NDH802 - Assignment 2

Group no.

- The assignment includes 2 questions, 6 points each.
- Question 1 is empirical (based on the data provided). Question 2 is theoretical (based on the assumption in the question).
- Submit your assignment via Canvas before 15:00 CET, April 27, 2021.
- Your submission should be an RMarkdown file with your solutions in words and/or R code. If you handwrite parts of your assignment, insert it as an image near the corresponding question(s). Name the file NDH802_Assignment2_GroupNumber.
- You should work in groups and contribute equally.
- You can copy my code, but make sure you understand it.
- You should not have the exact solutions and/or results with other groups.
- Results without code/justifications will not be graded.

Set things up

Set your working directory and fill in your group number. For example, if you are group 3, make it `our_group <- 3`. If you don't fill in your group number or fill in the wrong number, your assignment will **not** be graded.

```
#setwd("")
our_group <- 27
```

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. Hereby each and every group should have a unique `df`. Accordingly, your results should be different from other groups' and you should not be comparing them.

```
inference_dataset <- read.csv("https://cda.hhs.se/NDH802data.csv")
set.seed(our_group); df <- inference_dataset[sample(1:nrow(inference_dataset),
                           size = 1000000,
                           replace = FALSE), -1]; rm(inference_dataset)
```

Question 1.

Based on your data,

(a) What is the probability that a customer made no online transactions? Formally, let X denote the number of online transactions, find $P(X = 0)$. (1p)

(b) What is the probability that a customer made at least 3 online transactions? Formally, let X denote the number of online transactions, find $P(X \geq 3)$. (1p)

For Question 1c-e, assuming you randomly and independently sample 10 customers from your `df`.

(c) How many customers can you expect to make at least 3 online transactions? (1p)

(d) What is the probability that exactly 5 of them made at least 3 online transactions? (1p)

(e) What is the probability that no more than 5 of them made at least 3 online transactions? (1p)

Question 2.

Let B denotes the (true) average online basket value in SEK. Assume $B \sim N(\mu = 700, \sigma^2 = 150^2)$.

a. What is the probability that a random online basket is exactly 650 kr? Formally, find $P(B = 650)$. (1p)

b. What is the probability that a random online basket is less than 650 kr? Formally, find $P(B < 650)$. (1p)

c. What is the probability that a random online basket is more than 650 kr and less than 750kr? Formally, find $P(650 < B < 750)$. (1p)

d. Find the shortest range such that the probability is 50% that the average online basket values will fall in this range. Formally, find the shortest range $[b_1, b_2]$ such that $P(b_1 < B < b_2) = 0.5$ (1p)

Assume you randomly and independently sample 2 online baskets.

e. What is the probability that both of them are less than 650 kr? (1p)

f. What is the probability that one of them is less than 650 kr and the other one is from 650kr to 750kr? (1p)

Have fun and good luck!
Huong and Emelie