Applied Statistic with R

Fall 2019, ASDS, YSU

Happy first Stat Homework!
Hope you will enjoy solving our HW problems ☺

Homework Info

You will have both theoretical problems, and practical problems to be solved by using **R**. Please submit the paper version of your homework by the deadline, and submit your program codes in .R format file by the end of the same day using the Google Drive folder that instructor will share with you.

Applied Statistic with R

Fall 2019, ASDS, YSU

Homework No. 01

Due time/date: Section B: 9:28 PM, 11 September, 2019

Note: Please use **R** only in the case the statement of the problem contains (R) at the beginning. Otherwise, show your calculations on the paper. Supplementary Problems will not be graded, but you are very advised to solve them and to discuss later with the Instructor.

Problem 1. Describe the Population and a Sample for each of these cases. Suggest also a **concrete** Sampling Method[1]. Give an example of an Observation. Describe the Variables.

    a. We want to know the percentage of smokers in Armenia.

    b. We want to know the percentage of female smokers in Armenia.

    c. We want to study the relationship between the years of study and salary in Armenia.

    d. We want to study how dangerous are right-hand-drive (right-wheel) cars in the sense of accidents in Armenia.

Problem 2. Give a real-life example for each of the following type data:

    a. Univariate (1D), numerical variable;

    b. Univariate (1D), categorical variable;

    c. Bivariate (2D), both variables are numerical;

    d. Bivariate (2D), both variables are categorical;

    e. Multivariate (>2D), both numerical and categorical variables.

Problem 3. We want to make a statistic about the mean age and salary of all workers in Armenia. We choose 200 profiles of Armenians at random in Linkedin, write a message to the corresponding person and ask about if the person is working, and about the age and salary (also promising to keep the data confidential). Then we calculate means for obtained answers. Will the result be acceptable? Is this sampling method acceptable for the task? Explain your reasoning.

Problem 4. We have a population of size N, and want to choose a sample of size n ($n < N$).

    a. How many different samples we can have?

---

[1]Not of the form: I will use Systematic Sampling, but in the form: I will send emails to all ASDS program students asking the following questions ...

b. We will use random sampling (the probability of any object/person to be chosen from the population is the same). What is the probability that some fixed object/person will be chosen in our sample?

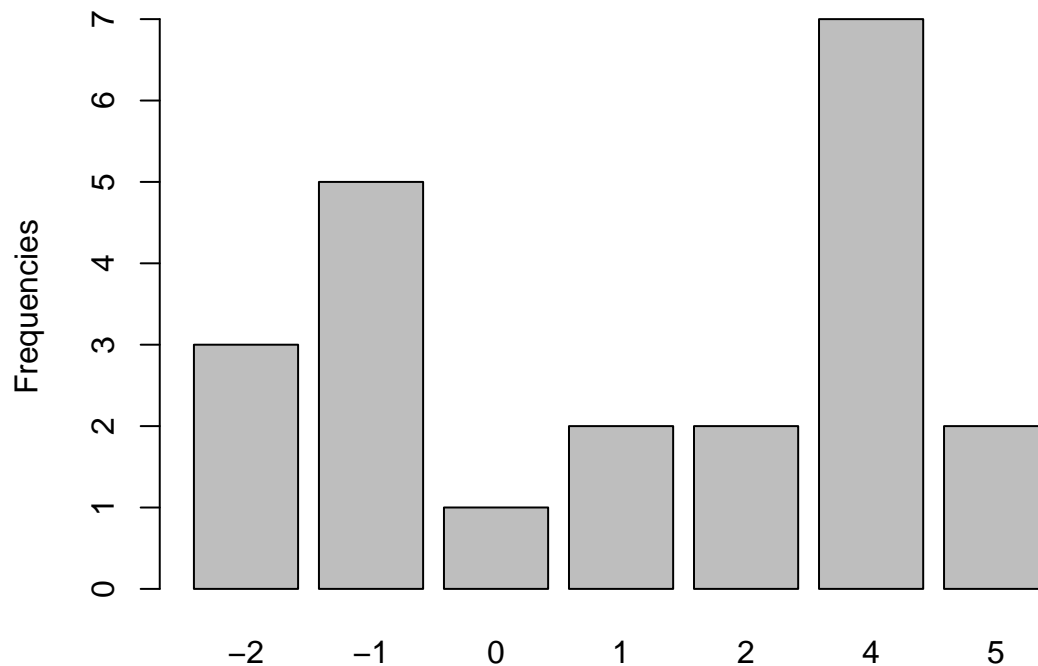**Problem 5.** Fig. 1 gives the BarPlot for some dataset:

Figure 1: Barplot for a dataset

a. How many observations we have?
b. What is the most frequent value in the dataset?
c. Construct the Frequency and Relative Frequency table for the dataset.
d. Can this data represent observations on the variable, which is, supposedly, normally distributed?

Problem 6. (**R**) Write an **R** code to:

– Show the first 3 and last 3 observations (rows) of the standard dataset *mtcars*;
– Find in this dataset the car with the maximal Horsepower;
– Calculate the Frequency Table for the Number of forward gears;
– Plot the BarPlot, Line Graph and the Polygon Plot for the *gear* Variable

Problem 7. (Supplementary) We want to make a survey concerning the popularity of cheating at AUA: we want to know how many students used cheating at least once during their study. Well, although we will not record any names, there is a good chance that some students will not answer honestly to the question. Give suggestion how to design the survey to get acceptable results. Explain your reasoning.

Problem 8. (Supplementary) We have asked two persons to make a coin-tossing experiment (120 tosses) to obtain a random sequence of H, T-s of length 120. The recorded responses are:

$$HHTTHTTHTHHTTHTTTHHTHTHTTHHTHT$$

$$THTHTHHTHHTTHHTHHHTHTTHTTHHTHT$$

$$THHTHTTHHTHTHHTTHHHTTHTHTHTTTH$$

$$THTHTHTHTHHTTHTTHTHHTTHTTHHTHH$$

and

$$HHHTTTHHHTHTHHHHHHTHTTTHHTTHHT$$

$$HTHTTTHHTHTTHHHTHTTTTHTHHHTHHT$$

$$THHHTHHTTHHHHHHTHHTHTHHHTTTTTT$$

$$TTHHTHHHTHHHHTHHHHTHTHHTHHTHTH$$

One of the persons sent a fake sequence (was too lazy to perform the experiment). Who? Explain!

Problem 9. (Supplementary, **R**) Write an **R** function that will make the Relative Frequency Table for a given 1D numerical Dataset x.