hw01-solution

February 19, 2020

1 Homework 1: Causality and Expressions

Please complete this notebook by filling in the cells provided. Before you begin, execute the following cell to load the provided tests.

```
[1]: # Don't change this cell; just run it.
     from client.api.notebook import Notebook
     ok = Notebook('hw01.ok')
     _ = ok.auth(inline=True)
```

```
======================================================================
Assignment: Homework 1: Causality and Expressions
OK, version v1.14.20
======================================================================
```

Successfully logged in as m.zareei@ieee.org

Reading: - Textbook chapters 1, 2, and 3

Deadline:

Homework 1 is due Tuesday, 25 Feb at 11:59pm.

Directly sharing answers is not okay, but discussing problems with the course staff or with other students is encouraged.

Start early so that you can come to office hours if you're stuck. Check the website for the office hours schedule. It is much better to use https://mzareei.youcanbook.me/.

Once you're finished, select "Save and Checkpoint" in the File menu and then execute the `submit` cell below. The result will contain a link that you can use to check that your assignment has been submitted successfully. If you submit more than once before the deadline, we will only grade your final submission.

```
[ ]: _ = ok.submit()
```

1.1 1. Scary Arithmetic

An ad for ADT Security Systems says,

> "When you go on vacation, burglars go to work [...] According to FBI statistics, over 25% of home burglaries occur between Memorial Day and Labor Day."

Do the data in the ad support the claim that burglars are more likely to go to work during the time between Memorial Day and Labor Day than at other times? Please explain your answer.

*Write your answer here, replacing this text.*

1.2 2. Characters in Little Women

In lecture, we counted the number of times that the literary characters were named in each chapter of the classic book, Little Women. In computer science, the word "character" also refers to a letter, digit, space, or punctuation mark; any single element of a text. The following code generates a scatter plot in which each dot corresponds to a chapter of Little Women. The horizontal position of a dot measures the number of periods in the chapter. The vertical position measures the total number of characters.

```
[2]: # This cell contains code that hasn't yet been covered in the course,
     # but you should be able to interpret the scatter plot it generates.

     from datascience import *
     from urllib.request import urlopen
     import numpy as np
     %matplotlib inline

     little_women_url = 'https://www.inferentialthinking.com/data/little_women.txt'
     chapters = urlopen(little_women_url).read().decode().split('CHAPTER ')[1:]
     text = Table().with_column('Chapters', chapters)
     Table().with_columns(
         'Periods',    np.char.count(chapters, '.'),
         'Characters', text.apply(len, 0)
         ).scatter(0)
```
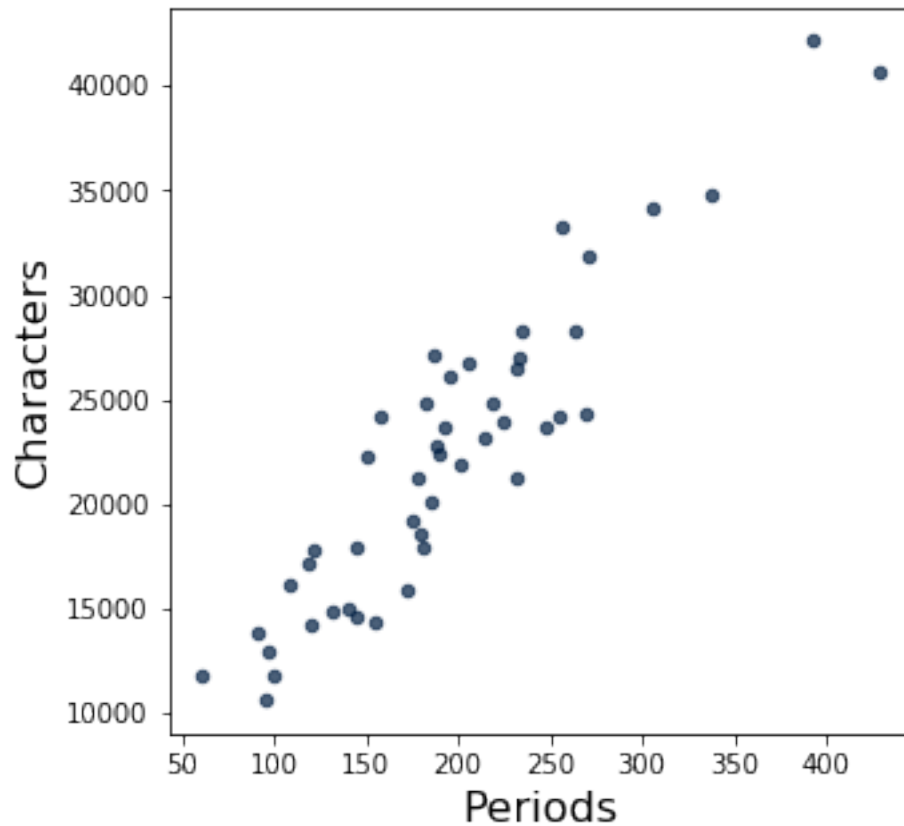
Question 1. About how many periods are in the chapter with the most characters per period? Assign either 1, 2, 3, 4, or 5 to the name `characters_q1` below.

1. 60
2. 80
3. 175
4. 250
5. 350

```
[3]: characters_q1 = 1 #SOLUTION
```

Question 2. About how many periods are in the chapter with the most characters? Assign either 1, 2, 3, 4, or 5 to the name `characters_q2` below.

1. 250
2. 390
3. 440
4. 32,000
5. 40,000

```
[4]: characters_q2 = 2 #SOLUTION
```

Check that your answers are in the correct format. This test *does not* check that you answered correctly; only that you assigned a number successfully in each multiple-choice answer cell.

```
[5]: _ = ok.grade('q2')
```

```
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Running tests

---------------------------------------------------------------------
Test summary
    Passed: 1
    Failed: 0
[ooooooooook] 100.0% passed
```

To discover more interesting facts from this plot, read Section 1.3.2 of the textbook.

1.3   3. Names and Assignment Statements

Question 1. When you run the following cell, Python produces a cryptic error message.

```
[ ]: 4 = 2 + 2
```

Choose the best explanation of what's wrong with the code, and then assign 1, 2, 3, or 4 to `names_q1` below to indicate your answer.

1. Python is smart and already knows `4 = 2 + 2`.

2. `4` is a number, and it doesn't make sense to make a number be a name for something else. In Python, "`x = 2 + 2`" means "assign x as the name for the value of `2 + 2`."

3. It should be `2 + 2 = 4`.

4. I don't get an error message. This is a trick question.

```
[6]: names_q1 = 2 # SOLUTION
```

Question 2. When you run the following cell, Python will produce another cryptic error message.

```
[ ]: two = 3
     six = two plus two
```

Choose the best explanation of what's wrong with the code and assign 1, 2, 3, or 4 to `names_q2` below to indicate your answer.

1. The `plus` operation only applies to numbers, not the word "two".

2. The name "two" cannot be assigned to the number 3.

3. Two plus two is four, not six.

4. The name `two` cannot be followed directly by another name.

```
[7]: names_q2 = 4 # SOLUTION
```

Check that your answers are in the correct format. This test *does not* check that you answered correctly; only that you assigned a number successfully in each multiple-choice answer cell.

```
[8]: _ = ok.grade('q3')
```

```
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Running tests

---------------------------------------------------------------------
Test summary
    Passed: 1
    Failed: 0
[ooooooooook] 100.0% passed
```

1.4 4. Job Opportunities & Education in Rural India

A study at UCLA investigated factors that might result in greater attention to the health and education of girls in rural India. One such factor is information about job opportunities for women. The idea is that if people know that educated women can get good jobs, they might take more care of the health and education of girls in their families, as an investment in the girls' future potential as earners.

The study focused on 160 villages outside the capital of India, all with little access to information about call centers and similar organizations that offer job opportunities to women. In 80 of the villages chosen at random, recruiters visited the village, described the opportunities, recruited women who had some English language proficiency and experience with computers, and provided ongoing support free of charge for three years. In the other 80 villages, no recruiters visited and no other intervention was made.

At the end of the study period, the researchers recorded data about the school attendance and health of the children in the villages.

Question 1. Which statement best describes the *treatment* and *control* groups for this study? Assign either 1, 2, or 3 to the name `jobs_q1` below.

1. The treatment group was the 80 villages visited by recruiters, and the control group was the other 80 villages with no intervention.

2. The treatment group was the 160 villages selected, and the control group was the rest of the villages outside the capital of India.

3. There is no clear notion of *treatment* and *control* group in this study.

```
[9]: jobs_q1 = 1 # SOLUTION
```

Question 2. Was this an observational study or a randomized controlled experiment? Assign either 1, 2, or 3 to the name `jobs_q2` below.

1. This was an observational study.

2. This was a randomized controlled experiment.

3. This was a randomized observational study.

```
[10]: jobs_q2 = 2 # SOLUTION
```

Question 3. The study reported, "Girls aged 5-15 in villages that received the recruiting services were 3 to 5 percentage points more likely to be in school and experienced an increase in Body Mass Index, reflecting greater nutrition and/or medical care. However, there was no net gain in height. For boys, there was no change in any of these measures." Why do you think the author points out the lack of change in the boys?

Write your answer here, replacing this text.

Check that your answers are in the correct format. This test *does not* check that you answered correctly; only that you assigned a number successfully in each multiple-choice answer cell.

```
[11]: _ = ok.grade('q4')
```

```
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Running tests

------------------------------------------------------------------------
Test summary
    Passed: 1
    Failed: 0
[ooooooooook] 100.0% passed
```

1.5   5. Differences between Universities

Question 1. Suppose you'd like to *quantify* how *dissimilar* two universities are, using three quantitative characteristics. The US Department of Education data on UW and Cal describes the following three traits (among many others):

Trait	UW	Cal
Average annual cost to attend ($)	13,566	13,707
Graduation rate (percentage)	83	91
Socioeconomic Diversity (percentage)	25	31

You decide to define the dissimilarity between two universities as the maximum of the absolute values of the 3 differences in their respective trait values.

Using this method, compute the dissimilarity between UW and CAL. Name the result `dissimilarity`. Use a single expression (a single line of code) to compute the answer. Let Python perform all the arithmetic (like subtracting 91 from 83) rather than simplifying the expression yourself. The built-in `abs` function takes absolute values.

```
[12]: dissimilarity = max(abs(13566 - 13707), abs(83 - 91), abs(25 - 31))   # SOLUTION
      dissimilarity
```

`[12]:` 141

Check your work by running the cell below.

```
[13]: _ = ok.grade('q5_1')
```

```
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Running tests

------------------------------------------------------------------------
Test summary
    Passed: 1
    Failed: 0
[ooooooooook] 100.0% passed
```

Question 2. The national average annual cost of college is $16,126.50. If the average annual cost of UW and Cal were measured as a percentage of the national average rather than a dollar amount, what would be the dissimilarity between these two universities? Assign either 1, 2, or 3 to the name `universities_q2` below.

1. 1
2. 8
3. 141

```
[14]: universities_q2 = 2 # SOLUTION
```

Check that your answer is in the correct format. This test *does not* check that you answered correctly; only that you assigned a number successfully to `universities_q2`.

```
[15]: _ = ok.grade('q5_2')
```

```
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Running tests

------------------------------------------------------------------------
Test summary
    Passed: 1
    Failed: 0
[ooooooooook] 100.0% passed
```

1.6   6. More Children Living at Home?

A USA Today article from 2006 includes this sentence: "Since 1970, the percentage of people ages 18 to 34 [in the United States] who live at home with their family increased 48%, from 12.5 million

to 18.6 million, the Census Bureau says."

Question 1. The word "percentage" isn't used correctly in the context of the rest of the sentence. What word should replace it?

SOLUTION: "number" would make more sense. (In fact, saying that the "percentage" increased is quite misleading, since it implies that people ages 18 to 34 are more likely to live at home, when in fact there might just be more people today than in 1970.)

Question 2. In at most 3 sentences, describe whether changes in the US population are relevant to the data in the USA Today article.

SOLUTION: The US population increased by around 45% over this period, and the 18 to 34 population increased by around 42%. So we'd expect the number of people living at home to have increased by about that much, even if nothing else changed! 42% is pretty close to 48%, so that can explain most of the increase. In other words, people 18 to 34 are slightly more likely to live at home today, but there hasn't been a dramatic increase, and the USA Today story was explaining a phenomenon that was mostly due to a general population increase.