Homework 1 Answer Key

DATA1220-55, Fall 2024

Sarah E. Grabinski

2024-09-19

Objectives

Homework 1 was assigned on date, due on date, and was worth 30 points. The material covered in Homework 1 was drawn from Chapter 1, *Introduction to Data*, of the textbook OpenIntro Statistics available on Canvas. This chapter discusses the basics of how to best collect, analyze, and draw conclusions from data. In addition, Homework 1 introduced the use of RStudio to create publication-ready data analyses. The objectives of Homework 1 were as follows.

- Help Sarah get to know you and better understand the specific needs of the class
- Register for Campuswire, where you will earn many of your participation points
- Identify different populations and sampling strategies
- Describe the reliability, validity, and generalizability of different types of data
- Perform a basic data analysis
  - Identify variable types
  - Visualize and interpret data
  - Communicate results
- Become familiar with creating markdown documents with R code in RStudio

Problem 1 - Survey

Survey

A Google Forms survey was sent to your JCU email. I estimate it will take 5-10 minutes to complete.

Points: 5

- All students completed the survey, so all students received 5 points.

Campuswire

Instructions for registering to our class Campuswire forum were sent to your JCU email. I have also posted our first discussion topic. Interacting with it will earn you participation credit. Make sure to check your notification settings so you don't miss anything you want to interact with!

Points: 5

- Students who signed up for Campuswire *and* interacted with the first discussion post received 5 points.
- Students who signed up for Campuswire but did *not* interact with the first discussion post received 2.5 points.

Problem 2 - Interpreting Studies

The Studies

Researchers in the UK wanted to answer the question of how much crime there was in Britain and whether it was going up or down. They used 2 different approaches to gather data for their investigation, but they need help determining the validity of their approach.

Data Set 1

The Crime Survey for England and Wales is a survey in which approximately 38,000 people are questioned about their experiences with crime. People surveyed are 16 years of age or older and were not living in communal residences. Answers are self-reported.

Data Set 2

UK Police keep administrative records of crimes they have investigated. Police use internal definitions of crimes and their discretion when creating these records.

Questions

1. In 1 sentence each, describe the study population of the data sets. (Points: 2)

- Students received 1 point if their answer for data set 1 described the study population as people 16 years of age or older not living in communal residences in England and Wales.
- Students received 1 point if their answer for data set 2 described the study population as people who report crimes to the police or who experience crime in the UK.

2. In 1 sentence each, describe the sampling strategy of the data sets. (Points: 2)

- Students received 1 point if their answer for data set 1…
- Students received 1 point if their answer for data set 2 indicated that no sampling strategy was used, as they have access to all the data that is available.

3. In 1 sentence each, describe the sampled population of the data set. (Points: 2)

- Students received 1 point if their answer for data set 1 described the study population as 38,000 people 16 years of age or older not living in communal residences in England and Wales.
- Students received 1 point if their answer for data set 2 indicated that no sampling strategy was used and/or that the data only contains crimes that were reported to police.

4. In 1 sentence, describe the target population of the study (Points: 1)

- Students received 1 point if their answer indicated that the target population was potential victims of crime in Britain

5. In a short paragraph (3-6 sentences), please describe… (Points: 3)

```
***a.***  ***the reliability of each data set***

***b.***  ***the validity of each data set***

***c.***  ***if conclusions based on each data set from the study population are generalizabl
```

- Students received all 3 points if they…
  - Indicated both data sets were self-report, which affects their reliability
  - Indicated data set 2 is missing unreported crimes and/or that police decide what qualifies as a crime, which affects its validity
  - Indicated that conclusions from data set 1 were not generalizable to people under the age of 16
  - Indicated that conclusions from data set 2 were not generalizable to unreported/underreported crimes and those likely to be victims of them

Problem 3 - Interpreting Data

The Data

The Child Health and Development Studies investigate a range of topics. One study, in particular, considered all pregnancies between 1960 and 1967 among women in the Kaiser Foundation Health Plan in the San Francisco East Bay area.

Questions

1. Read in the .csv document (Points: 1)

- Students received 1 point if the output in their HTML document indicates that they properly loaded the `.csv` file.

2. Print a summary of the data. (Points: 1)

- Students received 1 point if the output in their HTML document contains the numerical summary of the data created by the `describe()` function from the `Hmisc` package.

3. Complete the data dictionary. (Points: 3)

- Students received all 3 points if they…
-

4. Add the name of an explanatory (i.e. independent) variable to the x-axis of the plot and a response (i.e. dependent) variable to the y-axis of the plot. In 1 sentence, describe what you see. (Points: 3)

- Students received all 3 points if they…

  - Changed both the x and y variables shown in the plot as requested in the instructions
  - Successfully embedded the plot in their HTML document
  - Appropriately described the relationship between the two variables as positive, negative, or independent

5. BONUS: Add features such as titles, axis labels, colors, shapes, etc. to enhance your data visualization (Points available: 2)

- Students received # additional point for adding a regression line to help visualize the relationship between the two variables
- Students received # additional points for adding text features like titles and labels
- Students received # additional points for adding aesthetic features like colors and shapes
- Students received # additional points for adding anything to the plot which was not included in one of the previous lectures

6. Render your document as an HTML file (Points: 2)

- Students received 2 points if their document was properly rendered and submitted as a `.html` file.
- Students received 1 point if there was still a problem with their `.html` document that caused it to look meaningfully different from the example on Canvas

Last Update

This document was last updated on 2024-09-20.