STATISTICS IN THE RESEARCH PROCESS
Reference Sheet

Experimental Design

Analysis Plan

Develop an analysis plan to conduct your analyses in a targeted manner. It should do the following:

- *State a concrete **research question***
- *Identify the **population of interest***
- *Identify the **outcome/dependent variable***
- *Identify the **primary predictor/independent variable***
- *Identify **covariates** (mediators, effect modifiers, etc.)*
- *Describe the way in which **missing data** will be handled*
- *Describe which **model/variable selection** methods will be used*
- *List which **analyses** will be performed in what order.*

Power, Sample Size and Effect Size

Can be calculated for any hypothesis test--just find the power or sample size formula that matches your test

Quantities Included in Calculation

Power	1-β	Probability of detecting an effect (rejecting H_0) if there truly is an effect (if H_0 is false)
Sample Size	n	Number of subjects observed
Effect size	$\mu_1 - \mu_2$, etc.	Detectable difference in means, etc.
Variability	σ, σ^2, etc.	Standard deviation, variance, etc.
Significance Level	α	Probability of a Type I error; usually = 0.05

Need to know all but one of these quantities, then solve for the one you don't know:
e.g., use power, effect size, s.d., and significance level to solve for sample size
e.g., use sample size, effect size, s.d., and significance level to solve for power

Data Collection

Data Description

Descriptive/Sample Statistics

Describe the data using summary measures.

Sample Mean:	$\bar{x}$
Sample Standard Deviation:	s
Sample Variance:	s^2
Sample Proportion:	$\hat{p}$
Sample Odds:	$\frac{\hat{p}}{1-\hat{p}}$

Others…

Graphs are also valuable tools for describing data.

Statistical Inference

Estimation

Point Estimates: *use descriptive stats to estimate population parameters.*

The sample statistic…	…is the best estimator of…	…the population parameter.
$\bar{x}$	→	μ
s	→	σ
$\hat{p}$	→	p

These relationships are not always obvious. For example, the intuitive formula for *s* would have a denominator of n, but the actual formula, with a denominator of n-1, gives better estimates.

Interval Estimates: *Calculate a confidence interval using probability and standard error, a measure of the variability of our point estimate. A confidence interval:*

1) is a set of likely values for the population parameter
2) reflects the uncertainty associated with the point estimate

Hypothesis Testing

Use probability and estimates to accept or reject hypotheses about the population parameters. If the data collected would be highly improbable under the null hypothesis (H_0), it is rejected in favor of an alternative hypothesis (H_1). Hypothesis testing can be done with confidence intervals, p-values, or critical values.

Truth

Decision/Inference	H_0 is True	H_0 is False
Don't Reject H_0	**Correct Decision** Probability = **1– α**	**Type II Error** (false negative) Probability = **β**
Reject H_0	**Type I Error** (false positive) Probability = **α**	**Correct Decision** Probability = **1-β** = **power**

Conclusions/Dissemination

Summarize the results of your estimation and hypothesis testing

Include the following:
Point estimate
Confidence interval
p-value
*Decision to accept/reject null hypothesis

*Note: in practice, the decision may not be stated explicitly. However, it will be apparent in the discussion of the results.

STATISTICS IN THE RESEARCH PROCESS
Worksheet

Experimental Design

Analysis Plan

Develop an analysis plan to conduct your analyses in a targeted manner.

Research Question:
Population of Interest:
Handling of Missing Data:
Model Selection Strategy:
Analyses Planned:

	Outcome/ Dependent Variable	Predictor(s)/ Independent Variable(s)
Name		
Data Scale (continuous, binary, etc.)		

Power, Sample Size and Effect Size

Can be calculated for any hypothesis test--use the power or sample size formula for your test

Step 1: Decide which hypothesis test is appropriate (see Analysis Plan and Hypothesis Testing sections).
Step 2: Fill in any known values below.
Step 2: Once all but one of these quantities are known, use the appropriate formula to solve for the remaining unknown quantity.

	Known/Given Values	Value Solved For
Power (1-β)		
Sample Size (n)		
Effect size		
Variability (SE, SD, etc.)		
Significance Level (α)		

Data Collection

Data Description

Descriptive/Sample Statistics

Are any descriptive stats given in the question?

Sample/Descriptive Statistic	Value(s)
Sample Mean ($\bar{x}$)	
Sample Standard Deviation (s)	
Sample Variance (s^2)	
Sample Proportion ($\hat{p}$)	
Sample Odds $\left(\frac{\hat{p}}{1-\hat{p}}\right)$	
Other	

Graphs are also valuable tools for describing data.

Statistical Inference

Estimation

Point Estimates: *Use descriptive stats to get our best estimate of the population parameter*

Interval Estimates: *Use probability and standard error, a measure of the variability of our point estimate, to calculate a confidence interval.*

Sample Statistic	Value	Confidence Interval	Parameter Estimated
$\bar{x}$			μ
s			σ
$\hat{p}$			p
$\frac{\hat{p}}{1-\hat{p}}$			$\frac{p}{1-p}$

Hypothesis Testing

Use probability and estimates to accept or reject hypotheses about the population parameters

1) Define hypotheses
   H_0:
   H_1:
2) Choose significance level
   α =
3) Calculate test statistic value
   Test Stat =
4) Use test statistic to calculate p-value, or find test statistic critical value and compare critical value to calculated test statistic.
   p = or Crit. Val =
5) Decide whether or not to reject the null hypothesis

Conclusions/Dissemination

Summarize the results of your estimation and hypothesis testing

Include the following:
Point estimate
Confidence interval
p-value
*Decision to accept/reject null hypothesis

*Note: in practice, the decision to accept or reject the null hypothesis may not be stated explicitly. The other three parts will be stated, though, and the decision to accept/reject will be apparent in the discussion of the results.