

# Hypothesis Testing, Effect Sizes and Statistical Power

Brad Verhulst

Texas A&M University

# Statistical Power

**Statistical Power** – The probability of **correctly** rejecting the **Null Hypothesis**

Statistical Power depends upon:

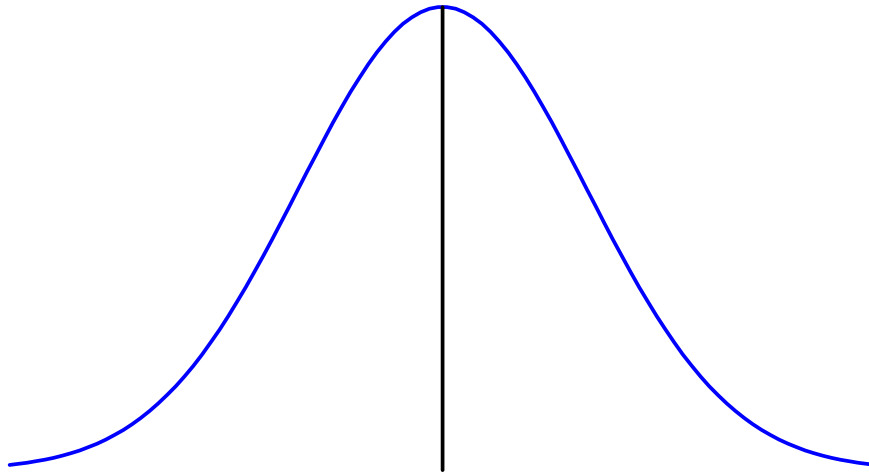
1. Sample size
2. The  $\alpha$  level
3. The  $\beta$  level
4. Effect size

# Hypothesis Testing

1. Define a null hypothesis
  - Often “No Difference”
2. Define your alpha level, or what would be considered evidence that the observed phenomenon was inconsistent with the null hypothesis
  - E.g.  $p < .05$  or  $p < 5 \times 10^{-8}$
3. Obtain data and conduct the analysis

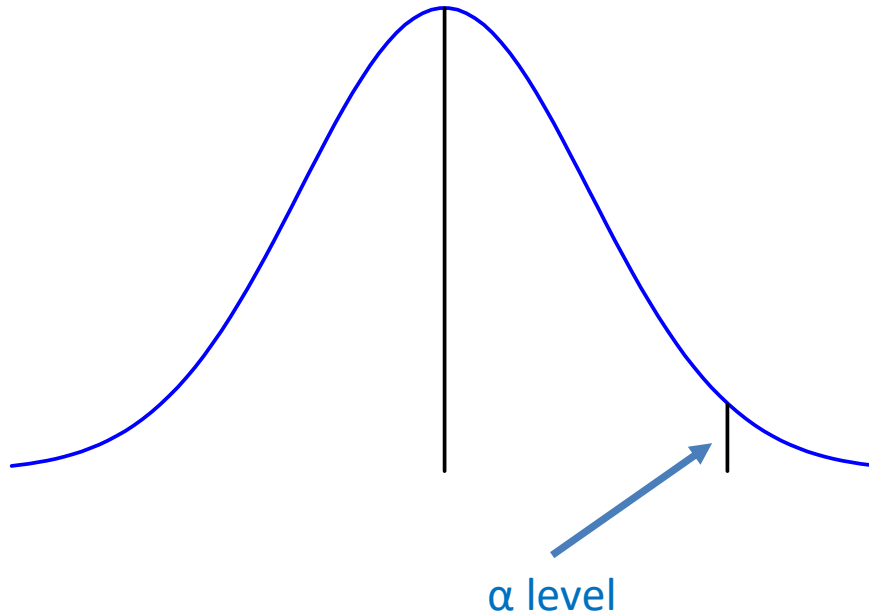
# Hypothesis Testing

Distribution of our test statistic  
under the Null Hypothesis



# Hypothesis Testing

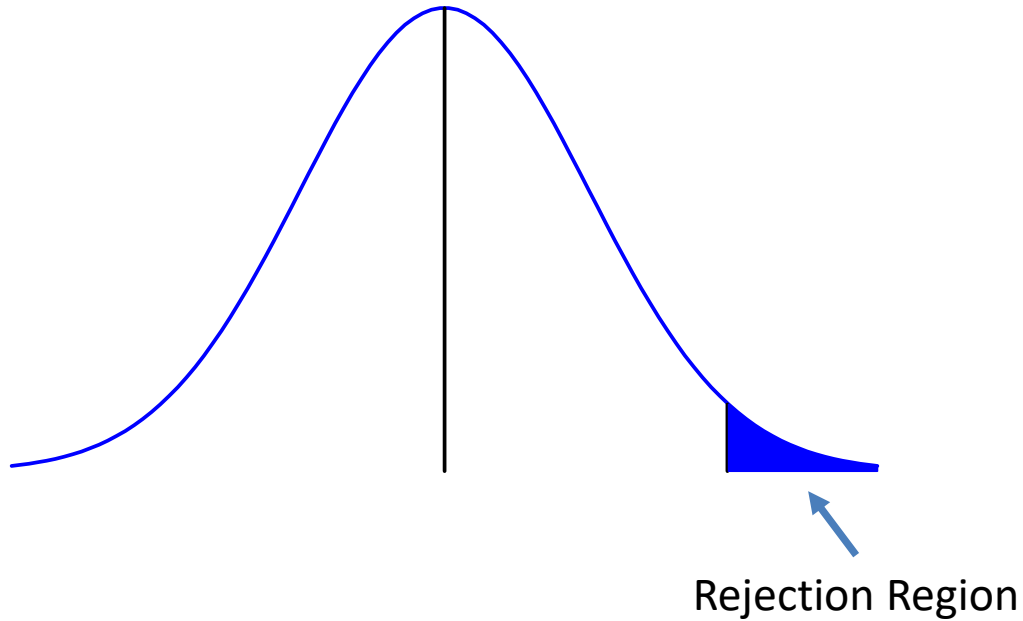
Distribution of our test statistic  
under the Null Hypothesis



Sufficient evidence that the observed statistic is  
not characteristic of the null distribution

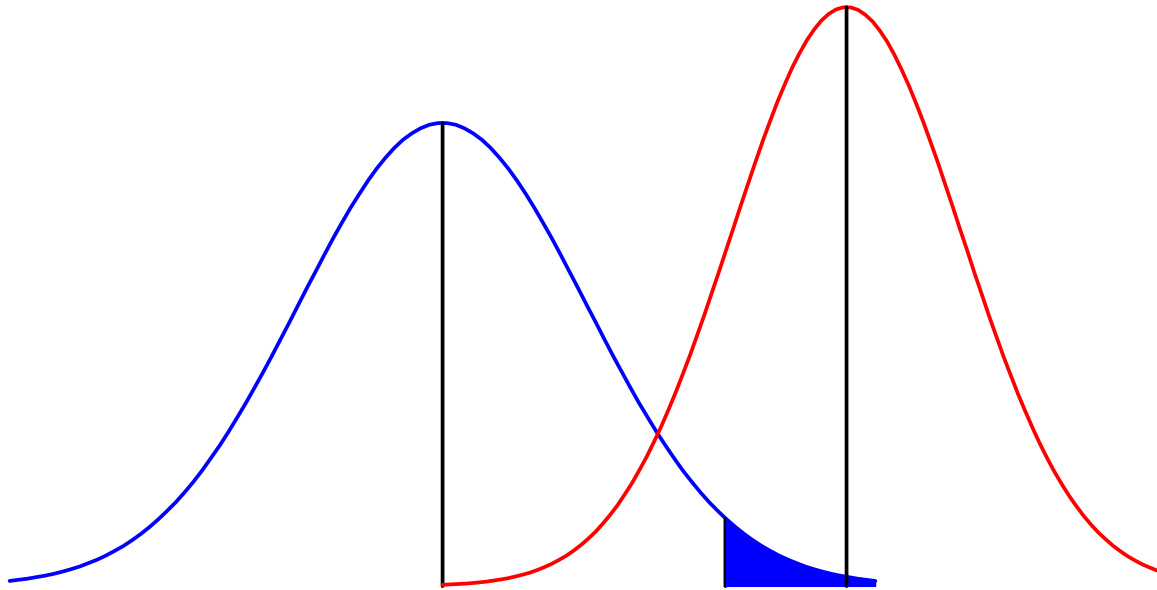
# Hypothesis Testing

Distribution of our test statistic  
under the Null Hypothesis



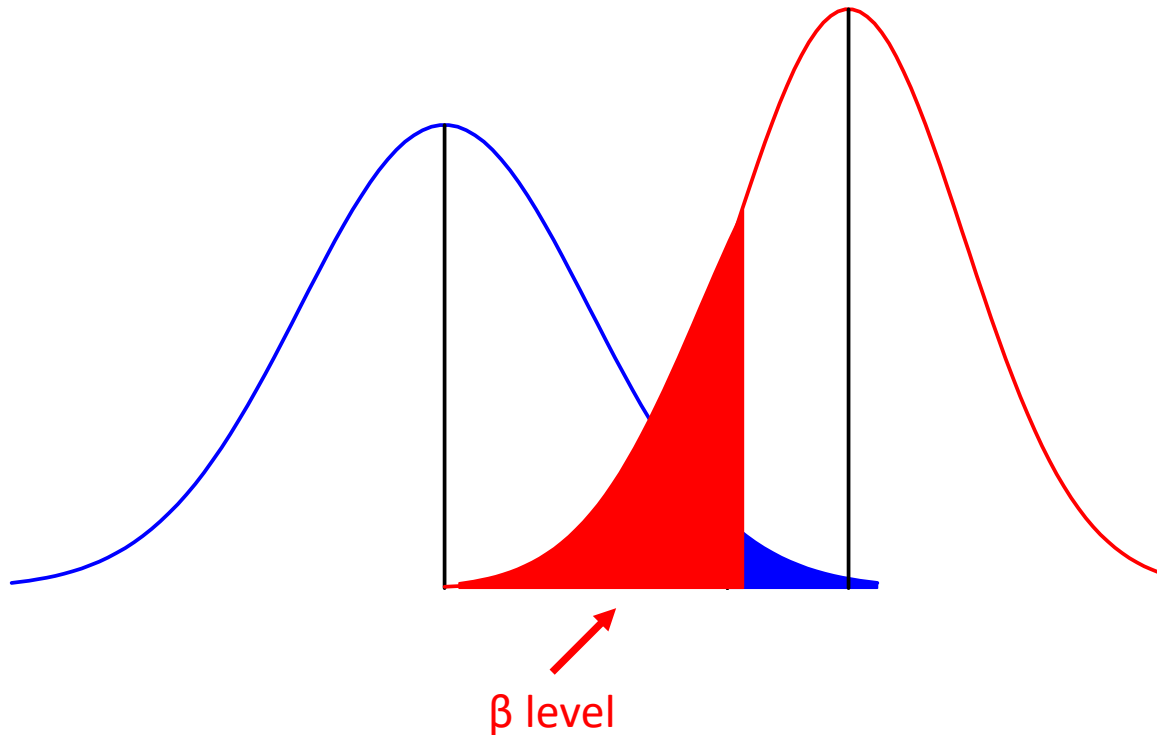
# Hypothesis Testing

Distribution of our test statistic  
under the Alternative  
Hypothesis



# Hypothesis Testing

Distribution of our test statistic  
under the Alternative  
Hypothesis



Sometimes, even if our statistic is drawn from the distribution under the alternative hypothesis it will not be sufficient evidence that it was not a null result



# Effect Size

---

The Earth Is Round ( $p < .05$ )

---

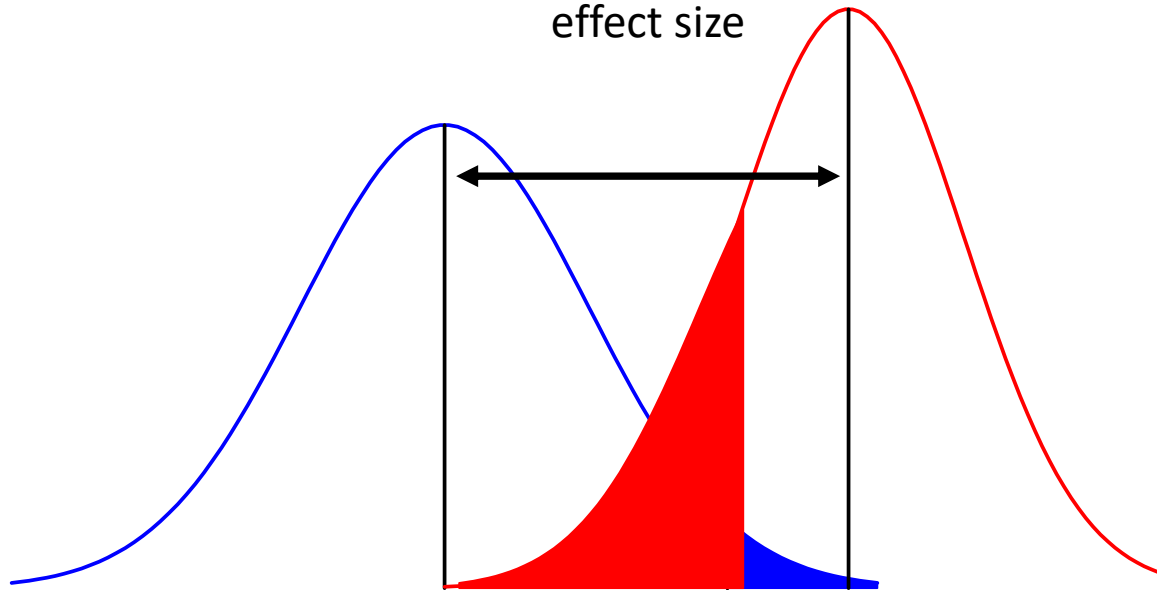
Jacob Cohen

# Effect Size

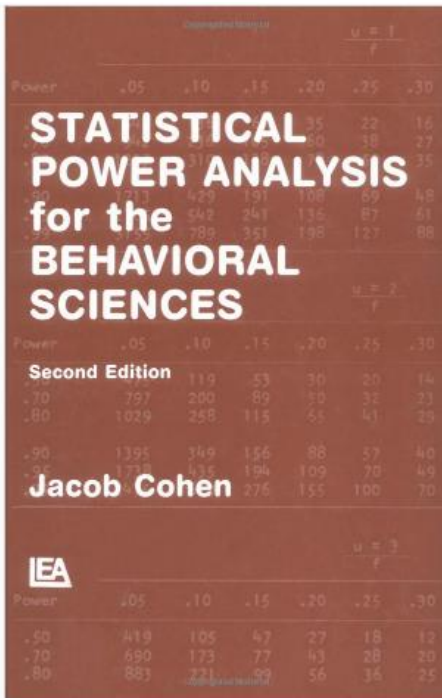
- Effect Size** – A measure of the strength of a phenomenon in the population, independent of sample size.
- Help communicate the result (in everyday language) if the scale is meaningful on a practical level
    - People taking Zyban (Bupropion) smoke 5 less cigs/day
    - People taking Liponox lose 28.16 lbs in 8 weeks
  - Effect sizes are agnostic to whether the effect is real or not
    - They are not associated p-values (as most effect size statistics do not deal with sample size)

# Effect Size

The distance between the mean of the null and the mean of the alternative distributions is the effect size



# Effect Size Conventions



Cohen's (1988) Standards for Interpreting Effect sizes

Be careful with interpreting Effect Sizes

- Most large effects are not necessarily theoretically interesting but tend to be quite obvious

	$d$	$R^2$	Interpretation
Small Effect	.25	.01	Tends to require statistical methods to be observed

# Statistical Power

**Statistical Power** – The probability of **correctly** rejecting the **Null Hypothesis**

Statistical Power depends upon:

1. Sample size
2. The  $\alpha$  level
3. The  $\beta$  level
4. Effect size

# Statistical Power

Two Common Uses of Power Analysis:

- 1. *a priori* (Prospective) Power Analysis:** Useful to consider design sensitivity and examine how many responses are necessary to “fairly” test the Null.
- 2. *post hoc* (Retrospective) Power Analysis:** Explore whether the effects you have can reasonably be expected to reject then null if it is actually false.

# Type I and II Errors

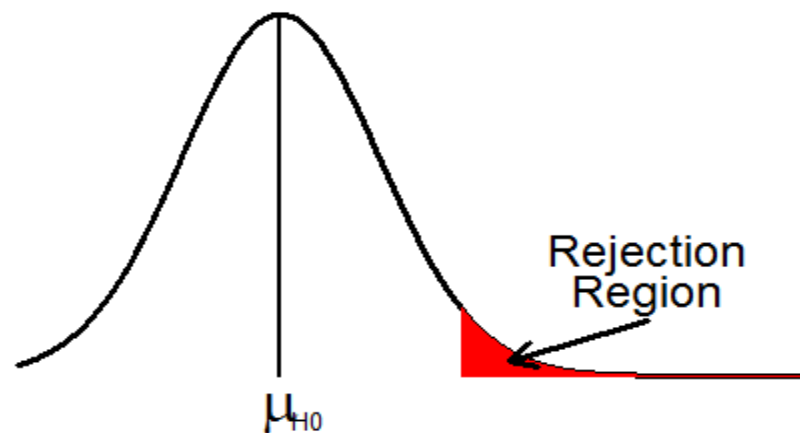
Your Decision	The TRUTH	
	$H_0$ True	$H_0$ False
Reject The Null	<i>Type I Error (<math>\alpha</math>)</i>	<i>Correct</i>
Fail To Reject the Null	<i>Correct</i>	<i>Type II Error (<math>\beta</math>)</i>

# Type I ( $\alpha$ ) Error

The rejection region is the space under the distribution that is greater than  $\alpha$  if:

Therefore,  $\alpha$  is equal to:

$$\alpha = P(Z > z_{\alpha})$$



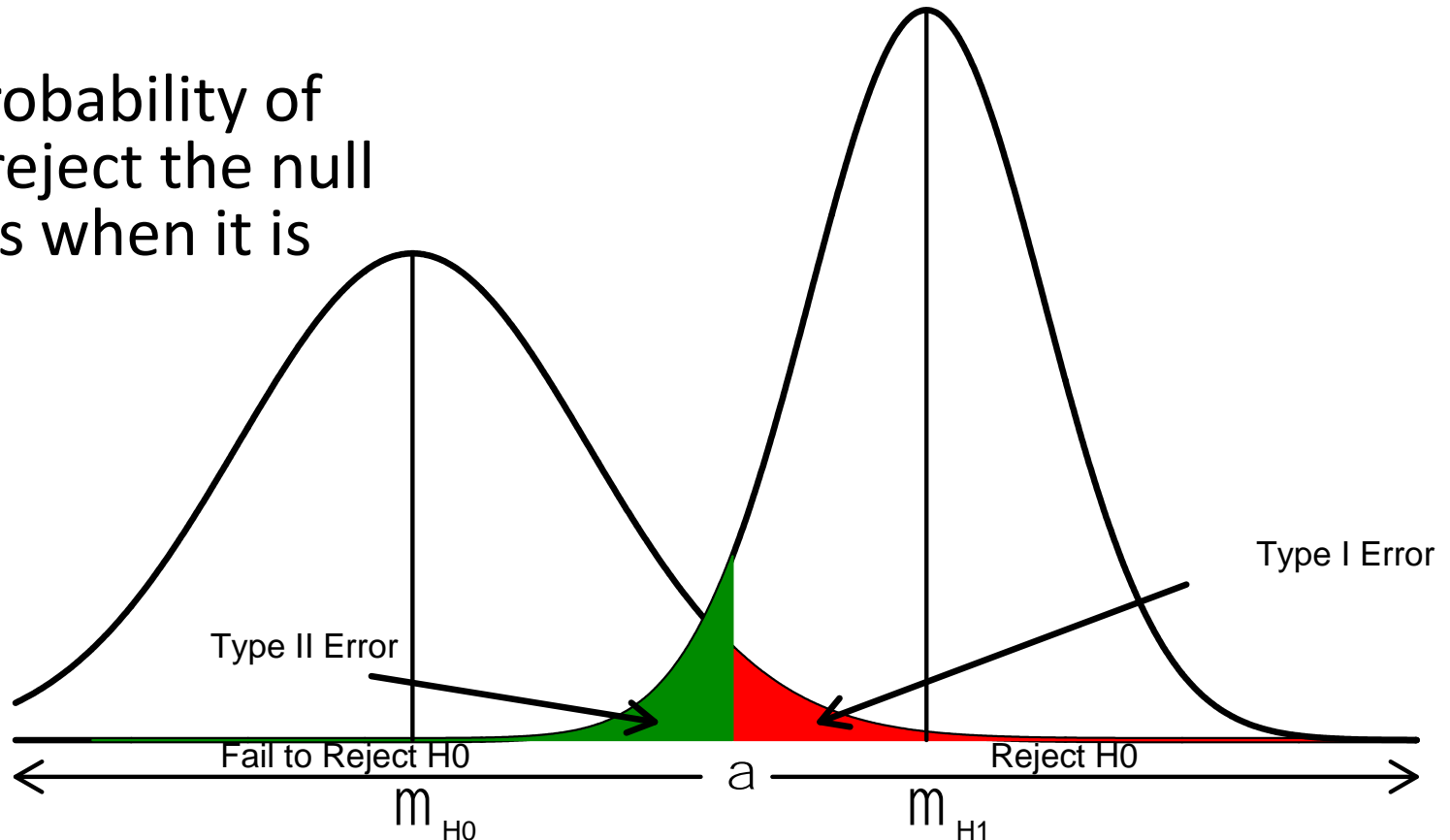
$\alpha$  is fixed exogenously (this is the level of significance that you set for your test)



# Type II ( $\beta$ ) Error

$$b = P\left(\frac{\hat{q} - q_a}{S_{\hat{q}}} \leq \frac{k - q_a}{S_{\hat{q}}}\right) = P\left(Z \leq \frac{k - q_a}{S_{\hat{q}}}\right), \text{ when } q = q_a$$

$\beta$  is the probability of failing to reject the null hypothesis when it is false.

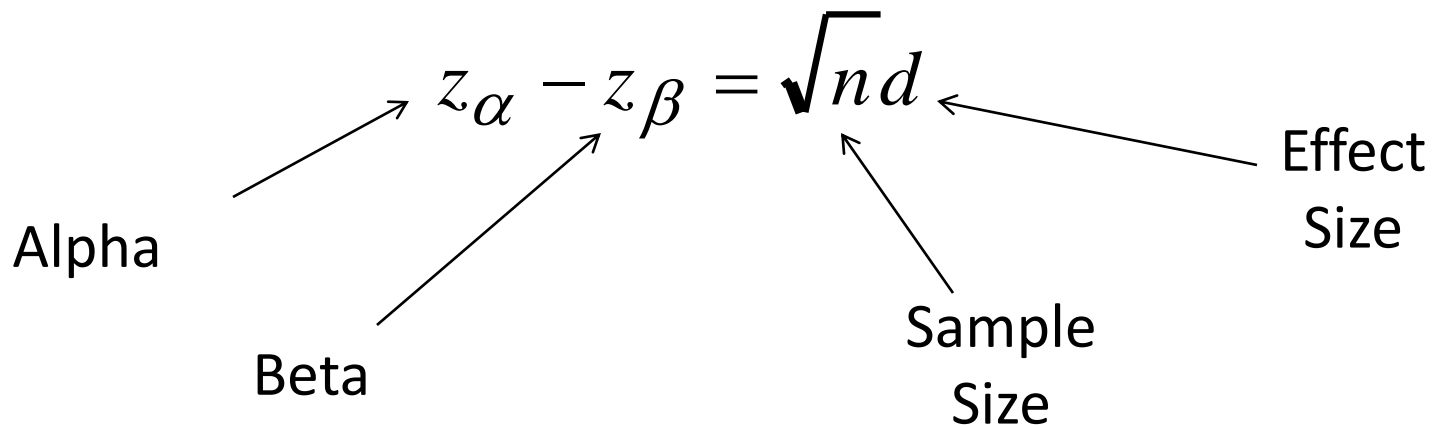


# Standard Conceptualization of Power

$$\mu_0 + z_\alpha \left( \frac{\sigma}{\sqrt{n}} \right) = \mu_1 - z_\beta \left( \frac{\sigma}{\sqrt{n}} \right)$$

- Rearranging the variables we get:

$$z_\alpha - z_\beta = \sqrt{n} \frac{(\mu_1 - \mu_0)}{\sigma} = \sqrt{nd}$$



# Alternative Methods for Calculating Statistical Power

## Monte Carlo Simulation

- Simulate a model numerous times (like 1000) and count how many times you observe a significant test statistic
- The proportion of significant results is your statistical power
- This can be very time consuming

- Non-Centrality Parameters

- Because we are using parametric tests, we are assuming a distribution of the test statistic (e.g. standard normal or Chi-squared distribution)
- We can leverage this assumption to more directly calculate statistical power (without doing a ton of repetitions)
- This can be done relatively quickly

# Non-centrality Parameters

- Non-centrality parameter (ncp)
  - the sum of the mean of the test statistic distribution under the alternative hypothesis and the degrees of freedom
- Two points that are especially important for statistical power:
  1. As the effect size gets larger, the mean of the test statistic gets larger, and therefore the ncp gets larger
  2. As sample size increases, the standard deviation of the sampling distribution of the test statistic gets tighter, and the ncp gets larger

# Calculating Power with NCPs

There are 4 steps to calculate power with NCPs:

1. Simulate (twin) data that corresponds with the alternative hypothesis.
  - This should be based on the literature, as far possible.
2. Fit the full and reduced models to the simulated data to obtain the  $X^2$  value from the likelihood ratio test
3. Calculate the average contribution of each observation to the the  $X^2$ 
  - Calculate the weighted non-centrality parameter (Wncp)
  - Divide the  $X^2$  by the total sample size from the simulation
4. Calculate power
  - Because the ncp increases linearly with sample size, all we need to do is multiply the ncp by a desired sample size to get the  $X^2$  value

# Quick Example

If  $X^2 = 10$  with 1000 observations, the  $W_{ncp} = 10/1000 = .01$ .

- Therefore, on average each observations contributes .01 to the ncp.
- If we had 2000, observations the  $X^2$  would be  $.01 * 2000 = 20$
- If we had 500, observations the  $X^2$  would be  $.01 * 500 = 5$

# A Power Calculator for the Classical Twin Design

Brad Verhulst<sup>1</sup> 

A series of demonstrations of common power analyses for twin models can be found in this file:

[powerScript.R](#)

The actual functions that are used to do a power analysis can be found in this file:

[powerFun.R](#)

Note: you will need to put the powerFun.R file in your current working directory or the powerScript.R will not be able to find the necessary functions