**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 right. 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: Complete the following questions from the textbook. Note that the answers to the odd numbered questions are at the end starting on page 384.

1. 1.3 – Air pollution and birth outcomes, study components. (page 19)
   1. Main research question:

Is there a correlation between air pollution exposure and having premature babies?

* 1. Units (same as cases /subjects):

143,196 births in years 1989-1993 in Southern California

* 1. List all variables, state if they are quantitative (numerical) or categorical, and state if they are discrete, continuous, nominal, or ordinal.

Continuous quantitative variables: carbon monoxide, nitrogen dioxide, and ozone measurements. Coarse particulate matter and length of gestation.

1. 1.4 – Buteyko method, study components. (page 19)
   1. Main research question:

Does the Buteyko method improve symptoms for asthma patients?

* 1. Units (same as cases /subjects):

600 asthma patients who rely on medication for asthma treatment and are of age 18-69.

* 1. List all variables, state if they are quantitative or categorical, and state if they are discrete, continuous, nominal, or ordinal.

Categorical, nominal: treatment group

Categorical, ordinal: quality of life, activity level, asthma symptoms, medication reduction. Because of being grouped into categories 0-10.

1. 1.15 – Buteyko method, scope of inference. (page 29)
   1. Intended population:

All asthma patients of age 18-69 who rely on asthma medication for treatment.

Actual population:

600 asthma patients of age 18-69 who rely on asthma medication for treatment

Sample:

600 asthma patients aged 18-69 who rely on asthma medication for treatment used in the study

* 1. Can we generalize to the intended population? Explain.

This depends. If the sample is randomly selected and can represent the entire population of asthma patients of age 18-69 who use medication, then the results can be generalized.

Can we establish causal relationships? Explain.

This study is experimental because the patients were randomly assigned so a casual relationship can be established.

1. 1.16 – Stealers, scope of inference. (page 29)
   1. Intended population:

All University of California Berkely undergraduates

Actual population:

129 University of California Berkely undergraduates

Sample:

The 129 University of California Berkely undergraduates used in the study

* 1. Can we generalize to the intended population? Explain.

This depends. If the undergraduates in the study were randomly sampled and can represent all of UC Berkely undergraduates, then the results can be generalized.

Can we establish causal relationships? Explain.

This study is observational so a casual relationship can’t be established.

1. 1.25 – Haters are gonna hate, study confirms. (page 31)
   1. Units (same as cases):

200 randomly sampled men and women

* 1. Response variable(s)

The participants’ reaction to the fictional microwave oven

* 1. Explanatory variable(s)

Dispositional attitude

* 1. Random sampling?

Yes.

* 1. Observational study or experiment? Explain.

Observational study because there is no random assignment to treatment.

* 1. Can we establish a causal link? Explain.

No, the participants weren’t randomly assigned.

* 1. Can we generalize to the population at large? Explain.

Yes. The participants were randomly sampled so the results of the study can be generalized.

1. 2.17: Income at the coffee shop. (page 59)
   1. Mean or median? Explain.

The median is better. The two new people who go to the coffee shop with $225,000 and $250,000 are the outliers to the other 40 patrons. The mean is substantially affected by the two extreme observations. The median doesn’t get effected as much since it is robust to outliers.

* 1. Standard deviation or IQR? Explain.

The IQR is better. The two extreme observations would affect the standard deviation and cause it to increase showing that it is not robust. The IQR is robust to outliers. So it best represents the amount of variability in the incomes of the 42 patrons at the coffee shop.

(TURN OVER)

Answer questions 7-10 based on the following scenario and dataset.

Students were surveyed in an introductory statistics course.[[1]](#footnote-1) The variables we will look at are *Pulse* which refers to pulse rate in beats per minute of the student and *Piercings* which is the number of piercings the student has.

1. Use JMP to create a histogram and numerical summaries (statistics) for both variables, *Pulse* and *Piercings*. Include a copy of your output either below or at the end of your assignment. For information on how to copy or print output from JMP, go to Canvas, click on the JMP module, and open the JMP-general document. The JMP Preferences document may also be of use to you. Note that you will likely use the *Analyze, Distribution* analysis in JMP, however there are other ways to get the same results.

A picture containing chart

Description automatically generated

1. Describe the distribution of *Pulse* using the histogram and appropriate statistics. Note you should consider the shape when choosing which statistics to report.

The shape of the distribution of *Pulse* is unimodal, bell-shaped and symmetrical, thus approximately normal. The center (mean) *Pulse* is approximately 70 beats per minute. The range of the middle 50% of beats per minute is 16 (IQR). In this distribution, there are 4 outliers total, 2 that are higher and 2 that are lower.

1. Describe the distribution of *Piercings* using the histogram and appropriate statistics.

The shape of the distribution of *Piercings* is unimodal and skewed to the right. The center (median) is 0 piercings. The range of the middle 50% of piercings is 3 (IQR). In this distribution, there are 4 outliers.

1. What is the standard deviation of *Pulse*? Interpret the value of the standard deviation in context.

The standard deviation of *Pulse* is 12.21. The average difference from the mean pulse of 69.57 of students surveyed is 12.21.

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