\rho}_1$. **1/6.**