Unit 5E

Correlation and Causality

Slide 5-1

Correlation between two data variables

- No correlation: There is no apparent relationship between the two variables.
- Positive correlation: Both variables tend to increase or decrease together.
- Negative correlation: One variable increases while the other decreases.
- Strength of a correlation: The more closely two variables follow the uphill trend or downhill trend, the stronger the correlation.

Weight and Price of Diamond



A scatter diagram shows that higher diamond weight generally goes with higher price.

Median Age of Marriage



Each point in the scatterplot represents one state in the US and the District of Columbia.

Correlation (Positive or Negative?)



A scatter diagram shows that higher life expectancy generally goes with lower infant mortality.

Describe the plot



This scatterplot shows no trend because the points seem to follow no predictable pattern. For every age group, we can find relatively fast and relatively slow runners. Marathon running speed does not seem to be related to age of runner.

Example: Strength of Association



Is there a stronger association between height and weight or between waist size and weight?

Possible Explanations for a Correlation

- 1. The correlation may be a *coincidence*.
- 2. Both variables might be directly influenced by some *common underlying cause*.
- 3. One variable may be a *cause* of the other.

Consider the negative correlation between infant mortality and life expectancy. Which of the three explanations for correlation applies?

The negative correlation is probably due to a common underlying cause – the quality of health care. In countries where health care is better in general, infant mortality is lower and life expectancy is higher.

Example

- State whether the variables are positively or negatively correlated (or no correlated). State whether the correlation is most likely due to coincidence, a common underlying cause, or a direct cause.
- Height of a person and how frequently he or she attends religious services.

Example

- State whether the variables are positively or negatively correlated (or no correlated). State whether the correlation is most likely due to coincidence, a common underlying cause, or a direct cause.
- State spending on education and SAT scores of high school students.

Example

- State whether the variables are positively or negatively correlated (or no correlated). State whether the correlation is most likely due to coincidence, a common underlying cause, or a direct cause.
- Gas Mileage and tire pressure.