

Problem Set 1

POL 501

September 17, 2024

Instructions

Contents Covered

- *From the textbook*: Chapters 1 and 2 (Introduction to Data, Summary Statistics, and Graphs).
- *R Lessons*: Introduction to R, Dataframe Manipulation, Summary Statistics, and Graphs.
- See the PDF file “*Glossary of Terms and Concepts*” for a review of key theoretical concepts.

General Instructions

- *Justify all of your answers.* Failing to do so will have score deductions.
- If you use formulas, always write them down and justify their appropriateness in each case.
- **Unless otherwise instructed**, you can use calculators or spreadsheet software to compute your answers (that is, to come up with the accurate calculation of a given arithmetic operation).
- Yet, you **must clearly describe how you arrived at each result by stating the appropriate formula**. Correct “numbers” in answers, but without justification or computed with the wrong formula, will have score deductions.

Submission Instructions

- You must submit your answers to this problem set as a PDF or Word file generated by R markdown.
- You must use the R Markdown template I uploaded in Brightspace to write your answers.
- You must upload your PDF or Word File in the corresponding assignment in Brightspace.
- **Due Date:** [September 25th](#).

Questions

Question 1

A study exploring the effects of political corruption on tax compliance found that in countries with less political corruption, private firms were less likely to engage in tax fraud.

What type of relationship is described in the study?

- (a) Positive relationship
- (b) Negative relationship
- (c) Non-linear relationship.
- (c) No relationship

Question 2

A study found that as the number of community events increases from a few to a moderate number, community cohesion improves. However, when the number of events increases from moderate to high, community cohesion decreases.

What kind of association is this?

- (a) Positive association
- (b) Negative association
- (c) Non-linear association
- (d) The variables are independent of each other

Question 3

A study of one recent primary for the Republican party revealed the following data. Researchers were surprised to observe the results.

Candidate Names	Total Votes Won	Campaign Spending (\$)
Candidate A	10,000	500,000
Candidate B	15,000	300,000
Candidate C	8,000	700,000
Candidate D	12,000	400,000

According to this data table, what kind of association is found between “spending” and “votes”? (Hint: Use a scatterplot to determine your answer)

- (a) Positive association
- (b) Negative association
- (c) Non-linear association
- (d) The variables are independent of each other

Question 4

Consider the level of satisfaction with local government services as expressed through a survey using ratings: very unsatisfied, unsatisfied, neutral, satisfied, very satisfied. This question assesses the perceived effectiveness of services such as public transportation, parks, and emergency responses.

Which kind of variable type is this?

- (a) Regular Categorical (Nominal)
- (b) Ordinal Categorical (Ordinal)
- (c) Numerical (Discrete)
- (d) Numerical (Continuous)

Question 5

In a psychological study examining stress triggers, participants were categorized by their primary work environment settings, such as 'open-plan offices', 'private offices', and 'remote work from home'. Researchers sought to determine if these settings influenced reported stress levels during work hours.

Which is the explanatory variable and which is the response?

- (a) Work environment setting is the response, and stress level is the explanatory variable.
- (b) Stress level is the response, and work environment setting is the explanatory variable.

Question 6

For this question, you are **expected to use R**.

A company surveyed different neighborhoods to measure the level of social trust and the number of social activities. Social activities include concerts, social benefit events, sports, and other community gatherings. Social trust is defined as the level of trust residents have in their community and neighbors, measured on a scale from 0 to 100. The results are shown below:

Neighborhood	Number of Social Activities (per month)	Social Trust Level (out of 100)
N1	1.0	21.48
N2	2.9	48.90
N3	4.8	76.20
N4	6.7	96.73
N5	8.6	96.87
N6	10.5	98.58
N7	12.4	102.14
N8	14.3	85.35
N9	16.2	59.21
N10	18.1	37.10
N11	20.0	1.2

Table 1: Survey Data from Various Neighborhoods

Parts of Question 6

- a) Calculate the mean, median, and standard deviation for both the number of social activities per month and the social trust levels across the neighborhoods. Explain what these measures indicate about the data distribution of each variable.
- b) Create a histogram for both the number of social activities per month and social trust levels with five bins of equal size. Describe the shape of each distribution.
- c) Plot a scatter plot between the number of social activities and social trust levels. Describe any visible patterns that you find.
- d) Based on your calculations and plots in parts (a) to (c), hypothesize about the relationship between social activities and social trust. According to these data, what is the relationship between social activities and social trust? Can you think of an intuitive explanation of what could be the underlying process that explains this relationship?