AP Statistics Chapter 7 Notes #1
Chapter 7: Scatterplots, Association, and Correlation

Scatterplots

- display *quantitative* data and show patterns, trends, relationships, and even the occasional extraordinary value
- often used to determine if there is an ***association*** between two variables
- **direction**
  - negative: a pattern that runs from the upper left to the lower right
  - positive: a pattern that runs from the lower left to the upper right
- **form**
  - linear: a straight line relationship, appears as a cloud or swarm of points stretched out in a generally consistent, straight form
  - if the relationship isn't straight, but curves gently, while still increasing or decreasing steadily, we can often find ways to make it more nearly straight
  - if it curves sharply, there is much less we can say about it
- **strength** (of the relationship)
  - how closely the points follow the pattern
- **unusual features**
  - outliers: look for points standing away from the overall pattern
  - clusters or subgroups: like outliers, deserve special attention

Making a scatterplot:

- Assign each **QUANTITATIVE** variable to an axis
- Clearly label the axes with the variable and its units
- Each point is placed on a scatterplot at a position that corresponds to values of the two variables
- If both variables have values near or on both sides of zero, then the origin will be part of the display.
- If values are far from zero, there's no reason to include the origin (you can draw these with axes that don't quite meet.

TI Tips: p. 149, 155, 160

Explanatory or predictor variable: placed on the x-axis

Response variable: placed on the y-axis

Correlation:

- Measures the strength of the association between two **QUANTITATIVE** variables
- Between 0 and ±1 (positive association between 0 and +1; negative association between 0 and -1)
- NO UNITS! Data is standardized: $(z_x, z_y) = \left(\frac{x - \bar{x}}{s_x}, \frac{y - \bar{y}}{s_y} \right)$
- Standardizing data makes the origin the new center of the scatterplot and the scales on both axes the same

AP Statistics Chapter 7 Notes #1
Chapter 7: Scatterplots, Association, and Correlation

- *Correlation coefficient:*
  - $r = \frac{\sum z_x z_y}{n-1}$ (See p. 152)
  - Summarizes both strength and direction of a **LINEAR** association between two **QUANTITATIVE** variables

Conditions…

- *Quantitative Variables Condition:* Correlation applies only to quantitative variables. Check that you know the variables' units and what they measure.
- *Straight Enough Condition:* Must have a linear association
- *Outlier Condition:* Outliers can distort the correlation dramatically. When you see an outlier, it's often a good idea to report the correlations with and without the point

**Just Checking: p. 153
**Step-by-Step: p. 154-155

Correlation Properties:

- The sign of a correlation coefficient gives the direction of the association
- Correlation is always between -1 and +1. Correlation *can* be exactly -1 or +1, but these values are unusual in real data because they mean that all the data points fall *exactly* on a single straight line
- Correlation treats x and y symmetrically. The correlation of x with y is the same as the correlation of y with x
- Correlation has no units.
- Correlation is not affected by changes in the center or scale of either variable. Changing the units or baseline of either variable has no effect on the correlation coefficient. Correlation depends only on the z-scores, and they are unaffected by changes in center or scale
- Correlation measures the strength of the **LINEAR** association between the two variables. Variables can be strongly associated but still have a small correlation if the association isn't linear
- Correlation is sensitive to outliers. A single outlying value can make a small correlation large or make a large one small

Correlation Tables: see p. 158

Straightening Scatterplots:

- When a scatterplot show a bent form that consistently increases or decreases, we can often straighten the form of the plot by re-expressing one or both variables

**Read WCGW (What Can Go Wrong): p. 160-161

Lurking variable

- a hidden variable that stands behind a relationship and determines it by simultaneously affecting the other two variables

HW Chapter 7: See HW Log