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_o when H_o is true (α)

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

Type II Error:

Probability of retaining H_o when H_o is false (β)

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

Mean Difference	***Null Hypothesis*** Reject	Accept
Yes	correct	Type II error
No	Type I error	correct

Effect Size and Statistical Power

The Power of a test

The probability of correctly rejecting a false H_o.

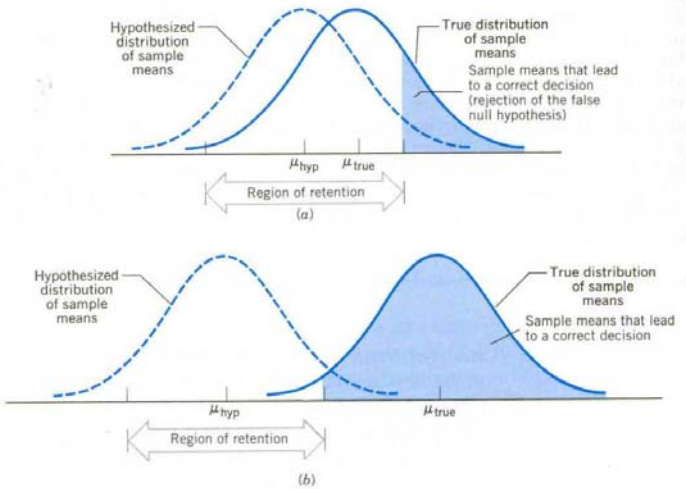
Power = 1 - β

Probability of type II error

PEP507: Research Methods

Factors Affecting Power

1. Size of the effect

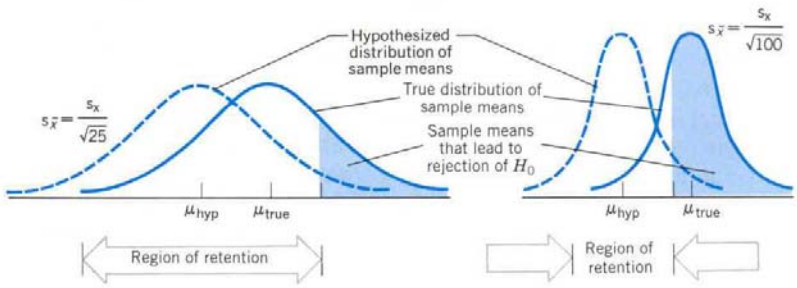


PEP507: Research Methods

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.



Factors Affecting Power

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.

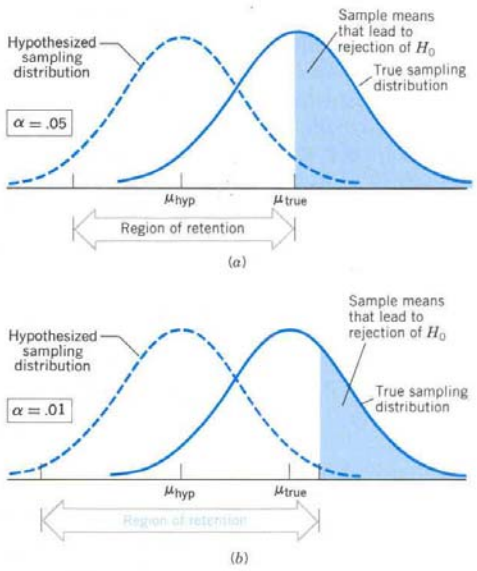
It is important for the researcher to realise that a considerable source of variance in the DV can be caused by the poor quality of the research design and/or methods used in the study.

Always be aware of the need to decrease variability in any variable that is caused by factors other than sampling *(eg. Instrumentation, Inconsistent research methods such as reward, motivation, explanations, etc.)*

Factors Affecting Power

4. Level of significance (α)

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

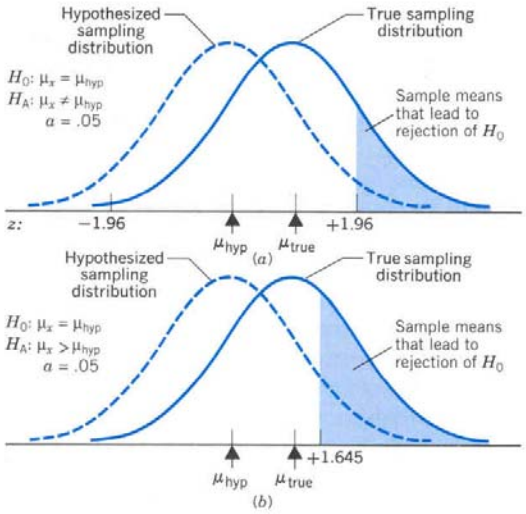


PEP507: Research Methods

Factors Affecting Power

5. One vs. two tailed statistical tests

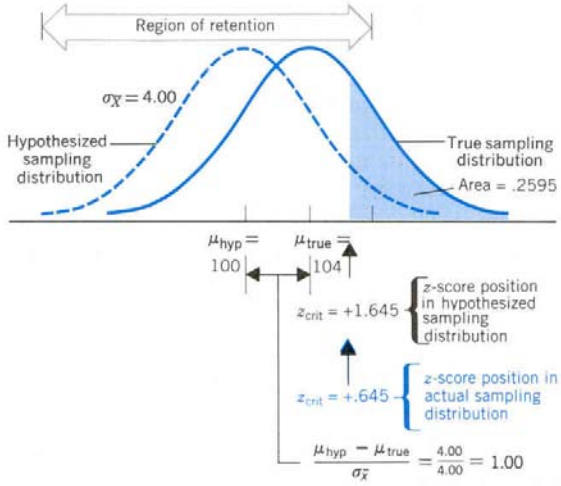
If past research and the logical understanding of the variable and intervention mandates that there is only one direction of the response, then a one-tailed statistical test will be more powerful than otherwise.



PEP507: Research Methods

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) = $(\mu_{true} - \mu_{hypo}) / \delta$

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.

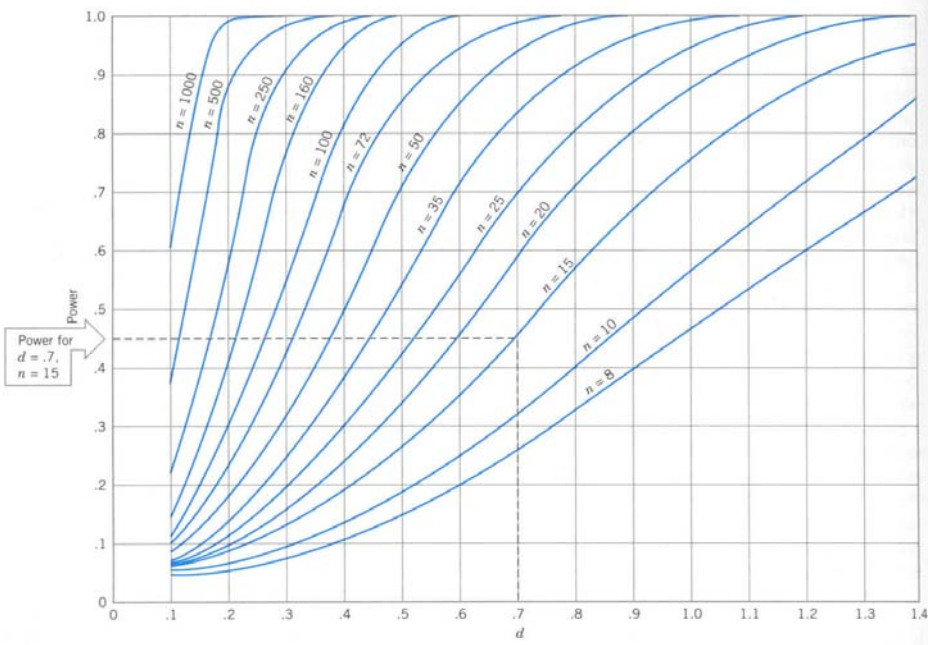
Problem:

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?

PEP507: Research Methods



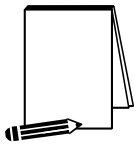
Note, this power curve chart is for t-test H_o:μ_1 - μ_2 = 0, independent samples, α = 0.05

PEP507: Research Methods

Effect Size and Statistical Power

Meaningful effect sizes vary with the topic. However, typically, an effect size > 0.3 is potentially meaningful.

Work by yourself and …



1) *Read your selected research article.*
2) *Identify the mean difference for an important DV.*
3) *Identify the standard deviation for the variable. If two SD's are given (eg. pre and post), then average these.*
4) *Compute the effect size.*
5) *Use the power curve provided and the sample size of the study to estimate statistical power.*
6) *Could there have been a type II error in this study?*

Estimating Power Using Computer Software

Power estimation is made easier by commercial software.

I use the software called "StateMate", produced by GaphPad.

www.graphpad.com