NDH802 - Assignment 3

Group no.

- The assignment includes 3 questions, 4 points each.
- Question 1 is purely theoretical (no data). Question 2 and 3 are empirical (based on the data provided).
- Submit your assignment via Canvas before 10:00 CET, May 21, 2021.
- Your submission should include (1) an RMarkdown file with your solutions in words and/or R code, (2) a knitted pdf file and (3) your handwritten solution if you have one. Name the files NDH802_Assignment1_GroupNumber.
- You should work in groups and contribute equally, or at least understand all parts of the assignment.
- You can copy my code, but make sure you understand it.
- You should not have the exact solutions and/or results with other groups.

Set things up

Set your working directory

```
#setwd("")
```

Load the package(s) you're going to use. If you don't use any packages, leave it as is.

```
library(tidyverse)
library(BSDA)
```

Run this code chunk to load data into your R Environment. The command will randomly select 1,000,000 rows of data from the original data set, i.e., every time you run the code, you have a new (unique) data set `df`. Accordingly, your results should be different from your friends and you should not be comparing them.

```
inference_dataset <-
  read.csv("~/Desktop/Data/NDH802/inference_dataset.csv")
df <-
  inference_dataset[sample(1:nrow(inference_dataset),
                           size = 1000000,
                           replace = FALSE), -1]
rm(inference_dataset)
```

Question 1. Sampling theory

Let X denote the height of all the Swedes in 2021. Assume $X \sim N(175, 5)$ (measured in cm).

(a) What is the probability that a random Swede is from 172cm to 177cm tall? (1p)

(b) Assume you randomly select 4 Swedes. What is the probability that *your sample mean height* is from 170cm to 180cm tall? (1p)

(c) Compare the results you have from (a) and (b). Why are they similar/different? (1p) Because they are technically different variables. (a) is X and (b) is $\bar{X}$

(d) What is the probability that *all of the 4 Swedes* you sample in (b) is from 170cm to 180cm tall? (1p) 0.021109

Questions 2 and 3 will be based on `df`.

Question 2. T-test

(a) Imagine you are the customer relationship manager. One of your colleagues argues that the average `value` of the customers is at least 4100. Based on your `df`, what do you say about this statement? (1p)

(b) Your colleague argues that the average `value` of the customers who return more than 40 items is no more than 4500. Based on your `df`, what do you say about this statement? (1p)

(c) Your boss would like to know if the multi-channel customers are more valuable than customers who shops offline only (based on the variable `value`). Perform the test at 95% confidence interval and explain the results in words. (1p)

(d) Your boss would like to know if the loyal and multi-channel customers are more valuable than the *not* loyal customers who shop offline only (based on the variable `value`). Perform the test at 95% confidence interval and explain the results in words. (1p)

Question 3. Correlation and regression

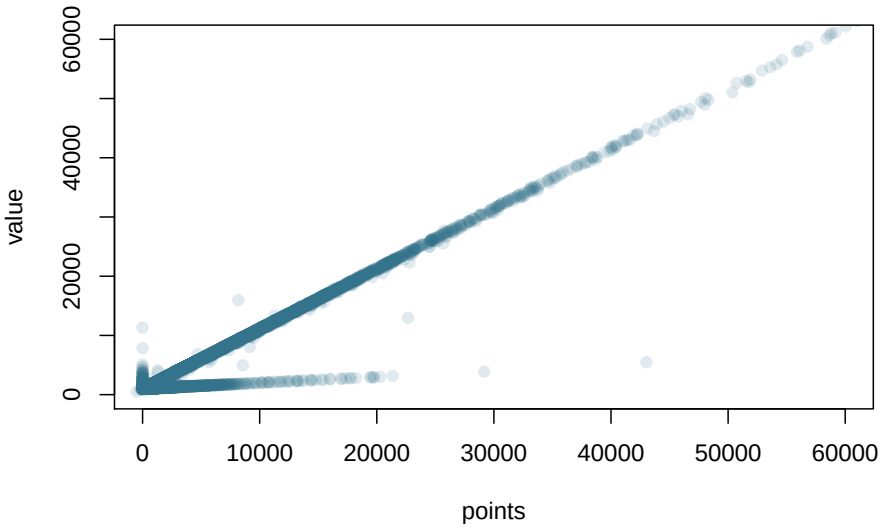
(a) Your objective for 2021 is to increase the customer `value`. In order to do that, you first aim to understand which factors (variables) are the most influential. Write your own linear regression equation (modify the one below), perform the estimation in R, print out your results. You are free to include any variables in `df` and any combination of them in yout model. (1p)

$$value = \beta_0 + \beta_1 var_1 + \beta_2 var_2 + \epsilon$$

(b) Due to the ongoing situation, your budget is now limited that you can only invest on one factor. What would it be? Why? (1p)

(c) Your boss would like a further investigation on the most valuable customers (whose `value`'s is more than 10,000). Perform (a) and (b) again on these customers. (1p)

(d) Based on the figure below, your colleage suggest that `point` causes `value`. Therefore, the retailer should give customer more points to increase their value. What do you say about this insight? (1p)



Have fun and good luck!
Huong and Emelie