CHAPTER 8 & 9

Reference: http://www.stat.ucla.edu/~vlew/stat10/archival/wi03/lectures/wi031019.pdf

Scatter plot

- • A scatterplot or scatter diagram is a two dimensional plot of data. The horizontal dimension is called x, and the vertical dimension is called y.

- • Each point on a scatterplot or scatter diagram shows two values, an x value and a y value. Each point represents a single case. A single case could be a single person or object, but a single case could be a matched pair (e.g. father-son, twins, husband-wife)

- ***Scatter diagrams only show association, but association does not mean causation***

Correlation

- The CORRELATION COEFFICIENT, denoted r, measures how close the data are to a straight line or in other words it measures the strength of association.

- Formula Your text gives you a very long formula for calculating the correlation coefficient (pp 132-134) and I am not certain how useful it is. Instead, read the technical note on p. 134, the formula is reproduced here:

Properties of the correlation coefficient

- • -1 <= r <= 1
- • If r is close to 1 or -1, the data are close to a line
- • If r is close to 0, the data are not close to a line
- • Pictures! (see pages 127, 129 and 142 of your text)
- • The correlation r measures how close the data are to a line
- • r does NOT tell what percentage of the data fall on the line
- • The correlation between x and y is the same as the correlation between y and x. r(x, y)=r(y, x)
- • Invariant under addition. If some constant "a" is added to every one of the X or the Y values, the correlation is unchanged.
- • Invariant under multiplication: if all of the x or the y values are multiplied by some positive constant "b", the correlation is unchanged. The correlation can change very dramatically if only ONE of the data points is changed.