# Introduction

PSY 5102: Advanced Statistics for Psychological and Behavioral Research 2

## Basic Advice for Graduate Students

- What you will get out of graduate school is directly related to what you invest in the experience
  - Be actively engaged in your own education and training
- Decide what is important to you and make plans for reaching your goals
- Be honest about your strengths and weaknesses
- Be professional
- Use your time effectively while also taking care of yourself
- Be passionate and committed to the science of psychology
- Build a strong relationship with your major professor
- Most learning should be happening outside of the classroom
- Learn how to accept criticism
- Learn how to deal with failure
- Recognize that the world is a competitive place
  - …but remember that someone else's success is NOT your failure
- Publish your work and present at conferences
- Focus on producing high quality work

## Current Controversies in Statistics

- The focus on null hypothesis significance testing
  - This is NOT a new issue
  - It may not really tell us what we want to know
    - "In scientific inference, what we want to know is the probability that the null hypothesis ($H_0$) is true given that we have obtained a set of data (D); that is, $p(H_0|D)$. What null hypothesis significance testing tells us is the probability of obtaining these data or more extreme data if the null hypothesis is true, $p(D|H_0)$" (Kirk, 1996, p. 747)
  - Overreliance on *p*-values
    - Statistical significance (e.g., $p < .05$) is an arbitrary cutoff
    - There has been a suggestion to redefine/restrict statistical significance to "$p < .005$" instead of "$p < .05$"
  - Confusion about the meaning of *p*-values (i.e., probability that the outcome could happen if the null hypothesis is true)
    - This is NOT the same thing as effect size
    - Statistical significance is NOT the same thing as theoretical or practical significance
  - Focus more attention on effect sizes and confidence intervals
  - Bayesian vs. frequentist approaches to hypothesis testing
    - Bayesian approaches focus on probability distributions concerning hypotheses (how much evidence already exists for a hypothesis?) rather than sampling distributions of data

## Current Controversies in Statistics

- Complex decisions regarding how to analyze data (e.g., the garden of forking paths)
  - 29 research teams were given the same data and asked to decide whether soccer referees were more likely to give red cards to dark-skin-toned players than to light-skin-toned players. The teams found very different answers to this basic question (Silberzahn et al., 2018)
- The concern is that researchers may actually make a lot of these decisions after looking at the data but believe/pretend they are just following an existing plan
  - Large number of *potential* associations/comparisons (e.g., fishing expedition, p-hacking, researcher degrees of freedom)
    - The decision to look at main effects with or without interactions
    - The decision to control for other variables in analyses
  - Which data points to exclude?
    - Common exclusion criteria: attention checks, univariate outliers, multivariate outliers via Mahalanobis distance, long-string analysis (invariant response patterns), and inter-item standard deviation (inconsistent response patterns)
  - How to combine data or code responses?
    - Example: How to compare "single" women with women who are "in a relationship"? These decisions may seem simple but there are actually a lot of options available which could lead to important differences (e.g., how to classify a woman who has been dating someone for two weeks?)
  - When to combine data from separate studies?
    - Example: You collect data from two samples. In Study 1 (college students), you find statistically significant results that are consistent with your hypothesis. In Study 2 (community sample), you find results that are similar but not quite significant ($p$ = .09). Combine them? Why or why not?

## Current Controversies in Statistics

- Reproducibility problems in psychology
  - THE IMPORTANCE OF REPLICATION!
    - Some large-scale replication projects within psychology have only been able to replicate about half of the studies they have considered
  - File-drawer problem
  - A lot of studies in psychology are underpowered
  - Researchers often "confuse" *prediction* (acquisition of data to test ideas about what will occur) and *postdiction* (use of data to generate hypotheses about why something occurred)
    - HARKing (hypothesizing after the results are known)
  - Focus on open science practices (e.g., pre-registration, making data sets publically available)

## Ideas for Improving Research in Psychology

- Honesty and transparency
  - Researchers have a lot of flexibility in their decisions as long as they are honest and transparent about what they have done
  - Focus on your long-term reputation within the field rather than any short-term advantages of being dishonest (e.g., a splashy result)
- Focus on finding the truth rather than supporting your hypothesis
  - Researchers should be passionate about answering research questions but they should be cautious about being "too passionate" about supporting a particular idea
- Replication!
  - Popper (1959/2002) defined a scientifically true effect as one that "can be regularly reproduced by anyone who carries out the appropriate experiment in the way prescribed" (pp. 23–24)
  - Both direct replications and conceptual replications are important
- Increasing sample sizes
  - Studies in psychology are often underpowered which may add a lot of "noise" to the existing literature because this inflates false positives
- Avoiding piecemeal publication whenever possible
  - May create the impression that the evidence for an effect is more robust than it actually is
  - Important to note when multiple publications use data from the same study
- Open science practices are helpful
  - Pre-registration of studies, sharing materials, sharing data and analyses