**HOMEWORK 1**

*Introductory Statistics*

For this reading and assignment, use the free online textbook, *OpenIntro Statistics, 4th Edition*. To see the textbook, go to <https://www.openintro.org/book/os/>. Click on *Free – OpenIntro Statistics PDF*, and then click on *Read Free Sample* on the left. A pdf version of the textbook should open.

Reading: This assignment focuses on content from Chapters 1 and 2.1. Read all of Chapter 1 and Section 1 of Chapter 2.

Assignment: Questions 1-6 come from the textbook. If you are unsure about what the book is really asking about, try to answer some of the similar odd numbered questions. Answers to the odd numbered questions are at the end of the book, starting on page 384.

1. 1.4 – Buteyko method, study components. (page 19)

* 1. Main research question:
     1. Does the Buteyko Method significantly reduce asthma symptoms?
  2. Cases (same as subjects/units):
     1. 600 asthma patients aged 18-69 who relied on medication
  3. List all variables, state if they are quantitative or categorical, and state if they are discrete, continuous, nominal, or ordinal.
     1. Number of participants, continuous. Age, continuous. Breathing method assignment, nominal. Quality of life, discrete. Activity, discrete. Asthma symptoms, discrete. Medication reduction, discrete.

2. 1.16 – Stealers, scope of inference. (page 29)

1. Intended population:
   1. Everyone with a socio-economic class.
2. Actual population:
   1. Berkeley undergraduates
3. Sample:
   1. 129 Undergraduates
4. Can we generalize to the intended population? Explain.
   1. No. Students at Berkeley do not properly represent everyone.
5. Can we establish causal relationships? Explain.
   1. No. This is not an experiment

3. 1.18 – Cats on YouTube. (page 30)

The text lists 4 quantities about this study. These are labelled a through d. For each identify which is a unit (i.e., case or observation), a variable, a statistic, or a parameter?

1. Parameter

b. Statistic

c. Unit

d. Variable

4. 1.34 – Exercise and mental health. (page 35)

1. Type of study?
   1. Controlled experiment
2. What are the treatment and control groups?
   1. Treatment: exercise twice a week
   2. Control: don’t workout
3. Does this study use blocking?
   1. Yes
4. If so, what is the blocking variable?
   1. Age
5. Do the results establish a causal relationship between exercise and mental health?
   1. Yes. The sampling was supposedly random, and that the two groups were assigned randomly
6. Can the conclusions be generalized to the population at large?
   1. No, sample size is not large enough

5. 2.16 – mean or median (p 59) This problem gives 4 scenarios.

This question only considers scenario (a).

1. Is this distribution symmetric, right skewed, or left skewed?
   1. Right skewed
2. Which better represents a typical observation: the mean or the median?
   1. Median. There is a “meaningful number of houses that cost more than $6,000,000”
3. Which better represents the variability in the observations: the standard deviation or the IQR?
   1. IQR. Data is not normally distributed. Has outliers

6. 2.16 mean or median (p 59). This question only concerns scenario (b)

1. Is this distribution symmetric, right skewed, or left skewed?
   1. Symmetric
2. Which better represents a typical observation: the mean or the median?
   1. Mean. Relatively normal data
3. Which better represents the variability in the observations: the standard deviation or the IQR?
   1. Standard deviation. Normalish distribution

Questions 7 and 8 are based on the following study and dataset.

A survey was conducted of students in an introductory statistics course.[[1]](#footnote-1) The variables we will look at are *Exercise*, the number of hours exercised per week, and *SAT*, which is their SAT score. The SAT is a college admission exam, like the ACT, popular outside the Midwest. The data set is StudentSurvey.jmp, available on the Assignments canvas page.

Read the data into JMP and create a histogram and numerical summaries (descriptive statistics) for both variables, *Exercise* and *SAT*. The JMP *Analyze, Distribution* is one way to obtain these results, but there are others.

7. Using the *Exercise* variable:

1. Describe the shape of its distribution
   1. Right skeweed
2. Describe the location (center) of this distribution using the most appropriate measure.
   1. Median is 8
3. Describe the spread of this distribution using the most appropriate measure.
   1. Interquartile range is 7

8. Using the SAT variable:

1. Describe the shape of its distribution
   1. Roughly normal, symmetric
2. Describe the location (center) of this distribution using the most appropriate measure.
   1. Mean is score 1204
3. Describe the spread of this distribution using the most appropriate measure.
   1. Standard Deviation is 121
4. Interpret the value of the standard deviation in the context of this study.
   1. The average difference of SAT scores from the average is roughly 121.
   2. Roughly 68% of scores is within one standard deviation of the mean

1. http://www.lock5stat.com/datapage.html [↑](#footnote-ref-1)