HW2_answers

August 19, 2019

1 Homework #2

2 ARE 106 Summer Session II

2.1 Name:

2.2 SSID:

Please put your name and SSID in the corresponding cells above.

The homework is worth 13.5 points.

For each of the following questions, show as much of your steps as you can (without going overboard). If you end up getting the wrong answer, but we spot where you made a mistake in the algebra, partial credit will be more readily given. If you only put the final answer, you will be marked either right or wrong.

Answer questions in the correct cell. For problems where you have to input math, make sure that you know that it's a markdown cell (It won't have a `In: []` on the left) and make sure you run the cell by either pressing `Ctrl + Enter` or going to `Cell -> Run Cell`. Alternatively, write all your answers and then go to `Cell -> Run All Cells` after you're done.

When you are finished export your homework to a PDF by going to `File -> Download as -> PDF`.

2.2.1 Exercise 1: Statistics Review

For the following problems, write down the answer in the markdown cell below the question. If you have trouble writing math, either double-click on the markdown cell to see how it was written or consult this cheatsheet:

Assume that there are two random variables X and Y with expected values μ_x and μ_y, and variances σ_x and σ_y. Let the covariance between X and Y be: σ_{xy}. Note that the covariance between these two variables need not be 0 (i.e. they aren't necessarily independent).

Any others (a, b, c etc...) can be considered non-random parameters.

For each of the following problems, rewrite the expressions in terms of $\mu_x, \mu_y, \sigma_x, \sigma_y$ and $\sigma_{x,y}$ when possible. If not possible, write what you think the answer is and explain why.

- i. $E(aX + bY)$

- ii. $Var(bY + c)$

- iii. $Cov(aX + b, c^2Y + d)$

- iv. $E(b)$, remember b is non-random

- v. $Var(a)$, remember a is non-random

Please write your answers here. If you need to use more than one line, you may do so.

- i. $E(aX) + E(bY) = a\mu_x + b\mu_y$

- ii. $b^2 \sigma_y$

- iii. $ac^2 \sigma_{xy}$

- iv. b

- v. 0

2.2.2 Exercise 2: Statistics Review Continued

This exercise uses the same instructions as question 1.

- i. What is the definition of $Cov(X, Y)$ in terms of expectations? **Hint: check the statistics review slides from the first class**.

- ii. Rewrite this expression so that it has $E(XY)$ on one side. Does $E(XY) = \mu_x \mu_y$? Why or why not? **Hint: μ_x and μ_y are not random!**

Please write your answers here. If you need to use more than one line, you may do so.

- i. $Cov(X, Y) = E\left[(X - \mu_x)(Y - \mu_y)\right]$

- ii. $Cov(X, Y) = E(XY) - \mu_x \mu_y - \mu_x \mu_y + \mu_x \mu_y = E(XY) - \mu_x \mu_y$

Therefore: $E(XY) = Cov(X, Y) + \mu_x \mu_y$
In order for this to be true, $Cov(X, Y) = 0$.

2.2.3 Exercise 3: Deriving the R^2.

Suppose that we have an estimated regression model $y_i = \hat{b}_0 + \hat{b}_1 x_i + e_i$, where $\hat{b}_0, \hat{b}_1$ are estimated OLS coefficients. Let $\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i$, so that:

$$y_i = \hat{y}_i + e_i$$

Use the definition for $Var(y_i)$ to derive the expression for the R^2. What expression has to be equal to 0 in order for this to be true?
Please write your answer here. If you need to use more than one line, you may do so.
$Var(y_i) = Var(\hat{y}_i + e_i) = Var(\hat{y}_i) + Var(e_i) + 2Cov(\hat{y}_i, e_i)$
But we know that $Cov(\hat{y}_i, e_i)$ must be equal to 0, so:
$\frac{1}{N} \sum_i^N (y_i - \mu_y)^2 = \frac{1}{N} \sum_i^N (\hat{y}_i - \bar{\hat{y}}_i)^2 + \frac{1}{N} \sum_i^N (e_i - \bar{e})^2$
We can see that this is the same as $TSS = ESS + SSR$. Since $R^2 = \frac{ESS}{TSS}$, if we divide the above expression by TSS, we get:
$1 = R^2 + \frac{SSR}{TSS}$
So $R^2 = 1 - \frac{SSR}{TSS}$

2.2.4 Exercise 4: Ordinary Least Squares

Suppose you have some data x_i and you would like to see how it affects y_i, your dependent variable. You think that this relationship can be estimated linearly, so the model you write down is:

$$y_i = b_0 + b_1 x_i + e_i$$

- i. If you wanted to estimate this relationship using OLS, what would be the expressions for $\hat{b}_0$ and $\hat{b}_1$? *You do not need to show your work for this question, you can just write it.*

- ii. Now suppose that the independent variable, x_i, is made up of really small number (ex. many of the observations are .000001). You decide to multiply this variable by $1,000$ so that the variable doesn't have quite so many 0's after the decimal point (making it easier to understand). Let's see how that would change the OLS coefficients.

  - a. What would your estimating equation look like now? (Hint:It will be similar to equation (1), but with one change)
  - b. Write the minimization problem that you need to solve.
  - c. Solve the minimization problem and give the answers for the expressions for the resulting, new OLS coefficients.
  - d. Which coefficient is bigger or smaller?

Please write your answer here. If you need to use more than one line, you may do so.

- i. $b_0 = \bar{y} - b_0 - b_1 \bar{x}$

  $b_1 = \frac{\sum_i^N (X - \bar{X})(Y - \bar{Y})}{\sum_i^N (X - \bar{X})^2}$

- ii.

  - a. $y_i = b_0 + b_1 1000 x_i + e_i$
  - b. $\min_{b_0, b_1} \sum_i^N (y_i - b_0 - b_1 1000 x_i)^2$
  - c. The first order conditions are:

  $$(b_0) : 2 \sum_i^N (y_i - b_0 - b_1 1000 x_i)(-1) = 0$$

  $$(b_1) : 2 \sum_i^N (y_i - b_0 - b_1 1000 x_i)(-1000 x_i) = 0$$

  Solving for b_0, we get:
  $\sum_i^N y_i - \sum_i^N b_0 - \sum_i^N b_1 1000 x_i = 0$
  $N\bar{y} - N b_0 - b_1 1000 N \bar{x} = 0$
  So $\hat{b}_0 = \bar{y} - b_1 1000 \bar{x}$
  Now substituting that into the b_1 first order condition: $\sum_i^N (y_i - (\bar{y} - b_1 1000 \bar{x}) - b_1 1000 x_i)(x_i) = 0$
  $\sum_i^N (y_i - \bar{y}) x_i - 1000 b_1 \sum_i^N (x_i - \bar{x}) x_i = 0$
  So we get:
  $\hat{b}_1 = \frac{1}{1000} \frac{\sum_i^N (y_i - \bar{y}) x_i}{\sum_i^N (x_i - \bar{x}) x_i}$

  - d. From this we can see that the new coefficient is now smaller than before.

2.2.5   Exercise 5: Python Coding

Suppose we have a list of numbers `my_list  = [1,3,5,6,7]`. Write a program that takes this list, and multiplies it by 2, but if the number is 3, print out, "this is the number 3."

So what we're looking for, is a function that takes in a list and outputs a list.

Input: `[1,3,5,6,7]`

Output:

```
[2,6,10,12,14]
"this is the number 3"
```

Remember to comment your code and to initialize your lists first (you might need to initialize an empty list first). A comment can be made by using the # symbol. Also remember that to tell the function what you want to output, you need to include a `return` statement in the function.

```python
[5]: ### Please put your answer here:

## Initialize list
my_list = [1,3,5,6,7]

## define function that multiplies each part of the list by 2 and prints␣
 ↪something out for 3.
def f(x):

    ## initialize empty list for collection of returned values
    list_collector = []
    for thing in x:

        list_collector.append(2*thing)
        if thing==3:
            print("this is the number 3")

    return list_collector

f(my_list)
```

```
this is the number 3
```

```
[5]: [2, 6, 10, 12, 14]
```