Everything you need to know about Chapter 7: Maggie Linebaugh

Scatterplots, Association, and Correlation

Scatterplots – a time plot, they are the most common and effective displays of data. They are the best way to observe the relationship between to *quantitative* variables. This means that both the x–axis and y-axis are labeled with numerical values.

ASSOCIATION OF SCATTERPLOTS: When describing a scatterplot's association, one should first start with **direction** of the scatterplot; this may be either positive or negative. A pattern that runs from upper left, to lower right is said to be negative. A pattern running from lower left, to upper right is called positive. **Form** is also necessary when describing scatterplots, if the relationship is *linear* then it will appear as a cloud or swarm of points stretched out in a generally consistent, straight form. If the relationship isn't straight, and curves gently we can find ways to make it more nearly straight. **Strength** of the scatterplot is also essential when describing a scatterplot. A scatterplot that shows a strong association shows little scatter around the underlying relationship. Finally, when looking at scatterplots one must also define **unusual features.** Often these features may be **outliers.** Outliers are points that are significantly straying away from the overall patter of the scatterplot.

ROLES FOR VARIABLES: There are two types of variables, **explanatory** or **predictor variables**, and **response variables**. The explanatory variable should be placed on the *x-axis* and the response variable should be placed on the *y-axis*. Generally the response variable will be what depicts the relationship between the two variables.

CORRELATION: **correlation is not the same as association, and does not imply causation.* Correlation is a numerical measure of the direction and strength of a linear association. Correlation coefficients may be between -1 and +1, no higher or less. A correlation of -.98 would show a *negative, strong* association. A correlation of+.34 would show a weak, positive association. Before finding the correlation of a scatterplot, one must first define the necessary conditions. First, the **quantitative variables condition**: are both variables defined by numerical values, not by categorical values. Second is the **straight enough condition**: Is the plot linear, if not, calculating the correlation should be reconsidered. Last is the **outlier condition:** are there any significant outliers? If there are, you should find the correlation with and without that outlier.

To find correlation: Correlation may be found by hand, but the easiest way is through the calculator. Go to **catalog**, scroll down to the ds and look for **DIAGNOSTICS ON**. After entering data into L1 and L2 and turning on the diagnostics, go to **STAT> CALC,** choose option **8: LinReg (a+bx)** and hit ENTER. What should show up is a list of y=, a=, b=, r^2= and r=. R represents the correlation coefficient. R^2 is the variability.

A Glance at Straightening Scatterplots: When scatterplots are not linear enough, one may re-express the data. This includes squaring the y values, or the x values, or both. Re-expression however is more widely discussed in chapter 10.