Problem sets for BUEC 333

I will indicate the relevant exercises for each week at the end of the lecture. Numbered exercises are back-of-chapter exercises from Stock and Watson. Try to complete the exercises before going to the tutorials. In the tutorials, the TAs will help you if you have any difficulties.

Probability

1. **An unfair die.** You are about to roll an unfair die, and are interested in the random variable "Number of eyes that show", denoted by X. The probability of "rolling a 1" is one-half (1/2), in other words:

$$P(X = 1) = 1/2.$$

All the other outcomes (2,3,4,5) have equal probability, i.e.

$$P(X = 2) = \cdots = P(X = 6).$$

   a) Write down the probability distribution of X
   b) Consider the event A, "throwing more than 3". Write A in terms of elements of sample space. Compute $P(A)$.
   c) Consider the event B, "throwing an odd number". Write B in terms of elements of sample space. Compute $P(B)$.
   d) Compute the variance of this RV X. Compare this to the variance of a fair die throw.

2. **Three coin flips (Final exam, Summer 2013).** You are about to flip 3 coins. The random outcome is what the coins are showing. For example, you could throw HHT: first two coins show heads, third shows tails. Or HTH, TTT, etcetera. This experiment can lead to different random variables. Let us look at two of them:

X_1**:** "the number of heads showing";

X_2**:** "the number of heads showing **minus** the number of tails showing".

Answer the following questions:

   a) What is the probability distribution for X_1? Note that you first need to determine the sample space.
   b) What is the probability distribution for X_2?

3. **A Bernoulli RV.** Let X be a random variable, with sample space $\Omega = \{0, 1\}$. Let the probability that the RV X take the value 0 be denoted as $P(X = 0) = p$.

   a) Compute $P(X = 1)$.
   b) Compute $E(X)$.
   c) Compute $\text{var}(X)$.

4. **Variance.** In class, we considered the random variable X with sample space $\Omega = \{-1, 1\}$ and associated probabilities $\{1/2, 1/2\}$ and found that the variance of this RV was 1 [Check this]. This setup corresponds to the probability distribution in Table 1.

Outcome x	$P(X = x)$
-1	1/2
1	1/2

Table 1: Probability distribution.

We also considered the modified RV in Table 2, and found the variance to equal 4.

Outcome x	$P(X = x)$
-2	1/2
2	1/2

Table 2: Probability distribution.

Compute the variance of the RVs described in Tables 3 and 4.

Outcome x	$P(X = x)$
-1	1/2
2	1/2

Table 3: Probability distribution.

Outcome x	$P(X = x)$
-1	1/10
1	9/10

Table 4: Probability distribution.

5. **Monty Hall II.** "You are a contender on a game show. There are **10** doors. Behind one is a car. Behind the other **9** doors is a goat. You pick door No. 1 (the door on the left). The game show host, who knows what is behind each door, opens **8** of the other **9** doors, revealing **8** goats. The host was instructed to open doors with goats only. The host asks you whether you want to switch to the other closed door."
   a) The value of the prize you receive is a random variable. Derive the probability distribution if your strategy is to switch.
   b) Derive the probability distribution if your strategy is **not** to switch.
   c) Should you switch?

6. **Monty Hall III.** "You are a contender on a game show. There are **10** doors. Behind one is a car. Behind the other **9** doors is a goat. You pick door No. 1 (the door on the left). The game show host, who knows what is behind each door, opens **1** of the other **9** doors, revealing **a** goat. The host was instructed to open a door with a goat. The host asks you whether you want to switch to one of the other closed doors."
   a) The value of the prize you receive is a random variable. Derive the probability distribution if your strategy is to switch.
   b) Derive the probability distribution if your strategy is **not** to switch.
   c) Should you switch?

7. **Russian roulette II.** Consider the following modification of the Russian roulette example: "Your enemy challenges you to play Russian Roulette with a 6-cylinder pistol (meaning it has room for 6 bullets). He puts **3** bullets into the gun in consecutive slots, and leaves the next **3**

slots blank. He spins the barrel and hands you the gun. You point the gun at yourself and pull the trigger. It doesn't go off. Your enemy tells you that you need to pull the trigger one more time, and that you can choose to either spin the barrel at random, or not, before pulling the trigger again. Spinning the barrel will position the barrel in a random position."

- What would you do: spin again, or pull the trigger without spinning? Derive the relevant conditional probabilities.

8. **Independence.** This exercises requires self-study of the section "Independence" in Chapter 2 of Stock and Watson. Let X and Y be two RVs, both with sample space $\{10, 20\}$. You know that $P(X = 10) = 0.3$ and $P(Y = 10) = 0.4$.

   a) Assume that X and Y are independent. What is $P(X = 10, Y = 20)$?
   b) Assume that X and Y are independent. What is $P(X = 10|Y = 20)$?
   c) If X and Y are not independent, can you compute $P(X = 10, Y = 20)$?

9. **Linear functions of random variables.** Solve Exercise 2.3 in Stock and Watson (page 56).

10. **A fair die and an unfair die.** Imagine the following part of a game. There are two dice. One of them is fair, and one of them is unfair. First, you randomly (blindfolded) select a die. Second, you throw that die. Let the RV X be the number of eyes that you ultimately throw, and let the RV Y denote which die you picked in the first step (0: fair die, 1: unfair die). 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 5: Two-step game

. Answer the following questions:

   a) Given that you have selected the unfair die, what is the probability of throwing a 6?
   b) Given that you throw a 4, what is the probability that you selected the fair die?
   c) What is the probability of throwing a 3, $P(X = 3)$?
   d) Compute $E(X|Y = 1)$ and $E(X|Y = 0)$.
   e) Using the answer under (c), and the law of iterated expectations, compute $E(X)$.
   f) Compute $\text{var}(X|Y = 1)$.

11. **Computer crashes.** Consider the joint probability distribution in Table 2.3 (A).

   a) Compute $E(M|A = 0)$ and $E(M|A = 1)$.
   b) What do you conclude? / What does this mean?
   c) Compute $\text{var}(M|A = 0)$.

12. **Covariance and correlation.** Do exercise 2.9(c) from Stock and Watson, pages 29 and 30.

13. **Practice with the Normal.** Do exercise 2.10 (Stock and Watson, page 58).

14. **Normal squared.** Let X be a random variable, and $X \sim \text{Normal}(3, 4)$. Compute $P(X^2 > 4)$.

15. **Sample that is identical but not independent (i.n.i.d).** Do exercise 2.26 (a) and (b), from Stock and Watson, page 62.

16. **Waiting for the bus.** [Questions (a)-(c) repeat lecture material]. You are taking the bus from stop A to stop B. A bus comes by stop A every 15 minutes. You arrive at a bus station. Assume that you have no information about how long ago the last bus came by, for example from the number of people currently waiting. Let X_1 be a random variable that represents the number of minutes you are going to wait for the bus.

    a) Why is X_1 a random variable? Why is X_1 a continuous random variable?

    b) What is the smallest value X_1 can take? What is the largest value?

    c) Draw the the probability density function (pdf) of X_1 (let's call it $f_1(x)$) in a graph. Make sure to note the values on the vertical axis.

    d) Consider the random variable X_2: X_2 comes from a similar situation as X_1, but now there is a bus every **10** minutes. Add the pdf for X_2, say $f_2(x)$, to the graph you drew under (c).

    e) Compute $P(1 \leq X_1 \leq 3)$ and $P(1 \leq X_2 \leq 3)$.

    f) In one graph, draw $F_1(x)$ and $F_2(x)$, the cumulative distribution functions (cdfs) of X_1 and X_2.

17. **Kahnemann and Tversky.** A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50 percent of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50 percent, sometimes lower. For a period of 1 year, each hospital recorded the days on which more than 60 percent of the babies born were boys. Which hospital do you think recorded more such days?

    A. the smaller hospital

    B. the bigger hospital

    C. about the same (within 5 percent of each other)

18. **Three coin flips (Final exam, Fall 2013).** You are about to flip 3 coins. The random outcome is what the coins are showing. For example, you could throw HHT: first two coins show heads, third shows tails. Or HTH, TTT, etcetera. Consider the RV's:

    X_1**:** "the number of heads showing on the **first two** flips";

    X_2**:** "the numer of heads showing **times** the number of tails showing".

    a) What is the probability distribution for X_2? (Hint: You first need to determine the sample space.)

    b) Compute $\mathrm{E}(X_2)$

    c) Compute $\mathrm{Var}(X_1)$.

    d) Compute $\mathrm{Cov}(X_1, X_2)$.

    e) Compute $\mathrm{E}(X_2 | X_1 = 2)$.

19. [Stock and Watson, 2.5] "In September, Seattle's daily high temperature has a mean of 59 degrees Fahrenheit. The standard deviation is 9 degrees Fahrenheit." Remember that, to convert from degrees Fahrenheit to degrees Celsius, we need to subtract 32 and then multiply by 5/9, so

$$T_C = \frac{5}{9} \times (T_F - 32).$$

    For the questions that follow, indicate which formulas you are using.

    a) What is the mean of Seattle's daily high temperature in September in degrees Celsius?

b) What is the standard deviation of Seattle's daily high temperature in September in degrees Celsius?

c) What is the variance of Seattle's daily high temperature in September in degrees Celsius?

20. **(Exam, Fall 2012).** Consider the joint probability distribution of the random variables X_1 and X_2 given in Table 6.

    a) What is the probability that $X_1 = 0$ given that $X_2 = 2$?
    In other words, what is $P(X_1 = 0|X_2 = 2)$?

    b) What is the conditional expectation of X_1 given that $X_2 = 3$? In other words, what is $E(X_1|X_2 = 2)$?

Statistics

1. **Central Limit Theorem.** Do Stock and Watson exercises 2.17 and 3.1.

2. **Confidence intervals.** Exercise 3.4 (a)-(c) on Stock and Watson, page 97, which requires you to do 3.3 (a)-(b) on the same page.

3. **Confidence intervals (2).** Do exercises 3.8 and 3.13(a).

4. **A weighted estimator.** Do exercise 3.11 (Stock and Watson, page 99).

5. **Comparing estimators (Final exam 2014).** Suppose you want to know the mean value of Y, $\mathrm{E}(Y) = \mu_Y$. You have a random sample of size n, $\{Y_1, \cdots, Y_n\}$. For simplicity, assume that $n = 2$. Then, the sample average $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, simplifies to $\bar{Y} = (Y_1 + Y_2)/2$.

    a) Is the sample average unbiased for μ_Y? Explain.

    b) Now, consider $\tilde{Y} = \frac{1}{4}Y_1 + \frac{1}{4}Y_2$. What is the variance of $\tilde{Y}$?

    c) On the basis of the variances, do you prefer $\bar{Y}$ or $\tilde{Y}$?

    d) What is wrong with $\tilde{Y}$?

    e) Now, consider the estimator $\check{Y} = \left(\mu_Y + \bar{Y}\right)/2$. Why is this not a good estimator? (Hint: the answer has nothing to do with efficiency, unbiasedness, or consistency.)

6. Stock and Watson 3.2.

7. Stock and Watson 3.14.

8. Stock and Watson 3.15.

9. Stock and Watson 2.7.

10. **Most basic statistic question (Midterm 2014).** Assume that you have a random sample $(2, 3, 4)$ for a RV X with mean μ_x and variance σ_x^2.

    a) For this sample, compute $\bar{X}$.

    b) For this sample, compute the sample variance.

    c) Construct a 90%-CI for μ_x based on the sample mean and sample variance, assuming that the sample is very, very large.

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

Table 6: Joint probability distribution of X_1 and X_2.

Student	GPA	ACT
1	2.8	21
2	3.4	24
3	3.0	26
4	3.5	27
5	3.6	29
6	3.0	25
7	2.7	25
8	3.7	30

Table 7: Students's ACT and GPA scores

i	X_i	Y_i
1	-3	-3
2	-2	$-k$
3	-1	-1
4	1	1
5	2	k
6	3	3

Table 8: Artificial data.

d) What is the use of making the assumption that n is "very very large"?

e) What is the interpretation of the numbers in your answer to (c)

11. **Scaled sample mean (Exam, Fall 2012).** Let X be a random variable with mean $E[X] = \mu$. We have a random sample of observations on X, namely $(X_1, \cdots, X_n)$. Let $\bar{X}$ denote the sample average, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Now, let $\widetilde{X} = \frac{1}{n-2}\sum_{i=1}^{n} X_i$ be the scaled sample average. Is $\widetilde{X}$ unbiased?

12. Stock and Watson, Exercise 3.9.

Simple linear regression: Mechanics and fit

1. (Wooldridge, Exercise 2.3) Table 1 presents a random sample of 8 observations on students' ACT (American College Test) and GPA (grade point average) scores. Consider the model

$$GPA_i = \beta_0 + \beta_1 ACT + u_i$$

   a) Compute the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_0$.
   b) Does $\hat{\beta}_0$ have a useful interpretation?
   c) What is the interpretation of $\hat{\beta}_1$?
   d) For each of the 8 observations, compute $\hat{u}_i$. Then, compute $\sum_{i=1}^{8} \hat{u}_i$
   e) What is the predicted value of GPA when $ACT = 20$?
   f) How much of the variation in GPA for these eight students is explained by ACT?

2. Consider the data in Table . Note that this data has some observations that depend on k.
   a) Compute R^2 if $k = 2$.
   b) Can you compute $\hat{\beta}_0$ and $\hat{\beta}_1$ for arbitrary values of k?
   c) Compute R^2 if $k = 1$.
   d) Compute R^2 if $k = 3$.

e) Compute R^2 if $k = 10$.

f) Can you explain what happens when k becomes very large?

3. Stock and Watson, 4.1 (a)-(c).
4. Stock and Watson, 4.2.
5. Stock and Watson, 4.3.
6. Stock and Watson, Exercise 4.9.

Simple linear regression: Estimation and inference

1. Stock and Watson, Exercise 4.4.
2. Stock and Watson, Exercise 4.6.
3. Stock and Watson, Exercise 4.7.
4. Stock and Watson, Exercise 5.2 (a)-(d).
5. Stock and Watson, Exercise 5.3
6. (Final, Summer 2014) Consider the following estimated regression equation that describes the relationship between a student's weight and height:

$$\widehat{WEIGHT} = 100 + 6\,HEIGHT$$

   a) A student has height 5. What is the regression's prediction for that student's weight?

   b) In the sample, the sample average of HEIGHT is 4. What can you say about the sample average for WEIGHT?

   Now, an additional variable is included, is ID, a student's SFU ID. Obviously, this is a nonsensical variable that is not in any way related to a student's weight. The new estimated regression equation is

$$\widehat{WEIGHT} = 100 + 6\,HEIGHT + 0.02\,ID$$

   c) Someone's weight has nothing to do with their SFU ID. Still, the R^2 went up from 0.74 to 0.75. How is this possible?

   d) On the other hand, the adjusted R-squared $\bar{R}^2$ went down from 0.73 to 0.72. Explain how it is possible that the R^2 can go up while the $\bar{R}^2$ goes down.

   e) If the student's SFU ID number is not related to a student's weight, should the estimated coefficient not be equal to 0? How could it be that it is 0.02?

7. (Final, Summer 2012) This question tests your understanding of the concepts involved in regression analysis. Consider the model with one regressor,

$$Y_i = \beta_0 + \beta_1 X_i + u_i.$$

   a) Write down the population regression function. Write down the sample regression function.

   b) What does Least Squares Assumption 1, $\mathrm{E}\,(u_i|\,X_i) = 0$, mean? In your answer, use the words "other factors".

   c) Describe the other two "Least Squares Assumptions"? You only need to use one sentence for each assumption.

   d) Is β_1 a random variable? Is $\hat{\beta}_1$ an estimator? Is u_i a random variable? Is $\hat{u}_i$ a random variable? (First: 4x yes/no. Then: short explanations if you are not sure about your yes/no answers.)

Prediction and forecasting

1. Exercise 14.1, page 571.
2. 14.7 (a), (b), and (d), on page 574.
3. Consider the time series $(-1, 1, -1)$.
   a) Compute the first sample autocovariance.
   b) Compute $\hat{\rho}_1$.
   c) Can you construct a time series that has its first sample autocovariance equal to zero?
4. What is the difference between a prediction and a forecast?

Introduction to multiple linear regression

1. Exercises 6.1-6.4.
2. Exercise 6.5.
3. Exercise 6.6.
4. Exercise 6.7.
5. Exercise 6.8.
6. Exercise 6.9.
7. [Final, Summer 2014] Let W_i be an individual's wage, let $EDUC_i$ be their education, and let F_i be a dummy variable that equals 1 if and only if that person is female. Furthermore, let M_i be a dummy variable that equals 1 if and only if that person is male. Consider the following model:

$$W_i = \beta_0 + \beta_1 EDUC_i + \beta_2 F_i + \beta_3 M_i + \epsilon_i.$$

   a) What is the problem with using OLS to estimate the parameters in this regression equation?
   b) How do you solve this problem?
8. [Final, Summer 2014] Consider the following estimated regression equation that describes the relationship between a student's weight and height:

$$\widehat{WEIGHT} = 100 + 6.0\, HEIGHT$$

   a) A student has height 5. What is the regression's prediction for that student's weight?
   b) In the sample, the sample average of HEIGHT is 4. What can you say about the sample average for WEIGHT?

      Now, an additional variable is included, is ID, a student's SFU ID. Obviously, this is a nonsensical variable to include: it is not in any way related to a student's weight. The new estimated regression equation is

$$\widehat{WEIGHT} = 101.5 + 5.98\, HEIGHT + 0.02\, ID$$

   c) Someone's weight has nothing to do with their SFU ID. Still, the R^2 went up from 0.74 to 0.75. How is this possible?
   d) If the post office box number is not related to a student's weight, should the estimated coefficient not be equal to 0? How could it be that it is 0.02?

Topics in linear regression

1. Exercise 5.2, skip (b).
2. Exercise 5.5, (a) and (c).
3. Exercise 5.7
4. Exercise 5.10.
5. [Final, Summer 2014] Let D_i be a dummy variable. Consider the model that consists of the equation

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 D_i X_i + u_i$$

and the standard OLS assumptions

   a) Draw a graph to visualize this model.
   b) What is the interpretation of β_3?

6. Exercise 7.1, page 242.
7. Exercise 7.2, page 242.
8. Exercise 7.6, page 244.
9. During the lectures, we have discussed that, in the model

$$\log Y_i = \beta_0 + \beta_1 \log X_i + u_i$$

you can interpret β_1 as approximately measuring the expected percentage change in Y_i given a one-percent change in X_i, c.p. Similarly, we know that in the model

$$\log Y_i = \beta_0 + \beta_1 X_i + u_i,$$

we can interpret $\beta_1 \times 100\%$ as the expected percentage change in Y_i for a one-unit change in X_i. What is the interpretation of β_1 (or $\beta_1 \times 100\%$) in the model

$$Y_i = \beta_0 + \beta_1 \log X_i + u_i,$$

known as the linear-log model?

10. Exercise 8.7 (SW page 300).

Instrumental variables

1. Exercise 12.5.
2. Exercise 12.9.
3. Exercise 12.10.

Panel data

1. (Final, Summer 2014). Consider the example used in the chapter on panel data. We have a panel data set on $n = 48$ U.S. states during $T = 7$ periods, from 1982 up to and including 1988. The total number of observations is 336.

   a) Is this a balanced panel? Explain.

b) For each state, in each time period, let Y_{it} denote the number of annual traffic deaths per 10000 in the population. Let X_{it} denote the beer tax in 1988 U.S. dollars. Temporarily ignore the data after 1982, so that we have a cross-section of 48 states. The estimated regression line gives

$$\hat{Y}_{i,1982} = 2.01 + 0.13X_{i,1982}.$$

If the Least Squares assumptions hold for this regression, how would you interpret the 0.13? Include in your answer: "tax on beer".

c) Alternatively, we can use fixed effects regression to estimate the effect fixed effects regression line is

$$\begin{aligned}\hat{Y}_{i,t} &= \hat{\alpha}_i - \underset{(0.29)}{0.66}X_{i,t}.\end{aligned}$$

How would you interpret the -0.66?

d) Consider the results for the fixed effects regression. Do you think that the Least Square assumptions hold, i.e. do you believe that the 0.13 in the first result comes from an unbiased estimator? if YES: explain what causes the difference between 0.13 and -0.66. If NO: explain why the Least Square assumptions are unlikely to hold.

2. This question is about the code that we used during the lecture of Wednesday, July 17. You can find it on the course website or directly through [this link]. We are going to investigate the output of the commands on lines 236 and 250. Remember that the data set we are using is a random sample of 935 "young men" in the U.S. in 1980. A list of variables and descriptions is

```
[...]
# 1. wage                          monthly earnings
# 2. hours                         average weekly hours
# 3. IQ                            IQ score
# 4. KWW                           knowledge of world work score
# 5. educ                          years of education
# 6. exper                         years of work experience
# 7. tenure                        years with current employer
# 8. age                           age in years
[...]
# 12. urban                         =1 if live in SMSA
# 13. sibs                          number of siblings
# 14. brthord                       birth order
# 15. meduc                         mother's education
# 16. feduc                         father's education
[...]
```

Here is the first regression and its output:

```
> summary(lm(IQ~brthord, data=wageData.2))

[...]

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 105.6634     0.8699 121.470  < 2e-16 ***
brthord      -1.6640     0.3129  -5.318 1.34e-07 ***
---

[...]
```

a) What is the question we are trying to answer?

b) What results would you draw from this output, if you knew that all the OLS assumptions were satisfied?

c) Assumption (1) is unlikely to hold. Why?

d) To solve that problem, we are going to include additional regressors, namely the education of both parents ($feduc$ and $meduc$).

   i. Could you make an argument that would support $\text{cov}(feduc, brthord) \neq 0$ or $\text{cov}(meduc, brthord) \neq 0$?

   ii. Can you make an argument to support the statement "The regression coefficients on $feduc$ and $meduc$ are likely to be different from zero.

e) The output from that regression follows:

```
> summary(lm(IQ~brthord+feduc+meduc,data=wageData.2))

[...]

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  85.4160     2.6170  32.639  < 2e-16 ***
brthord      -1.0544     0.3712  -2.840 0.004647 **
feduc         0.9780     0.1970   4.963 8.83e-07 ***
meduc         0.8602     0.2333   3.687 0.000245 ***
---

[...]
```

If the OLS assumptions are valid, how do you interpret these results?

f) Do you believe that the conclusion under (e) is correct? Why / why not?