**Overview of Some Advanced Statistical Methods**

- Traditionally, our science has relied on t-tests, ANOVA, and regression to study our questions of interest.
    - Social and cognitive science: focus on experiments, t-tests, and ANOVA
        - Example: drug treatment for depression.
            - Research question: is the drug an effective treatment for depression?
            - Hypothesis: yes, the drug effectively decreases symptoms of depression
            - Two groups: placebo control and drug treatment
            - We do a t-test to see if there is a significant difference between the groups. If there is, then it supports our hypothesis (careful- does not prove the hypothesis!)
        - ANOVA lets us have multiple treatments with multiple levels
            - Can examine direct effects and interactions
    - With non-experiments (most common in personality & developmental research), we rely on correlations and linear regression
        - Includes 1 or more continuous variables
            - ANOVA is a special case of regression
        - Example: predicting health from self-reports of happiness
            - Research question: do higher levels of happiness predict better health outcomes?
            - Regression equation: $Y = b_0 + b_1X_1$
        - A nice feature: we can add multiple variables, and look at the effect of one variable on another while controlling for the effect of other variables
        - Beware: people like to talk about this in causal language – but no basis for this
            - Usually cross-sectional studies (so does happiness → health, health → happiness, or something else?)
            - True experiments are the only way to determine causality

- Increasingly, we are seeing a host of other statistics appear in the literature.
    - Why?
        - Increased computing capability
        - Directly deals with violations of assumptions
        - Fancy looks good
    - Are these complex statistics the way to go?
        - <u>Some positives:</u>
            - Can look at relationships over time
            - Better addresses our research questions
            - Probably better reflects the complexity of real life phenomena
        - <u>Some negatives</u>
            - Simple methods are more powerful
            - Aim for parsimony (even in the fancy models)
            - Gets complicated very quickly
            - A major concern: the black box

- Some stats you might encounter…
    - **Hierarchical linear regression**
        - This is your basic type of regression. We are predicting an outcome (Y) from one or more predictors ($X_1$, $X_2$, … $X_k$). Essentially, assuming there is a linear relationship between two variables, we are finding the line of best fit, and using this to predict an outcome. We can include control variables, and consider interactions between variables. In each case, we are looking at how much of the variance in Y is explained by X, holding the effect of the other variables in the equation constant.
    - **Path modeling**
        - An extension of regression, this essentially puts regression into visual form. An advantage is that we include error in the model.

- **Structural equation modeling (SEM)**
  - Here, we are directly modeling theoretical relationships between variables. We can include both observed and unmeasured (latent) variables. Generally, this is more confirmatory – we are supporting one theoretical model over another (although in practice a lot of exploration occurs). Factor analysis, path modeling, and regression are all special cases of SEM. You will see this more and more over the years ahead.

- **Factor analysis**
  - The goal of factor analysis is to reduce a large set of variables into a few scales or dimensions. We assume there is a linear relationship between an underlying factor and the observed variables that make up the factor. There are both confirmatory and exploratory types. This is used a lot in personality research (and is the basis of how the Big 5 were formed)

- **Cluster analysis**
  - This is similar to factor analysis, except instead of grouping variables, we are grouping people, and then describing what makes one group of people similar to each other and different from another group of people. For example, I might create clusters of physical activity (very active, sedentary, in the middle) and look at what characteristics distinguish the groups.

- **Logistic regression**
  - In regular regression, we have a continuous outcome. Logistic regression is a type of regression that has a dichotomous outcome (e.g., live/die). The question here is, what are the odds of one outcome or the other, based on a specific predictor. Like regression, we can include multiple predictor and control variables and look at interactions between predictors.

- **Survival analysis**
  - Survival analysis lets us look at two things whether a particular event occur and if so, when. This is most relevant to survival – when do people die, but also applies to things like attrition. Like regression, we include variables to predict how long people will survive (or the hazard of dying at a particular age). An advantage of this is that we can include censoring, which is helpful when we don't observe the event for everyone in the sample.

- **Hierarchical Linear Modeling (HLM)**
  - This is another data analytic technique that is increasingly becoming more popular. HLM is used when nesting occurs. There are two ways this happens – when individuals are nested within groups (e.g., children in classrooms) or when observations are nested within a person (i.e., the same individual is measured over time). This allows us to look at dependent outcome, and consider how much variance is explained by individual factors (e.g., time, teacher) versus how much is explained by external factors (e.g., sex, early experiences, a treatment).

- **Meta-analysis**
  - This actually includes a host of techniques that are used to quantitatively combine the results of multiple studies, to find the average effect across studies, and look at moderators of this effect. This helps us move beyond a single study with a single sample and see what the science as a whole says about a particular phenomenon.

There of course are tons of other types of analyses that can be done, but these are probably the ones you will see most often. Each one has advantages and disadvantages – each is simply a tool within our statistical toolbox.