HOW REGRESSION WORKS

Regression is a method designed to create a **FORMULA** that uses some information to **PREDICT/EXPLAIN** an outcome, using **DATA**.

This is how to perform regression:
- Have a "what's the formula?" question.
- Collect data.
- Look at the pattern – usually with a scatterplot – to choose a formula.
- Get a computer to calculate the numbers and p-values.
- Check the p-values.
- Choose your final formula.

This is what regression means:
- It tells you a formula for how an outcome varies based on other information.
- It does NOT tell you if some things CAUSE others, only how to calculate them as accurately as possible.
- The computer output will tell you p-values and confidence intervals to answer other types of questions.

More details:
- *DESCRIBING A RELATIONSHIP:*
  - Scatterplot describes relationship – and helps choose a good formula
  - Correlation coefficient (r) measures how strong a linear relationship is.
    Ranges from -1 (perfect negative) to 0 (no relationship) to 1 (perfect positive).
    Ignores how steep the slope is, only says how close to a line.
- *FINDING AND INTERPRETING THE FORMULA:*
  - Computer program will use the data to find the numbers that make the formula fit best.
  - The coefficient says how much the outcome changes (on average) for a change of 1 in the explanatory variable.
- *LOOKING AT P-VALUES:*
  - The p-value that goes with the F-statistic in the ANOVA table tells you whether all the variables at once have a relationship with the outcome.
    Low p-value means the relationship is "significant".
  - The p-value for each coefficient tells you whether that explanatory variable appears to have a relationship with the outcome.
    Low p-value means the effect is "significant".
- *LOOKING AT CONFIDENCE INTERVALS:*
  - The confidence interval that goes with an explanatory variable tells you how large or small the real effect could be.

NOTE: Regression has assumptions that must be checked in order to use it properly, *especially* if you plan to use the p-values and confidence intervals.