

## Group Work - Chapter 1

1 The Tortilla and Cheese Organization (TACO) is a taco advocacy and lobbying group. They want to understand the relationship between taco consumption and grades. They begin by randomly selecting one stats class, one writing class, one business class and one nursing class at Metro State. They go to one class each week over four weeks (the stats class the first week, the writing class the second, etc) and ask each student attending the class their age, gender and "Knowing that tacos are delicious, how many tacos have you eaten in the past month?" At the end of the semester, they get the average grade for each class.

After analyzing the data, TACO finds that the stats class has a class GPA 0.1 higher than the other classes and had consumed more tacos, on average, than the other classes. The results were statistically significant. They issue a press release declaring, "Eating tacos leads to higher grades."

- (a) Identify the research question, the population being studied, the data used to answer the question and any potential problems or pitfalls with the study.

**Research question:** Is there a relationship between taco consumption and grades?

**Population studied:** Perhaps the intended population is college students, but the sample is drawn from Metro State students.

**Data:** Class GPA, tacos consumed in past month

**Problems or pitfalls:** There are many possible problems in this study, including...

- The data collected can't answer the research question
- Results are not practically significant
- Loaded question
- Self-reported data

- (b) For each variable used by the study, identify the type (quantitative/categorical, discrete/continuous, etc) and the level of measurement.

**Class GPA:** Quantitative, continuous, interval

**Number of tacos:** Quantitative, discrete, ratio

- (c) Identify the type of study and the sampling method used. For observational studies be sure to specify the type and for experimental studies assess whether good design guidelines were followed. Is the sample a simple random sample?

**Observational study, cross-sectional**

**Stratified/cluster sampling, not a simple random sample**

**2** A medical research company has developed what they hope is a promising new drug to treat a disease which has a life expectancy of 2 years after diagnosis. After conducting safety trials, they conduct a efficacy study. At four urban medical centers in different region of the US, they recruit a total of 176 patients who are newly diagnosed with the disease. The patients are randomly selected to either receive the new drug or to receive a placebo. All patients continue to receive the current standard of care. Neither the patients nor their doctors know which group they are in. Every patient in the study has a doctor visit every other month where vital measurements are recorded such as weight, blood pressure, blood sodium levels, as well as life style information such as marital status, education level and alcohol consumption. Also, if a patient dies during the study, the time between diagnosis and death is recorded.

After a six year study, 39 of the patients dropped out of the study or lost contact with the researchers. It is found that patients in the treatment group lived an average of 13 months longer. A parallel NIH funded study conducted at a university is still ongoing.

- (a) Identify the research question, the population being studied, the data used to answer the question and any potential problems or pitfalls with the study.

**Research question:** Does the new drug increase survival time for patients with the disease?

**Population studied:** Patients with the disease, particularly those being treated at the selected medical centers.

**Data:** Treatment (new drug or placebo), months until death or survival until end of study, bi-monthly measurements of weight, blood pressure, etc. and marital status, education, etc.

**Problems or pitfalls:** Possible problems...

- Generalizability (only patients who live near cities, for example)
- Study conducted by drug company, possible bias
- High percentage of missing data

- (b) For some of the variables used by the study, identify the type (quantitative/categorical, discrete/continuous, etc) and the level of measurement.

**Treatment:** Categorical, nominal

**Survival:** Categorical (died or did not die), nominal and quantitative (time since start of study of death), continuous, ratio

**Weight:** Quantitative, continuous, ratio

**Education level:** Categorical, ordinal

**Alcohol consumption:** Could be categorical, nominal (drinks or does not drinks), or categorical, ordinal (low, medium or high consumption) or quantitative, discrete, ratio (number of drinks in past week)

- (c) Identify the type of study and the sampling method used. For observational studies be sure to specify the type and for experimental studies assess whether good design guidelines were followed. Is the sample a simple random sample?

**Experimental, has replication, blinding and randomization**

**Convenience sample, not a simple random sample**

**3** Wanting to better understand the risk factors associated with various kinds of heart disease, researchers at the National Heart, Lung, and Blood Institute selected a small town in Massachusetts. Five thousand residents were selected for the study, an equal number of men and women and representative proportions for age groups. Every two years, study subjects undergo a detailed medical history, physical examination, and medical tests. Data collected include disease history for coronary heart disease, stroke and mental illness, vital stats such as weight, blood pressure and body temperature, lifestyle information such as history of smoking, diet and exercise, and age, gender and other biographical information.

The study is ongoing, but some of the results from it include smoking is associated with an increase risk of heart disease and exercise is associated with a decreased risk of heart disease, high blood pressure increases risk of stroke, and mental illness affects heart health.

- (a) Identify the research question, the population being studied, the data used to answer the question and any potential problems or pitfalls with the study.

**Research question: How are biological and lifestyle factors associated with heart and lung diseases?**

**Population studied: In the most generous light, the population is adults. However, since only a specific town is studied, it would not be incorrect to consider the population U.S. adults, or Massachusetts adults, or small town adults, etc.**

**Data: Many types of data are gathered, disease status, biological measurements and lifestyle indicators.**

**Problems or pitfalls: Possible problems...**

- **Generalizability (only one town is studied, it is unclear how representative it is)**
- **Some self-reported data**
- **Possible missing data**

- (b) For some of the variables used by the study, identify the type (quantitative/categorical, discrete/continuous, etc) and the level of measurement.

**Similar to previous question, many possible types of data and some, like diet or exercise, could be recorded in multiple ways.**

- (c) Identify the type of study and the sampling method used. For observational studies be sure to specify the type and for experimental studies assess whether good design guidelines were followed. Is the sample a simple random sample?

**Observational study, prospective**

**Cluster sampling (assuming they choose the town at random), not a simple random sample**