BUEC 333-D100, Test 1

June 13, 2016, 11:30-14:20

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, food, bottles with labels
- No bathroom breaks after the first person has handed in this 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.

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

Imagine the following game. There are two dice. One of them is fair, and one of them is unfair. First, you randomly select a dice. Second, you throw that dice. Let the random variable X be the number of eyes that you throw, and let the random variable Y denote which die you picked in the first step (0: fair dice, 1: unfair dice). Then, we have the following joint probability distribution:

		Y : which die	
		Fair ($Y=0$)	Unfair ($Y=1$)
Number of eyes, X	$X=1$	1/12	1/20
	$X=2$	1/12	1/20
	$X=3$	1/12	1/20
	$X=4$	1/12	1/20
	$X=5$	1/12	1/20
	$X=6$	1/12	1/4

Table 1: Two-step game

Answer the following questions:

1. What is the probability of choosing the fair dice? What is the probability of throwing a 3?
2. Given that you have selected the unfair dice, what is the probability of throwing a 6?
3. Given that you throw a 4, what is the probability that you selected the fair dice?
4. Compute $E(X|Y=1)$ and $E(X|Y=0)$.
5. Using the answer under (4), and the **law of iterated expectations**, compute $E(X)$.

2 Week 2: Sampling

Suppose that $Y_1, \cdots, Y_n$ are random variables with a common mean μ_Y, a common variance σ_y^2 and the same correlation ρ (so that the correlation between Y_i and Y_j is equal to ρ for all i and j where $i \neq j$).

1. Show that $\text{cov}(Y_i, Y_j) = \rho\sigma_y^2$ for $i \neq j$.
2. Suppose that $n = 2$. Show that $E(\bar{Y}) = \mu_Y$.
3. Suppose that $n = 2$. Show that $\text{Var}(\bar{Y}) = \frac{1}{2}\sigma_Y^2 + \frac{1}{2}\rho\sigma_Y^2$.
4. Is $\bar{Y}$ unbiased? (Reminder: justify your answer!)

3 Week 3: Statistics

In a random sample of 100 male and 100 female bus drivers, a variable *salary* is measured (in Canadian dollars). You find that the sample average salary for men is 2700, while the sample average salary for women is 2000. Furthermore, you compute the sample standard deviation to be 100 for men and 100 for women. Using these results, answer the following questions:

1. Denote by μ_{male} the expected salary of a male bus driver. Construct a 95% confidence interval for μ_{male}. (As far as you can get without a calculator.)
2. Test the null hypothesis $H_0 : \mu_{male} = 2000$ against the two-sided alternative $H_1 : \mu_{male} \neq 2000$. (Answer this question without computing the $p-$value: you do not have access to Table 1!)
3. Denote by μ_{female} the expected salary of a female bus driver. Construct a 95% confidence interval for $\mu_{male} - \mu_{female}$. Go as far as you can without a calculator.
4. What is the probability that $\mu_{male} - \mu_{female}$ is in the interval computed in (3)? If you did not answer the previous question, assume that the correct answer is $[781, 967]$.

4 Week 4: Linear regression (mechanics)

You are given the following random sample:

i	X_1	Y_i
1	0	4
2	2	8
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. What happens to the R^2 when you change Y_2 from 8 to 6? You can answer this question without repeating steps 1-2.

5 Week 5: Linear regression (statistics)

This question is about the Tennesee kindergarten experiment. Kindergarten students were **randomly assigned** to "regular" or "small" classes. At the end of the schoolyear, all students were given standardized tests. The variable $TestScore$ measures the score of a student on that standardized test. The variable $SmallClass$ is a binary variable that is equal to 1 if a student is assigned to a small class, and 0 otherwise. A regression of $TestScore$ on $SmallClass$ yields

$$\underset{}{\widehat{TestScore}} = \underset{(1.6)}{918.0} + \underset{(2.5)}{13.9} \times SmallClass\,.$$

Answer the following questions:

1. The underlying model is $TestScore_i = \beta_0 + \beta_1 \times SmallClass_i + u_i$. 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. Explain why Least Squares Assumption 1 holds in this example.
4. Construct a 95% confidence interval for β_1.
5. What is the interpretation of β_1?

6 Self-study

1. [Section 4.5] Under the Least Squares Assumptions, what is the expectation of $\hat{\beta}_0$? Is $\hat{\beta}_0$ consistent?
2. What does the abbreviation "i.i.d." stand for?

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:

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

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