

Group Project 2: Chapter 1 Review

Writer: _____ Speaker: _____

Facilitator: _____

Answer the following on a separate sheet of paper. When asked for an example, you should not give an example that you've seen in class or on homework--come up with new ones.

1. Give an example of when the median is the most appropriate measure of center for a data set.
2. Give an example of when the mean is the most appropriate measure of center for a data set.
3. What are the different ways of measuring spread? When do you use them?
4. What is an outlier? How do you recognize an outlier on a histogram? We've talked about various numerical measurements (mean, median, correlation, regression line, etc.) which of them are sensitive to the presence of outliers? Which of them are resistant?
5. Give an example of two variables that are likely to be strongly positively correlated.
6. Give an example of two variables that are likely to be strongly negatively correlated.
7. Give an example of two variables that are likely to have a correlation near zero.
8. Give an example of two variables that have a strong connection to each other but where the correlation will be near zero.
9. Give an example of two variables that are strongly correlated but where there is no causation (a change in one variable does not **cause** a change in the other variable).
10. If you were to calculate correlation by hand, what would you need to know about the data sets? (i.e. what are the ingredients that make up correlation?)
11. What are the ingredients that go into finding a least squares regression line?
12. What is the connection between correlation and regression?
13. What is the difference between what correlation r and correlation squared r^2 tell us?
14. Explain what a 2-way table is and what conditional and marginal distributions are.
15. What is Simpson's paradox?
16. (Extra-Credit) Give an example of Simpson's paradox.
17. On page 156, do problems I.2, I.11, I.15 - I.18