Assignment #3
October 22, 2022

This assignment is intended to test your understanding of the collection of data. Assignments should be submitted as a digitally generated LaTeX document (questions 9-11 can be done by hand in class). The datasets can be found on the jupyter server and listed below:

- movies.csv
- pokémon.csv

1. What is meant by bias in a sampling method?
2. Using a for loop write a script in R that generates the sampling distribution for the average Pokémon attack using each of the following sampling techniques:
   (a) Simple random sample
   (b) Stratified random sample
3. Using a for loop write a script in R that generates the sampling distribution for the variance in movie runtimes using each of the following sampling techniques:
   (a) Simple random sample
   (b) Stratified random sample
4. Taylor Swift wants to know how much fans enjoy her music. The section of her concert she wants to survey has 50 seats, 5 rows and 10 columns. The stage runs along the North edge of the venue. The fans rate the concert from 1-100, the scores are shown below:

Stage									
92	89	90	88	95	100	98	93	95	84
82	86	90	88	86	91	90	89	85	83
80	74	80	67	81	82	76	77	74	65
72	68	74	73	70	69	72	70	68	67
69	67	68	68	64	66	63	63	70	68

   (a) Discuss which sampling techniques could be used to select a sample of 10 people.
   (b) Write an R script that generates the sampling distribution of the scores for 100 simple random samples.
   (c) Write an R script that generates the sampling distribution of the scores for 100 stratified random samples (by row).
   (d) Write an R script that generates the sampling distribution of the scores for 100 stratified random samples (by column).
   (e) Compare the distributions from parts b-d side by side, and discuss which is preferred.
5. What is the difference between an observational study and an experiment? Describe a scenario when an experiment would be preferred over an observational study.

6. Alzheimer's disease results in a loss of cognitive ability beyond what is expected with typical aging. A local newspaper published an article with the following headline;

   Study Finds Strong Association Between Smoking and Alzheimer's

   The article reported that a study tracked medical histories of 21,123 men and women for 23 years. The article stated that, for those who smokes at least two packs of cigarettes a day, the risk of developing Alzheimer's disease was 2.57 times the risk for those who did not smoke.

   (a) Identify the explanatory and response variable in the study.

   (b) Is the study described an observational study or an experiment.

   (c) Exercise status (regular weekly exercise versus no regular weekly exercise was mentioned in the article as a possible confounding variable. Explain how exercise status could be a confounding variable in the study.

7. A preliminary study conducted at a medical centre in St. Louis has shown that treatment with small, low-intensity magnets reduces the self reported level of pain in polio patients. During each session, a patient rested on an examining table in the doctor's office while the magnets, embedded in soft pads, were strapped to the body at the site of pain. Sessions continued for several weeks, after which pain reduction was measured.

   A new study is being designed to investigate whether magnets also reduce pain in patients suffering from herniated disks in the lower back. One hundred male patients are available for the new study.

   (a) Describe an appropriate design for the new study. Your discussion should address treatments used, methods of treatment assignment, and what variables would be measured. Do not describe how the data would be analyzed.

   (b) Would you modify the design above if, instead of 100 male patients, there were 50 male and 50 female patents available for the study? If so, how would you modify your design? If not, why not?