

**Chapter 5: Summarizing Bivariate Data**

## Notetaking Guide

**Section 5.1 – Correlation**

- **Pearson’s Sample Correlation Coefficient**

What does the correlation coefficient represent?

- **Properties of  $r$**

Summarize the five properties of  $r$

- 1.
- 2.
- 3.
- 4.

- **The Population Correlation Coefficient**

What symbol is used to represent the population correlation coefficient?

How is the population correlation coefficient similar to the statistic  $r$ ?

- **Correlation and Causation**

How is association different from causation?

How can causation be established?

Give an example of two variables that have a strong correlation, but where an assumption of causation would be absurd.

**Homework: # 1, 2, 3, 5, 6, 9, 10, 12, 13**

**Section 5.2 – Linear Regression: Fitting a Line to Bivariate Data**

- **Fitting a straight Line: The Principle of Least Squares**

What are the roles given to  $x$  and  $y$  in a regression analysis?

What is meant by the least-squares line or line of best fit?

What is the *danger of extrapolation*?

- **Regression**

Why is the least-squares line often called the *sample regression line*?

What symbol is used to denote the regression line?

What data point always lies on the regression line?

What does regression analysis refer to?

**Homework: # 15, 16, 17, 18, 19, 22, 23, 26**

**Section 5.3 –Assessing the Fit of a Line**

- **Predicted Values and Residuals**

What is a residual and how is it calculated?

- **Plotting the Residuals**

What is a *residual plot* and what is its role in assessing the fit of a line?

In what way could an observation be considered *influential*?

How can you recognize an *outlier* in a scatterplot?

- **Coefficient of determination**

What is represented by the *coefficient of determination* and what symbol is used to denote it?

- **Standard Deviation About the Least-Squares Line**

What is the interpretation of  $s_e$ ?

**Homework: # 29, 31, 32, 33, 36, 41, 42, 43**

### **Section 5.4 – Nonlinear Relationships and Transformations**

- **Polynomial Regression**

When is a polynomial an appropriate model for the data?

How does  $R^2$  for a quadratic regression differ from  $r^2$  for a linear regression?

Give some examples of polynomial regression models and variables that may be modeled by them.

- **Transformations**

Why would it be preferable to transform the data set rather than using a non-linear model?

How is a particular transformation selected and how is it applied to a set of data?

**Homework: # 44, 45, 49, 52, 53**

**Section 5.6 – Interpreting and Communicating the Results of Statistical Analyses**

- **Communicating the Results of Statistical Analyses**

Summarize in a paragraph the key concepts that should be covered in analyzing bivariate data.

- **What to look for in Published Data**

Look for a published article that contains a scatterplot online or in printed material like a magazine or newspaper. Perform an analysis applying the principles from this section in your book.

- **A Word to the Wise: Cautions and Limitations**

Write a brief summary of the things you should keep in mind when conducting or reading a statistical analysis of bivariate data.

**Homework: # 63, 64, 65, 66, 77, 78, 79**