Homework 1

Due: Monday 2024-09-09

DATA1220-55, Fall 2024

Table of contents

Below are details on Homework 1, covering the introduction to R and RStudio and Chapter 1 of OpenIntro Statistics. This homework is due by 6:00pm on Monday, 9/9/24. No credit will be lost for assignments received by 7:00pm to account for issues with uploading. 10% of the points will be deducted from assignments received by 9:00am on Tuesday, 9/10/24. Assignments turned in after this point are only eligible for 50% credit, so it benefits you to turn in whatever you have completed by the due date.

A Quarto template has been posted on Canvas for you to use for this assignment. Please use it. You can learn more about customizing Quarto documents here: https://quarto.org/docs/get-started/hello/rstudio.html

Please render your document as HTML and submit ***BOTH*** the `.html` file and `.qmd` file to the Homework 1 assignment on Canvas.

Problem 1 - Survey

Objectives

- 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

Survey

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

Points: 5

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

Note: I have accidentally set this up from my Case Western email. I don't believe that will affect you all, but please let me know if there are any issues.

Problem 2 - Interpreting Studies

Objectives

- Identify populations
- Identify sampling strategies
- Describe the reliability, validity, and generalizability of different types of data

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)
2. In 1 sentence each, describe the sampling strategy of the data sets. (Points: 2)
3. In 1 sentence each, describe the sampled population of the data set. (Points: 2)
4. In 1 sentence, describe the target population of the study (Points: 1)
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 generalizable to the target population

Problem 3 - Interpreting Data

Objectives

- Incorporate text and code into a polished document
- Read in a .csv (Comma Separated Values) data set
- Create a data dictionary
- Use `ggplot2` to create a plot and interpret it

The template has been partially completed to help you with this portion of the assignment. Please fill in the missing portions.

You may find helpful hints on the cheat sheets for the bonus portion here: https://ggplot2.tidyverse.org/

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)
2. Print a summary of the data. (Points: 1)
3. Complete the data dictionary. (Points: 3)
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)
5. BONUS: Add features such as titles, axis labels, colors, shapes, etc. to enhance your data visualization (Points available: 2)
6. Render your document as an HTML file (Points: 2)

Reminder

My office hours are MW 2:30-4:00pm immediately after class and Friday by appointment in Dolan E252. You may also post questions to crowdsource help (and answer them for your classmates! and upvote them!) on our Campuswire forum, earning additional participation credit in the meantime.

Updates

This document was last updated on 2024-09-04.