Inclusion and Exclusion Criteria

Inclusion criteria = attributes of subjects that are essential for their selection to participate.

*Inclusion criteria function remove the influence of specific confounding variables.*

*eg., fitness, menstrual cycle phase, use of oral contraceptives, risks for certain disease states, tobacco use, no prior exercise within 24 hrs, etc.*

Exclusion criteria = responses of subjects that require their removal as subjects.

*eg., failure to adhere to pre-test requirements, infection, evidence of altered training/fitness, etc.*

Experimental Designs: Preliminary Info.

Experimental Designs can be one of three different categories:

- **Between Groups** = different subjects in each group
- **Within Groups** or Repeated Measures = same subjects exposed to different interventions/control
- **Mixed Design** = some factor(s) Between Groups, some factor(s) Repeated Measures

There is also a differentiation based on the number of dependent variables studied and included in the statistical design.

- **Univariate** = one dependent variable
- **Bivariate** = one dependent and one or more independent variables
- **Multivariate** = more than one dependent variable and one or more independent variables
It is also important to distinguish how researchers control knowledge of treatments/interventions between themselves and the subjects

- **Single blind** = when either (not both) of the subjects or the researchers do not know the nature/specificals of the intervention(s).
- **Double blind** = when both the subjects and the researchers do not know the nature/specificals of the intervention(s). This requires that a third party be chosen to determine intervention sequences for each subject.

**Design Problems: Internal Validity**

**Internal Validity** = ability to interpret that measured changes were caused **solely** by the intervention.

To fully appreciate differences between designs, you must be aware of threats to internal validity.

Why is this design bad? O X O

There is no control group, and therefore no way to assess that the intervention was the sole cause of any change in measured variables.

What are the threats to internal validity?
**Threats to Internal Validity**

<table>
<thead>
<tr>
<th>TABLE 8.5 Major Confounding Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maturation</td>
</tr>
<tr>
<td>History</td>
</tr>
<tr>
<td>Testing</td>
</tr>
<tr>
<td>Instrumentation</td>
</tr>
<tr>
<td>Regression to the mean</td>
</tr>
<tr>
<td>Selection</td>
</tr>
<tr>
<td>Attrition (Mortality)</td>
</tr>
<tr>
<td>Diffusion of treatment</td>
</tr>
<tr>
<td>Sequence effects</td>
</tr>
</tbody>
</table>

**Biased and Unbiased Sampling**

**Participant Selection**

- Population
  - General Population
  - Target Population
  - Accessible Population
  - Sample

- A. General Population
- B. Target Population
- C. Accessible Population
- D. Sample
Biased and Unbiased Sampling

Sample = selected subset of a population

As it is typically impractical, if not impossible, to research an entire population, we need to sample from the population.

What is an unbiased sample?
One where every member of the population has an equal chance of being included in the sample.

Do we ever really know all people from a given population?

Work in groups of 2-3, and …

1) Identify 2 to 3 populations that are of interest in your field.
2) For each population, state a) how you could or could not sample from it, b) how you would obtain a sample, and c) how biased your sampling really is.

Types of Sampling

Simple Random Sample = when every member of the population has an equal chance of being included in the sample.

Random sampling is important because:

1. Helps control threats to internal and external validity
2. Can control for many variables simultaneously
3. It is the only control procedure that can control for unknown factors
Types of Sampling

Sample of Convenience = when, through convenience, sampling occurs from only a subset of the intended population.

Volunteerism (ad hoc sampling) = when sampling is based to a large extent on individuals volunteering to participate in the study. (due to ethical reasons mandated by human subjects review committees, this is hard to avoid)

Systematic Sampling = When every n\textsuperscript{th} person is selected.

Stratified Random Sampling = Attempts to decrease sampling errors that exist even if using simple random sampling.

When a population is first divided into strata based on a different variable (eg. Gender), and then random sampling occurs from each strata.

- the same relative representation of each strata should occur
- more than one additional stratification variable can be used (eg. age, gender, ethnicity, wealth, geographical location, political bias, hours of television/day, etc.)

• Problem
you need access to and knowledge of the entire population to do this!!!
Types of Sampling

Free Random Assignment = using random number tables or computer generated random numbers

Matched Random Assignment = for smaller sample/groups sizes, subjects can be matched on certain characteristics, and then matched subjects can be randomly assigned

Balanced Assignment = ensuring that all group sizes, or sequences of trial orders, are equal

Cluster Sampling = when groups (clusters) of individuals are drawn rather than separate individuals (e.g., all students of randomly chosen APS 3rd grades; pregnant women from pre-natal classes)

Purposive Sampling = intentionally selecting specific individuals due to their traits.

Snowball Sampling = when subject recruitment is aided by the first participant.

Multi-Stage Sampling = really a multiple level stratified random sample. (e.g., Stratify all counties in US based on socio-economic issues, randomly select households from this list, and then randomly select household members. Used a lot in survey research)

Note:

• in reality, the sampling used is often a combination of several of these methods
• Extremely important to describe the characteristics of ad hoc samples
• Results should be generalised only to people who are like those used in the study.
Effect Size and Statistical Power

Prior to conducting the study (apriori), researchers should;

• estimate the size of a mean difference that is meaningful
• identify a type I error probability that is acceptable to them and the study/DV’s.
• identify a type II error probability that is acceptable to them and the study/DV’s.
• estimate the sample size needed to detect this mean difference, given the aforementioned type I and type II errors.

“With a large enough sample size we can detect even a very small difference between the value of the population parameter stated in the null hypothesis and the true value, but the difference may be of no practical importance. (Conversely, with too small a sample size, a researcher may have little chance to detect an important difference.)

Remember Type I and II Errors

Type I Error:
Probability of rejecting \( H_0 \) when \( H_0 \) is true (\( \alpha \))

**Stating that there is a difference when there really is not!!!**

Type II Error:
Probability of retaining \( H_0 \) when \( H_0 \) is false (\( \beta \))

**Stating that there is no difference when there really is!!!**

<table>
<thead>
<tr>
<th>Mean Difference</th>
<th>Null Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Reject</td>
</tr>
<tr>
<td>Yes</td>
<td>correct</td>
</tr>
<tr>
<td>No</td>
<td>Type I error</td>
</tr>
</tbody>
</table>

PEP507: Research Methods
Effect Size and Statistical Power

The Power of a test

The probability of correctly rejecting a false $H_0$.

$$\text{Power} = 1 - \beta$$

Probability of type II error

Factors Affecting Power

1. Size of the effect
## Factors Affecting Power

### 2. Sample Size
Increasing the sample size decreases the likely difference between the true population mean and the mean of your sample.

### 3. Variance of DV
As with a small sample size, high variance of the DV can make your sample mean more different from the true population mean.

It is important for the researcher to realise that a considerable source of variance in the DV can be caused by the poor quality of the research design and/or methods used in the study.

Always be aware of the need to decrease variability in any variable that is caused by factors other than sampling (e.g., Instrumentation, Inconsistent research methods such as reward, motivation, explanations, etc.)
4. Level of significance (\(\alpha\))

We tend to use \(p<0.05\) by convention, but no scientist is bound by this level of significance.

5. One vs. two tailed statistical tests

If past research and the logical understanding of the variable and intervention mandates that there is only one direction of the response, then a one-tailed statistical test will be more powerful than otherwise.
Calculating the Power of a Test

Although we do not really know what the mean of our DV will be after our intervention, we can estimate this based on past research, and our interpretation of what will be a meaningful effect.

Effect Size

By convention, we express the mean difference relative to the standard deviation of the variable within the population at question the effect size.

Effect size \( (d) = \frac{\mu_{\text{true}} - \mu_{\text{hypo}}}{\delta} \)

Important:

The effect size and not p value tells us of the magnitude of the effect.

You can have a minimal effect be significant if your sample is large enough!!!!!
Estimating Power and Sample Size

Typically, a researcher determines an acceptable minimal power (eg: 0.8), and then estimates the sample size needed to show an expected effect size to be significant.

Problem:
Computations of power are specific to research designs, and no single paradigm exists for power estimations. However, use of a t-test based power profile provides the researcher with some direction.

Let’s work on a problem of our own!!!

• Chin-ups completed before PE = 5 ± 3.2 (SD)
• Expected chins ups completed after PE = 8
• Effect size = (8-5) / 3.2 = 0.9375
• How many subjects do we need at power = 0.8 to allow this difference, if it occurs, to be significant?

Note, this power curve chart is for t-test $H_0: \mu_1 - \mu_2 = 0$, independent samples, $\alpha = 0.05$
Estimating Power Using Computer Software

Power estimation is made easier by commercial software.

I use the free software called “GPower v3.1”, available at the following URL.

wwwpsycho.uni-duesseldorf.de/abteilungen/aap/gpower3/