BUEC 333-D200, Test 1

June 15, 2016, 14:30-17:20

Caution

- Allowed on desk:
  - pen,
  - SFU ID,
  - water bottle without a label
- Not allowed: **anything** else. For example,
  - no pencil cases,
  - erasers,
  - pencils,
  - non-graphical calculator,
  - phone, ruler, whiteout, food, bottles with labels, ...
  - Penalty: **0% on this test.**
- Bathroom breaks:
  - No bathroom breaks in the first 30 minutes
  - No bathroom breaks after the first person hands in the test
- No student will be permitted to leave during the **last 15 minutes.**
- Once you **finish** this exam, **signal** it to us, and we will come to pick up your exam. Stay seated until somebody comes to collect your exam.
- We will **not answer questions about the exam**: if you think a question is unclear, or contains a mistake:
  - make a note about it in your answer to the question
  - answer the question as best as you can
  - notify us after you hand in the exam.

Instructions

- On the front page of **your answer sheet**, write (i) your name; (ii) your student ID.
- On the front page of **this document**, write: (i) your name; and (ii) your student ID.
- No explanation = no points. A correct answer **with correct explanation** earns 1 point for each subquestion.
- For a "compute" question, an explanation can consist of starting from **an appropriate formula**, and working towards the correct numerical answer.

1 Week 1: Probability theory

Consider the joint probability distribution of the random variables X_1 and X_2 given in Table 1.

	$X_2 = 1$	$X_2 = 2$	$X_2 = 3$
$X_1 = 0$	0.2	0.1	0.1
$X_1 = 1$	0.3	0.3	0

Table 1: Joint probability distribution of X_1 and X_2.

1. What is the probability that $X_1 = 0$ given that $X_2 = 2$?
2. What is the conditional expectation of X_1 given that $X_2 = 2$?
3. Compute $\text{Var}\,(X_1 | X_2 = 3)$.
4. Compute the covariance between X_1 and X_2. Get as far as you can without a calculator.

2 Week 2: Sampling

Suppose that $Y_1, \cdots, Y_n$ are independent random variables with a common mean μ_Y, but with different variances:

$$\begin{gathered}\text{Var}\,(Y_1) = \sigma_1^2 \\ \text{Var}\,(Y_2) = \sigma_2^2 \\ \vdots \\ \text{Var}\,(Y_n) = \sigma_n^2\end{gathered}$$

1. Are the random variables $(Y_1, \cdots, Y_n)$ i.i.d.?
2. Suppose that $n = 3$. Show that $E\left(\bar{Y}\right) = \mu_Y$.
3. Suppose that $n = 2$. What is the variance of the sample mean $\bar{Y}$?
4. For qeustions 4 and 5, assume that $n = 2$ and that $\sigma_1^2 = 1$ and that $\sigma_2^2 = 100$.
   Consider the estimator $\tilde{Y} = Y_1$. Is $\tilde{Y}$ unbiased?
5. Show that $\tilde{Y}$ is more efficient than $\bar{Y}$.

3 Week 3: Statistics

For this question, remember that a Bernoulli random variable Y has $\text{Var}\,(Y) = p\,(1-p)$, where $p = P\,(Y = 1)$. Therefore, given an estimator $\hat{p}$ for p, the estimator for the variance is $s_Y = \hat{p}\,(1 - \hat{p})$.

"Would you use the library more if the hours were extended?" In a random sample of 100 SFU freshmen, a sample proportion of $\hat{p}_1 = 0.5$ answered that they would use the library more. A random sample of 100 SFU sophomores was also obtained. Out of them, 80 responded that they would use the library more, for a sample proportion of $\hat{p}_2 = 0.8$.

1. Denote by p_1 the population proportion of SFU freshmen that would use the library more if hours were extended. Construct a 95% confidence interval for p_1. (As far as you can get without a calculator.)
2. Test the null hypothesis $H_0 : p_1 = 0.10$ against the two-sided alternative $H_1 : p_1 \neq 0.10$. (Answer this question without computing the $p-$value: you do not have access to the book's Table 1!)
3. What is the probability that p_1, the true population proportion, is in the interval computed in (1)? If you did not answer the previous question, assume that the correct answer is $[0.4, 0.6]$.
4. How would you test whether there is a statistically significant difference between freshmen and sophomores in their response?

4 Week 4: Linear regression (mechanics)

You are given the following random sample:

i	X_1	Y_i
1	0	8
2	2	4
3	1	6

Table 2: Sample for week 4 question.

For this sample:

1. Compute $\hat{\beta}_1$.
2. Compute the sum of squared residuals.
3. Without changing the values of X_1, X_2, X_3 and Y_1, give values for Y_2, Y_3 that give $R^2 = 0$.

5 Week 5: Linear regression (statistics)

The questions is about the results of a regression of the weight of infants (newborn babies) and the smoking behavior of the mother. We have data on $n = 1388$ pregnancies. The dependent variable is *bwght*, the weight of the baby when it is born, in ounces. The explanatory variable is *cigs*, the number of cigarettes smoked per day by the mother. We obtain the following results:

$$\widehat{bwght} = \underset{(20.6)}{119.77} - \underset{(0.1)}{0.5} \times cigs\,.$$

Answer the following questions:

1. The underlying model is $bwght_i = \beta_0 + \beta_1 \times cigs_i + u_i$. Give two examples of factors that are captured by u_i. Use factors that are specific to this example, so "it is an error term" or "other stuff" are not correct answers.
2. State Least Squares Assumption 1 using either math or words.
3. Construct a 95% confidence interval for β_1.
4. What is the interpretation of β_1, assuming that Assumption 1 holds?
5. What is the interpretation of β_1, assuming that Assumption 1 does not hold?

6 Self-study

1. [Section 4.5] Under the Least Squares Assumptions, what is the expectation of $\hat{\beta}_1$? Is $\hat{\beta}_1$ consistent?
2. What is a cumulative probability distribution? Use math and words.

A Cheat sheet

Key Concept 2.3

Let X, Y and V be random variables, let μ_x and σ_x^2 be the mean and variance of X, let σ_{XY} be the covariance between X and Y (and so forth for the other variables), and let a, b, and c be constants. The following facts follow from the definitions of the mean, variance, and covariance:

$$
\begin{aligned}
E(a+bX+cY) &= a+b\mu_X+c\mu_Y, && (1)\\
\operatorname{Var}(aX+bY) &= a^2\sigma_X^2+2ab\sigma_{XY}+b^2\sigma_Y^2, && (2)\\
E\left(Y^2\right) &= \sigma_Y^2+\mu_Y^2, && (3)\\
\operatorname{Cov}(a+bX+cV,Y) &= b\sigma_{XY}+c\sigma_{VY}, && (4)\\
E(XY) &= \sigma_{XY}+\mu_X\mu_Y, \text{ and} && (5)
\end{aligned}
$$

Other formulas

1. The OLS estimators:

$$
\hat{\beta}_0 = \bar{Y}-\hat{\beta}_1\bar{X}
$$

$$
\hat{\beta}_1 = \frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}
$$

2. The standard error of the regression:

$$
SER = \sqrt{SSR/(n-2)}
$$

3. Critical values for confidence intervals
   (a) 90%: 1.64
   (b) 95%: 1.96
   (c) 99%: 2.58