Notes About Correlation Coefficient

Scatterplot: A graphical display of two quantitative variables.

Describe the following when interpreting scatterplots:

1. **Direction**: Positive or Negative
2. **Form**: Linear, Curved Gently, Curved Sharply
3. **Strength**: Refers to how closely the points follow the pattern. Very strong, strong, weak, or moderate.
4. **Identify outliers**

After the class activity, copy the following scatterplots from the board:

Linear, Positive, Strong

Linear, Positive, Moderate

Linear, Positive, Weak

Linear, Negative, Strong

Linear, Negative, Moderate

Linear, Negative, Weak

Curved Sharply, Positive, Strong

Curved Gently, Negative, Weak

Curved Sharply, Moderate

Correlation coefficient: A number between -1 and 1 that measures the strength and direction of a linear relationship between two quantitative variables.

Before you use the correlation coefficient, you must check the following conditions:

Quantitative variables condition: correlation is ONLY APPROPRIATE for two quantitative variables

Straight Enough Condition: You can calculate the correlation for any pair of variables, but it measures the strength of linear relationships only! This will be misleading if your relationship is not liner, so make sure your association appears to be linear!

Outlier condition: correlation is SENSITIVE to outliers, so an outlier will heavily affect the correlation. Report the correlation with and without the outliers!

Properties of the Correlation Coefficient (r)

1. The magnitude of the correlation coefficient describes the strength of the relationship:
   $|r| < 0.3$: weak linear relationship
   $0.3 < |r| < 0.7$: moderate linear relationship
   $0.7 < |r| < 0.9$: strong linear relationship
   $0.9 < |r| < 1.0$: very strong linear relationship
2. The magnitude of r is not the slope!!
3. The sign of the correlation coefficient describes the direction of the scatterplot (positive or negative).
4. If there is no clear pattern in the scatterplot at all, the correlation coefficient is 0.
5. The correlation coefficient does not have units.
6. AND... correlation does not imply causation! Just because two variables are related doesn't mean one causes the other.

After the class activity, copy the following scatterplots from the board:

r=0.25 r=0.5 r=0.75

r=-0.25 r=-0.5 r=-0.75

r=0 r=1 r=-1

The sample correlation coefficient, r, is an estimate of the true correlation coefficient, ρ. ρ is a measure of the true linear relationship between two quantitative variables in the population. Next time, we will talk about making inference about ρ.

Below is output from the `pairs()` command in R. Describe each scatterplot (form, strength, and direction)

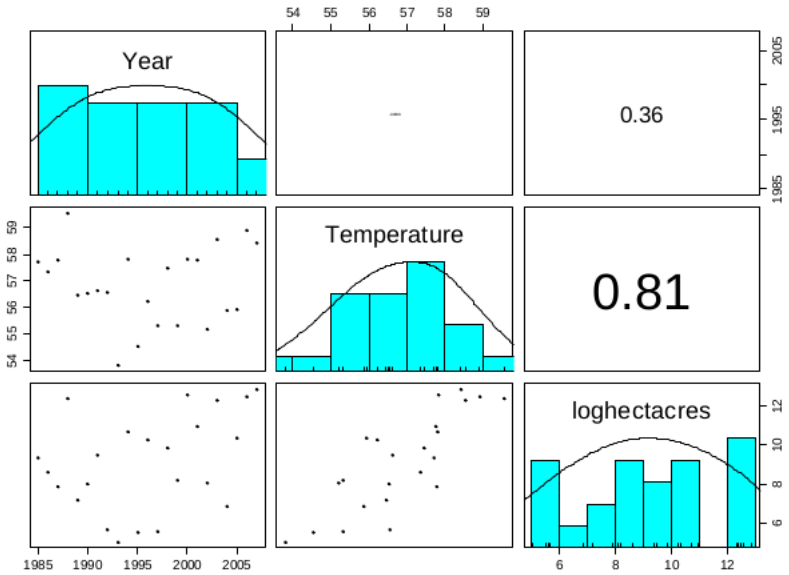

Figure 5.4

The sample correlation coefficient for males in the scatterplot below was found to be 0.1269 and the sample correlation coefficient for females was found to be -0.1679. Explain why the overall scatterplot is misleading.