Midterm BUEC 333

June 19, 2014, 12:30-14:20

Each question earns 1 point, and only if it is accompanied by a correct explanation. If the explanation is right, but the answer is wrong, you can still give 1 point, or you can give somewhere between 0.5 and 1 points.

1 Probability

You have a random sample $(X_1, X_2, \cdots, X_n)$ of measurements of X. The random variable X has mean $\mu_x = E(X)$ and variance $\sigma_x^2 = Var(X)$. The sample average is $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

1. Which one of the following three statements about $\bar{X}$ and μ_x is true?

   (a) $\bar{X} = \mu_x$.

   (b) $Var(X_1) = Var(\mu_x)$.

   (c) $E(\bar{X}) = E(X_1) = \mu_x$. **<- correct answer. E(X1)=muX because the draws are identically distributed. Showing that E[barX]=1/n \sum E[X_i] = 1/n * n * muX is not entirely necessary, but some kind of explanation is required.**

2. Show that $Var\left(\bar{X}\right) = \sigma_x^2/n$. Carefully state what you are doing at each step of the derivation. **Var(barX) = 1/ (n^2) \sum Var(Xi) + "some covariance term" = 1/ (n^2) \sum Var(Xi) = 1/ (n^2) * n * Var(Xi) = 1/n Var(Xi)**

Consider the joint probability distribution presented in Table 1.

3. For each of the following statements (a)-(h), state whether the statement is true. No explanation necessary. Correct answer: 0.25. No answer: 0. Incorrect: -0.25 points.

   (a) $P(X_2 = 1|X_1 = 3) = 1$ **true**

   (b) $P(X_2 = 1|X_1 = 2) = 0.4$ **true**

   (c) $P(X_2 = 3|X_1 = 2) = 0.1$ **false**

   (d) X_1 and X_2 are uncorrelated **false**

		X_1		
		1	2	3
X_2	1	0.1	0.2	0.25
	2	0.1	0.2	0
	3	0.05	0.1	0

Table 1: Joint probability distribution.

(e) X_1 and X_2 are independent **false**

(f) $E(X_1) > 1$ **true**

(g) $P(X_2 = 2) = 0.1$ **false**

(h) $P(X_1 \leq 2) = 0.75$ **true**

2 Statistics

Assume that you have a random sample $(2, 3, 4)$ for a RV X with mean μ_x and variance σ_x^2.

4. For this sample, compute $\bar{X}$. **3. Some use of the formula is required, not just "3"**

5. For this sample, compute the sample variance. **(2-3)^2+(3-3)^2+(4-3)^2 / (n-1) = 1.**

6. Construct a 90%-CI for μ_x based on the sample mean and sample variance. **[3-1.64*SE(Xbar), 3+1.64*SE(Xbar)] = [3-1.64/sqrt(3), 3+1.64/sqrt(3)]**

7. Interpret the numbers in your answer to question 6. **Incorrect answer: "contains μ_X with 90% probability". Correct: "If we would repeat this procedure many, many times, it would contain μ_X ~90% of the time.". Reasonable answers can be accepted, as long as they are NOT of the incorrect answer form.**

3 Regression analysis

8. This question is about the error term, u_i, in regression analysis.

   (a) Why do we include an error term in our regression equation? **possible correct answer: to capture other factors that influence Y. or: measurement error, nonlinearities, etc.**

   (b) What is the difference between an error term and a residual? **possible correct answer: one is a population quantity, one is a sample quantity. another correct answers": one measures distance from point to population regression line, the other to the estimated (sample) regression line. the answer should definitely contain the words "population" and "sample".**

9. 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. **Minimizes the sum of squared (vertical) distance of the points in the scatterplot to the estimated regression line, so $min \sum_{i=1}^{n} \hat{u}_i^2 =$.**

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 $n = 3$:

10. For this sample, Compute $\hat{\beta}_1$, the OLS estimator for the intercept β_1. **0.**

11. For this sample, Compute $\hat{\beta}_0$, the OLS estimator for the intercept β_0. If you did not manage to the previous question, assume that $\hat{\beta}_1 = 0.5$. **barY=beta0hat=4.**

X_i	Y_i	i
2	6	1
1	3	2
3	3	3

12. Compute the R^2. Interpret the number you obtain. **1, since the line is flat and X does not explain Y at all.**