Lecture 10: Power

GENOME 560

Doug Fowler, GS (dfowler@uw.edu)

Review - resampling

- We talked about two types of resampling methods
  - What are they?
  - How do they work?
  - What are they good for?

Goals of power analysis

- To understand what our experiments **can tell us**

- To understand what our experiments **can’t tell us**

- Consequently, to tweak our experimental design

- To give us a framework for comparing different statistical methods

Understanding Power Starts with Error

Decision	Actual Situation "Truth" H_0 True	H_0 False
Do Not Reject H_0		
Reject H_0		

Statistical Power

- This leads us to the idea of power, which is the probability of rejecting the null when the alternative is true (e.g. "true positive rate")

$$power = 1 - P(type\ II\ error)$$
$$power = 1 - \beta$$

Statistical Power

- This leads us to the idea of power, which is the probability of rejecting the null when the alternative is true (e.g. “true positive rate”)

$$power = 1 - P(type\ II\ error)$$
$$power = 1 - \beta$$

- It is common to aim for a power = 0.8 or greater (e.g. an 80% chance you’ll correctly reject the null hypothesis)

Statistical Power

- This leads us to the idea of power, which is the probability of rejecting the null when the alternative is true (e.g. "true positive rate")

$$power = 1 - P(type\ II\ error)$$
$$power = 1 - \beta$$

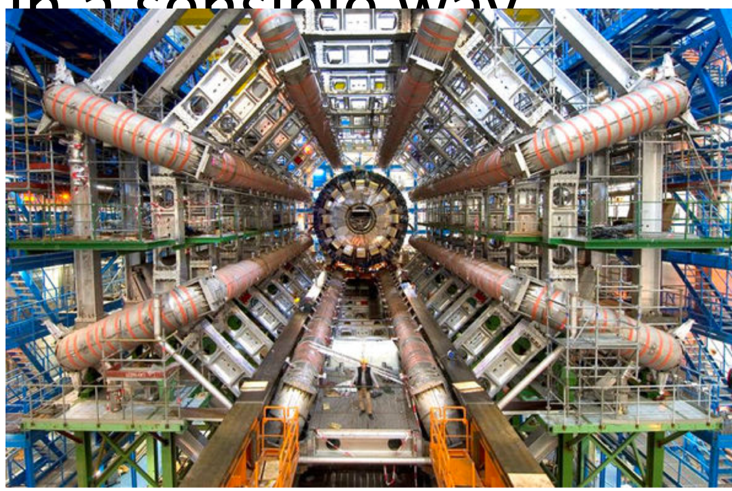
- We'll see that unlike α we can't just choose what the power we will have…

Why Analyze Power?

- Power analysis enables you to construct your experiment in a sensible way

- Especially i

humans/ani



- Nothing is worse than realizing you likely would never have succeeded in detecting the effect you were looking for in the first place…

A Graphical Example

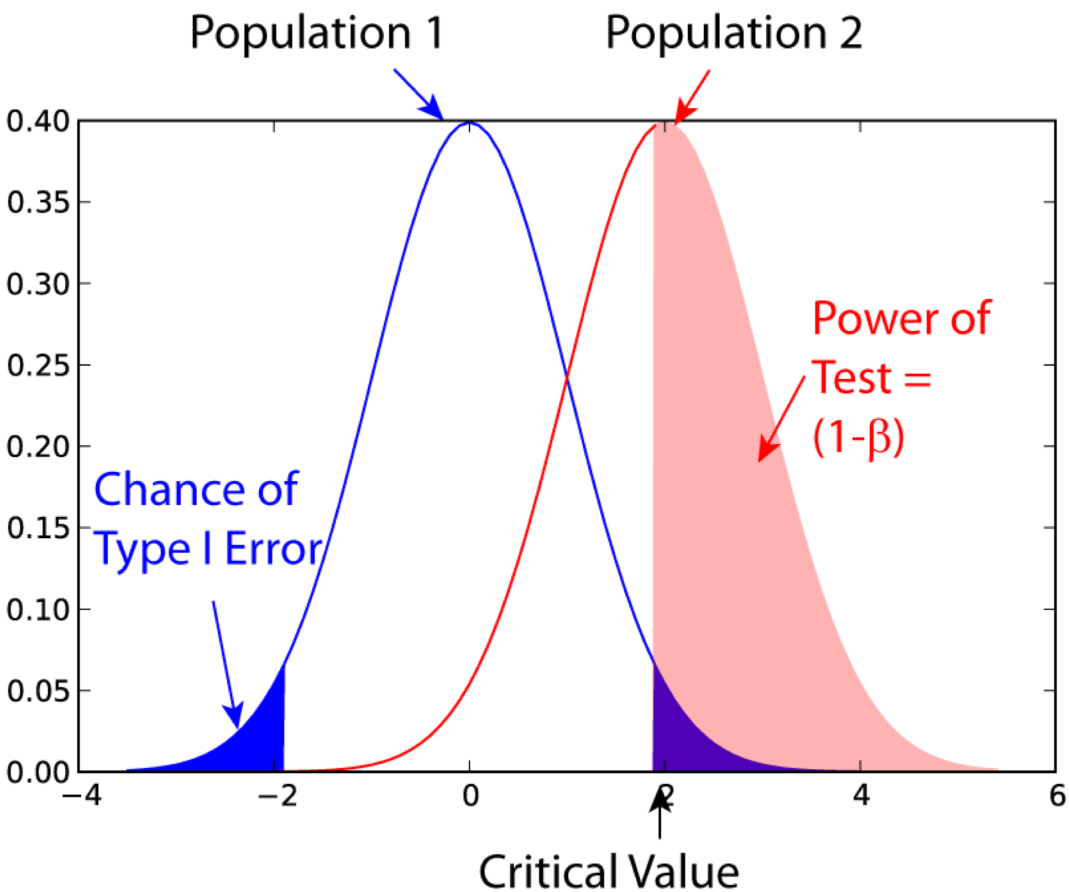
- Let's say we have measured the size of 50 cells that have been treated with a drug and 50 that have not. We would like to test whether cells are larger after drug treatment.

A Graphical Example

- We'll approximate the sampling distribution of the sample means with normal distributions:

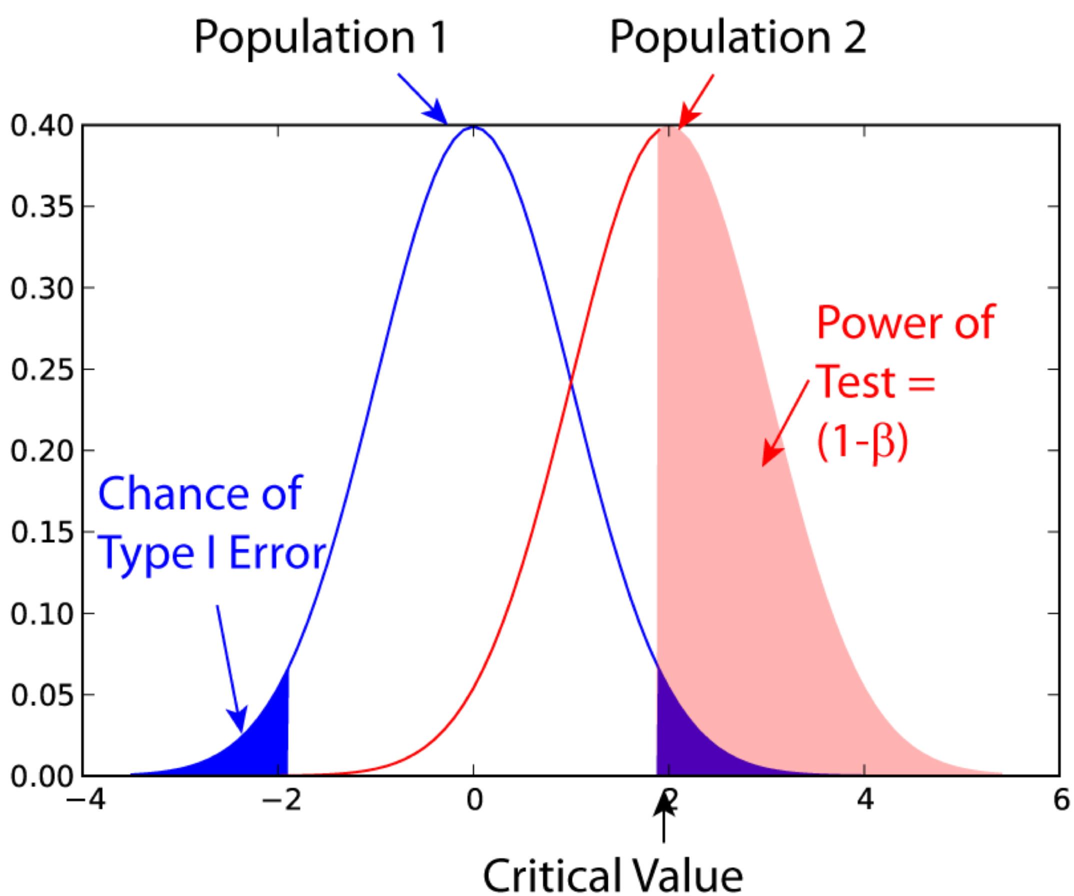
A Graphical Example

- Since they overlap, some samples will enable us to reject the null and others won't.



A Graphical Example

- What factors will impact power (our ability to detect the true difference in means)?



Components of Power Analysis

- To do a power analysis **before we do an experiment**, we need to know about these quantities – where do we get them?

Components of Power Analysis

- How do we get these?
  - σ^2 = variability/noise (as σ^2 decreases, power increases)
  - Δ = effect size/distance between H_0 and H_1 (as Δ increases, power increases)

Components of Power Analysis

- And α?
  - α = type I error rate (as α increases, power increases)

A Concrete Example

- Let's say we wish to test whether a particular transcript is present at 4 copies per cell

A Concrete Example

- Let's say we wish to test whether a particular transcript is present at 4 copies per cell

- To illustrate the concept of calculating power, we'll need a specific alternate hypothesis: let's say we want to know the power to detect a difference if there are are 6 copies per cell

$$H_0 : \mu = 4$$
$$H_1 : \mu = 6$$

A Concrete Example

- Let's say we wish to test whether a particular transcript is present at 4 copies per cell

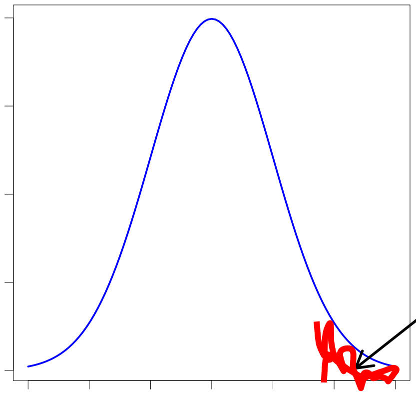
- From past data, we know that the standard deviation is 2. Let's examine power for a sample size of 4 and α = 0.05

$$
\begin{aligned}
&H_0 : \mu = 4 \\
&H_1 : \mu = 6 \\
&n = 4 \\
&\alpha = 0.05
\end{aligned}
$$

A Concrete Example

- Let's say we wish to test whether a particular transcript is present at 4 copies per cell

- The rejection threshold for H_0 is $Z>1.645$ where



Z = 1.645 is 5% of the area

$$H_0 : \mu = 4$$
$$H_1 : \mu = 6$$
$$n = 4$$
$$\alpha = 0.05$$

A Concrete Example

- Let's say we wish to test whether a particular transcript is present at 4 copies per cell

- Then, we find the hypothetical sample mean at the rejection threshold

$$H_0 : \mu = 4$$
$$H_1 : \mu = 6$$
$$n = 4$$
$$\alpha = 0.05$$

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - 4}{\frac{2}{\sqrt{4}}} = \frac{\bar{x} - 4}{1}$$

$$Z = 1.645 = \frac{\bar{x} - 4}{1}$$

$$\bar{x} = 5.645$$

A Concrete Example

- Let's say we wish to test whether a particular transcript is present at 4 copies per cell

- Then, we calculate the Z-statistic assuming H_1

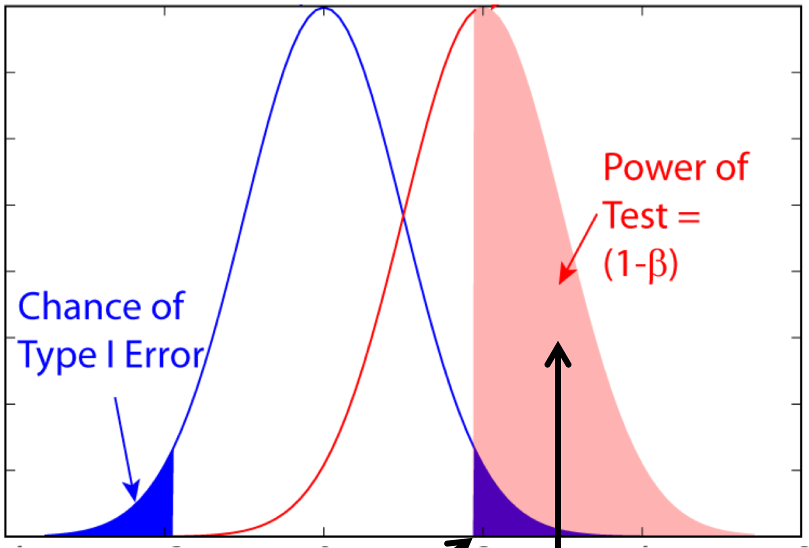
$$H_0 : \mu = 4$$
$$H_1 : \mu = 6$$
$$n = 4$$
$$\alpha = 0.05$$

$$Z = \frac{\bar{x} - \mu_1}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - 6}{1} = -0.355$$

A Concrete Example

- Let's say we wish to test whether a particular transcript is present at 4 copies per cell

- And use it to find β



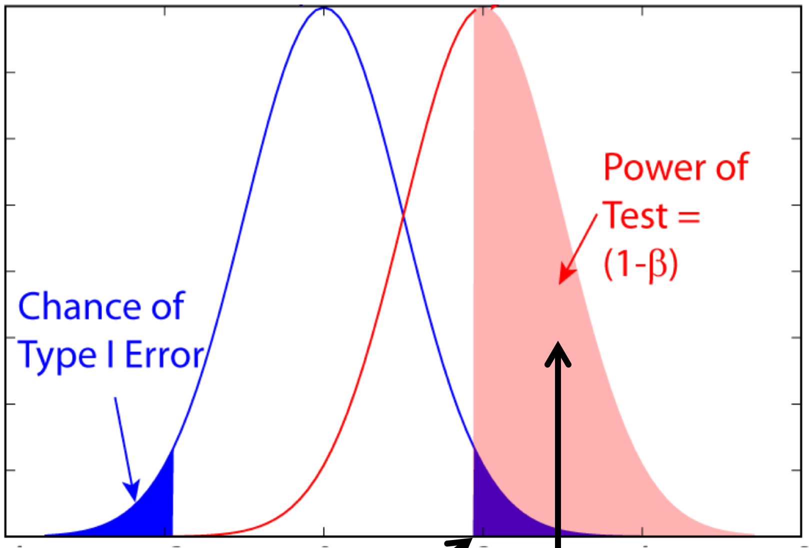
$$H_0 : \mu = 4$$
$$H_1 : \mu = 6$$
$$n = 4$$
$$\alpha = 0.05$$

$$Z = \frac{\bar{x} - \mu_1}{\frac{\sigma}{\sqrt{n}}} = \frac{5.645 - 6}{1} = -0.355$$

$$P(Z > -0.355) = 0.64$$

A Concrete Example

- This illustrates what we could do to improve power: increase sample size, change the effect size or increase α



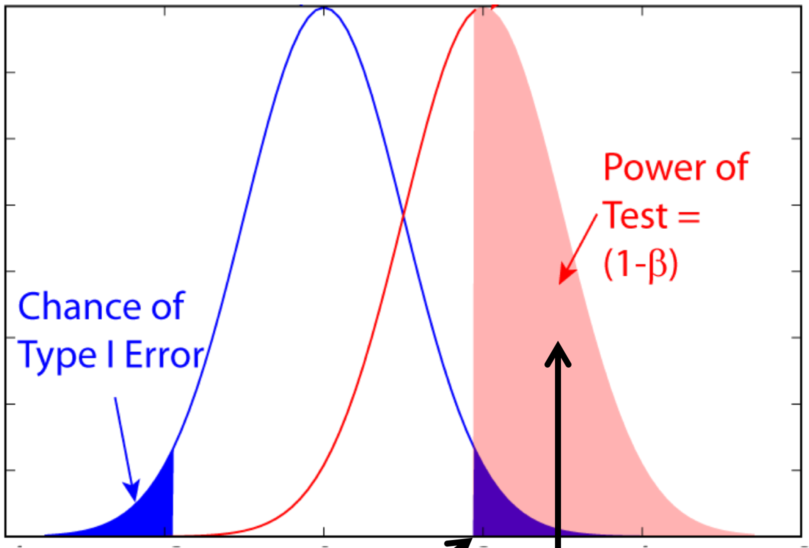
$$H_0 : \mu = 4$$
$$H_1 : \mu = 6$$
$$n = 4$$
$$\alpha = 0.05$$

$$Z = \frac{\bar{x} - \mu_1}{\frac{\sigma}{\sqrt{n}}} = \frac{5.645 - 6}{1} = -0.355$$

$$P(Z > -0.355) = 0.64$$

A Concrete Example

- It also makes clear that we can only calculate 1-β if we have a specific alternate hypothesis



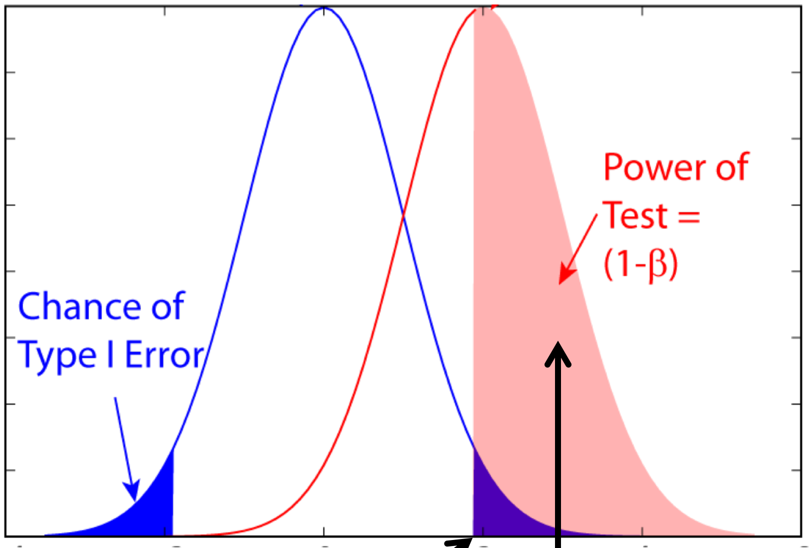
$$H_0 : \mu = 4$$
$$H_1 : \mu = 6$$
$$n = 4$$
$$\alpha = 0.05$$

$$Z = \frac{\bar{x} - \mu_1}{\frac{\sigma}{\sqrt{n}}} = \frac{5.645 - 6}{1} = -0.355$$

$$P(Z > -0.355) = 0.64$$

A Concrete Example

- Of course, we almost never have a specific alternative hypothesis…



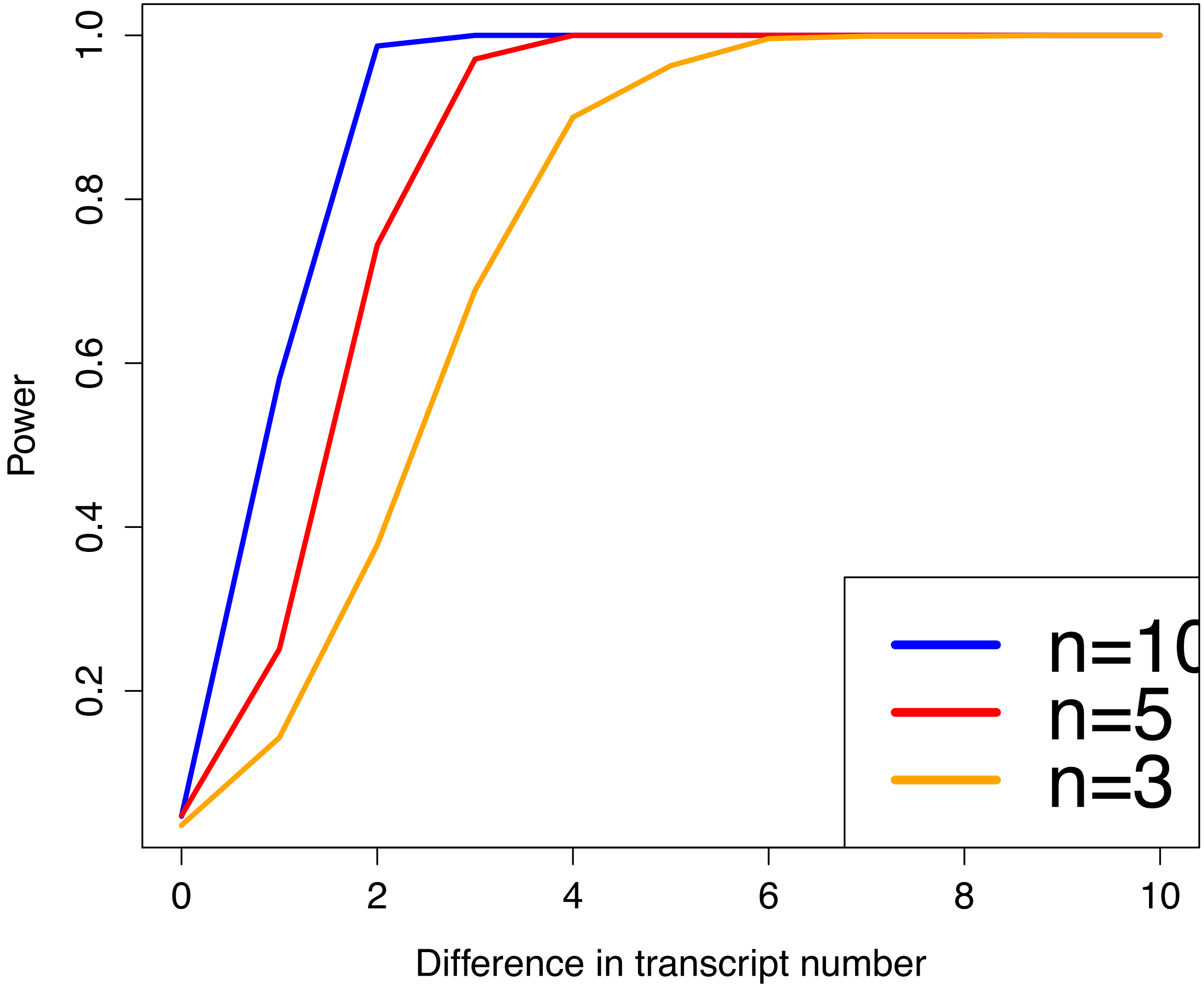
$$H_0 : \mu = 4$$
$$H_1 : \mu = 6$$
$$n = 4$$
$$\alpha = 0.05$$

$$Z = \frac{\bar{x} - \mu_1}{\frac{\sigma}{\sqrt{n}}} = \frac{5.645 - 6}{1} = -0.355$$

$$P(Z > -0.355) = 0.64$$

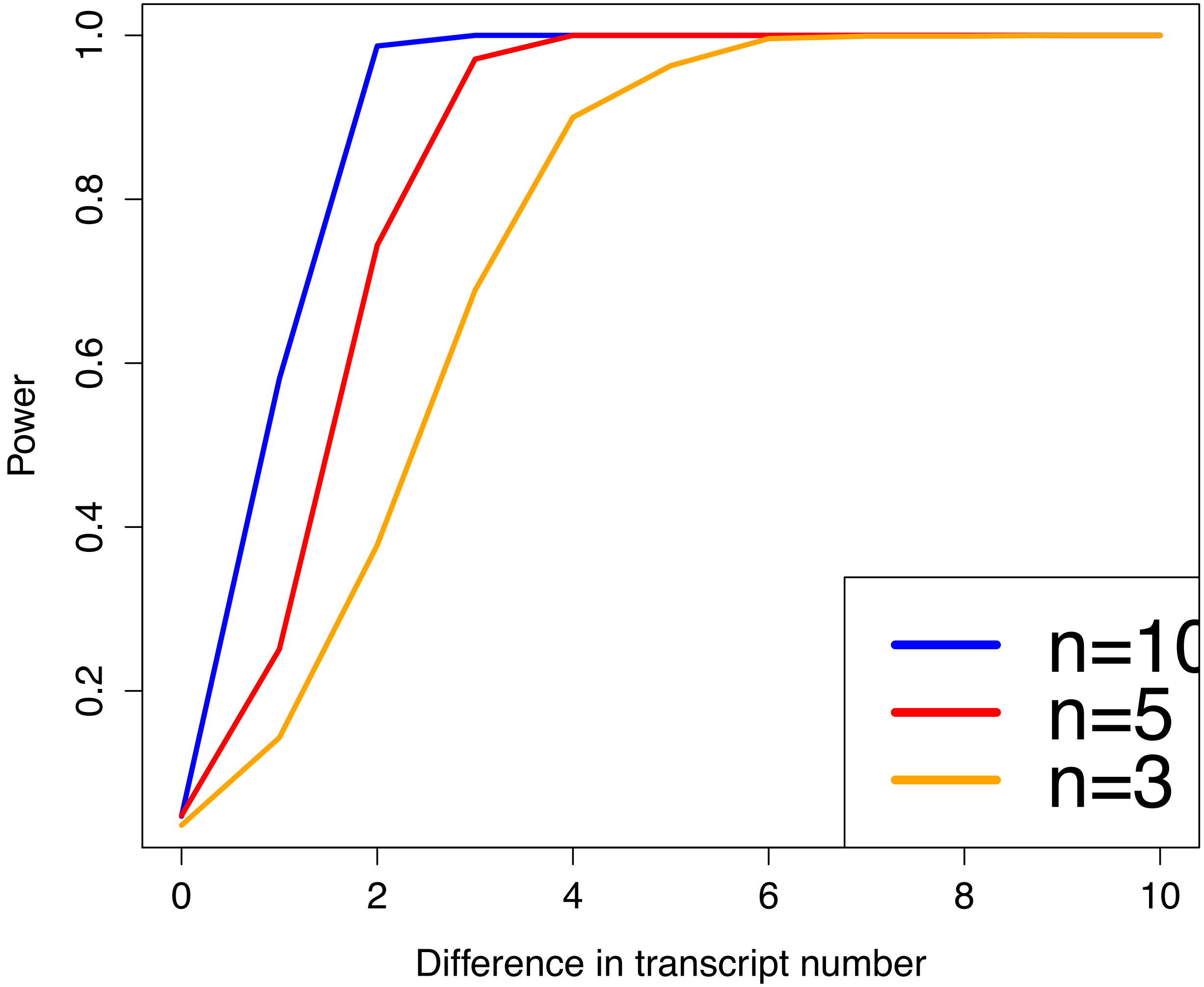
We Examine a Range of Alternate Hypotheses…

- …and generate a power curve



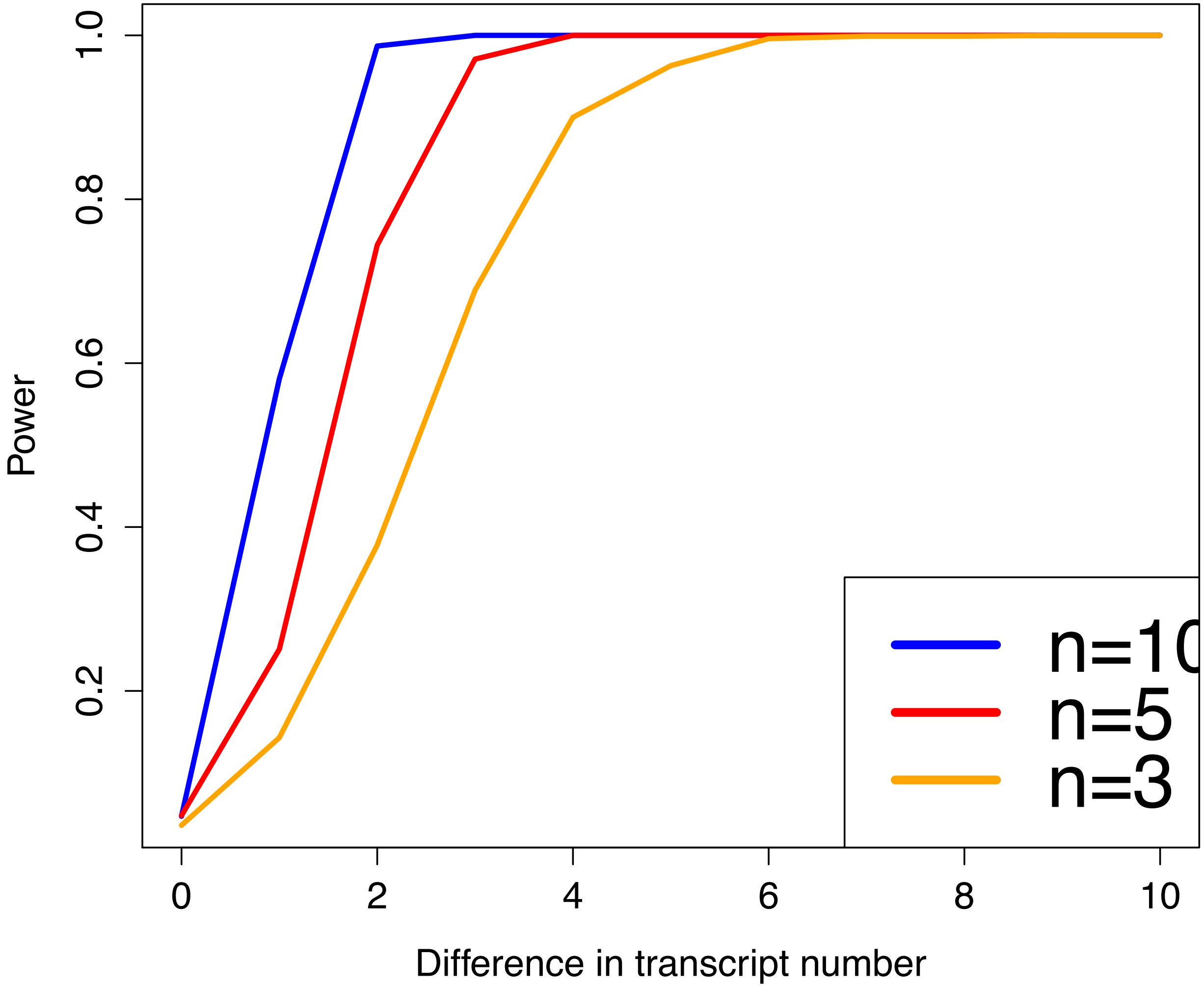
We Examine a Range of Alternate Hypotheses…

- Here we have power vs. effect size, but we could also look at α or variance



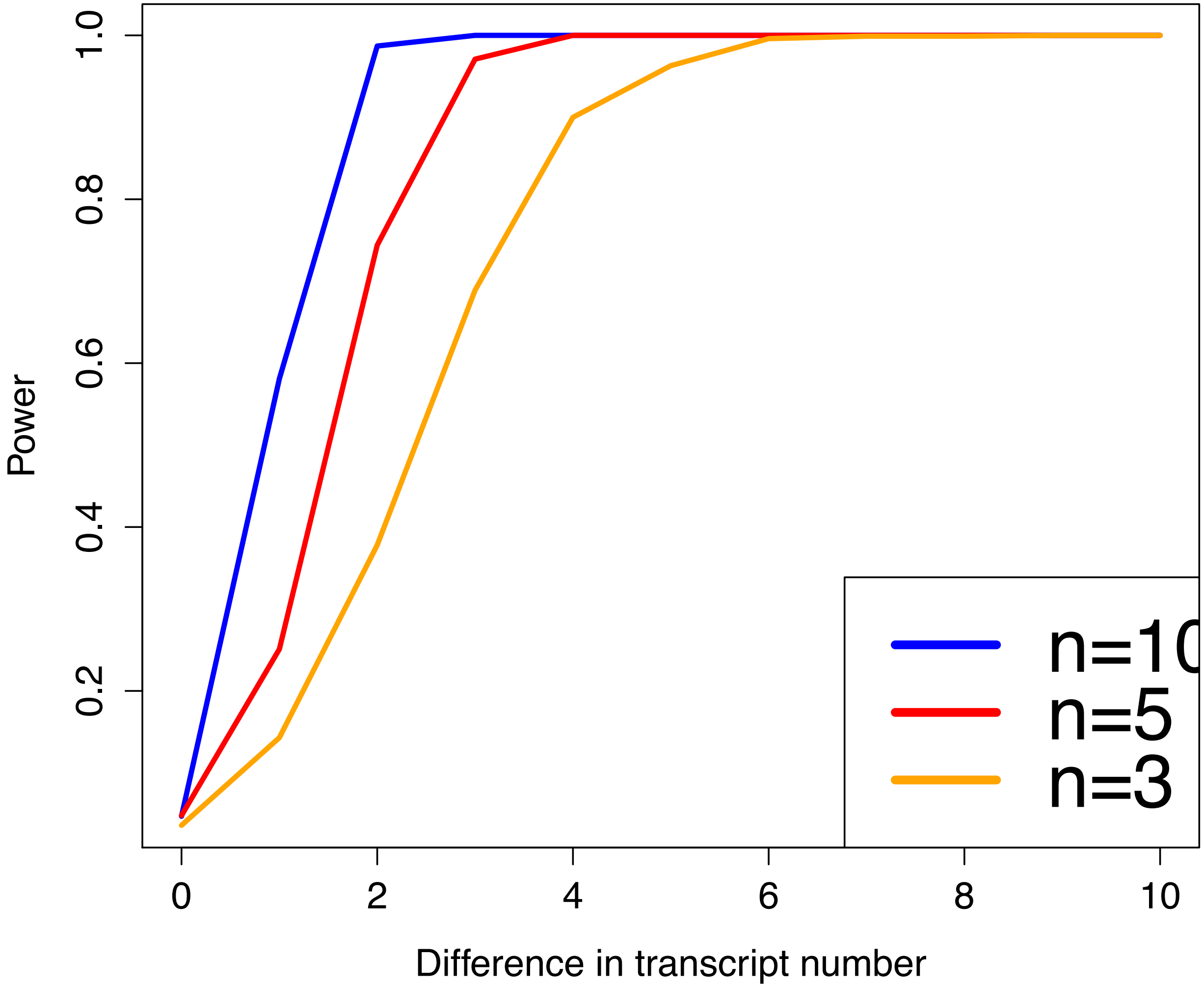
We Examine a Range of Alternate Hypotheses…

- Note that as the two distributions get closer power drops



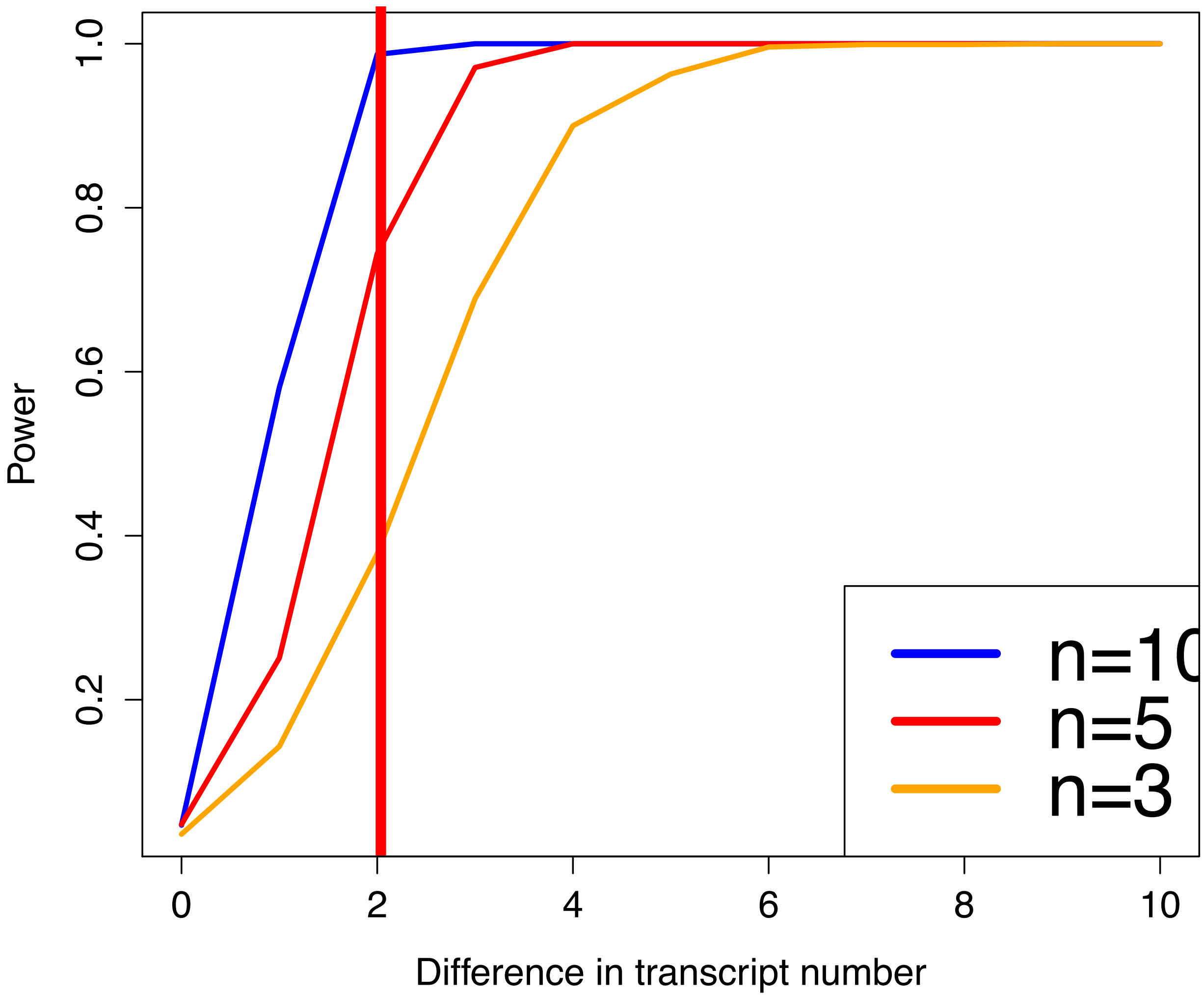
We Examine a Range of Alternate Hypotheses…

- What is 1-β equal to when the two distributions are identical?



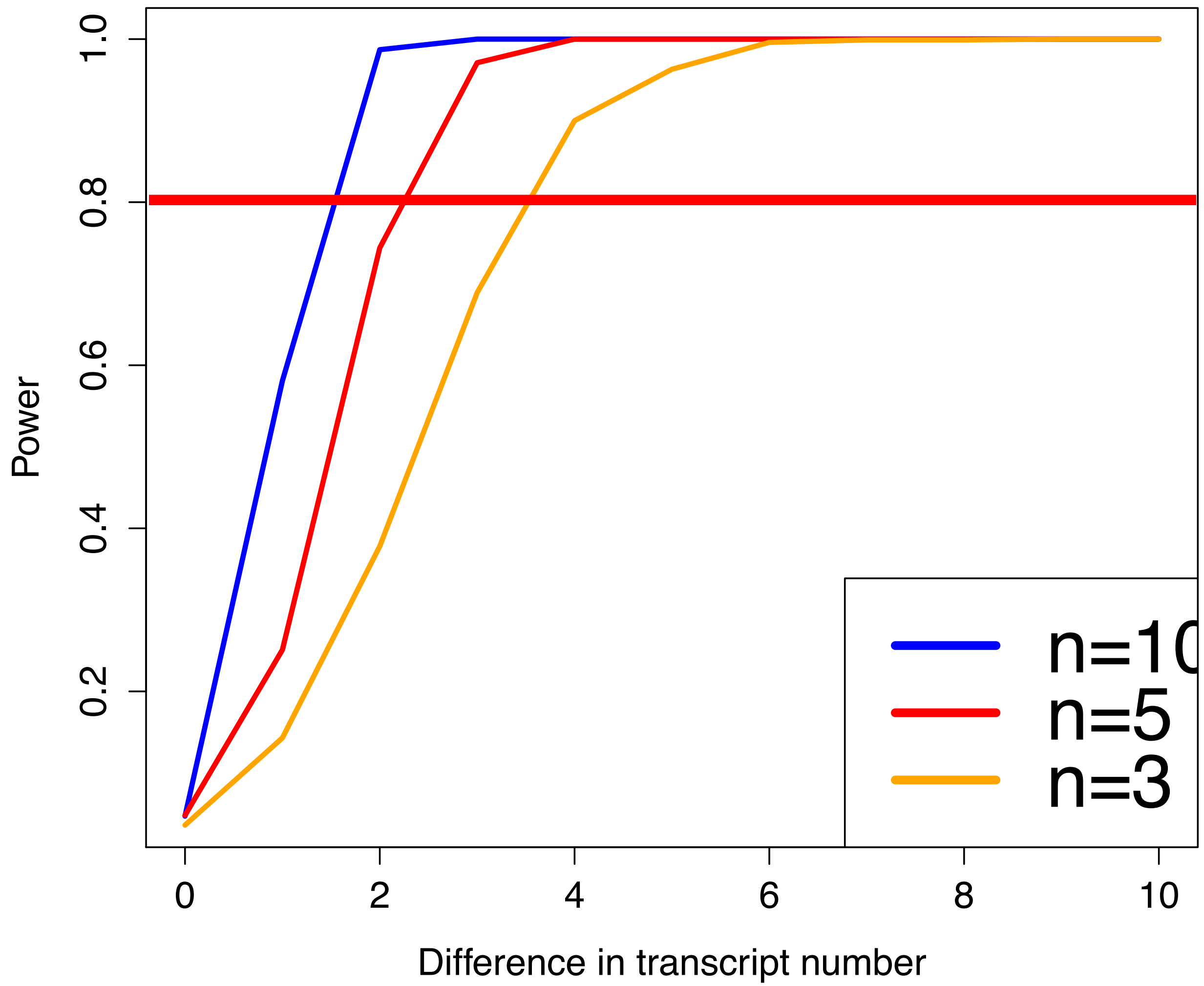
We Can Use Power Analysis to Determine Minimum Sample Size

- Given an effect size (let's say a difference of 2) we want to detect we can determine how many samples we'll need



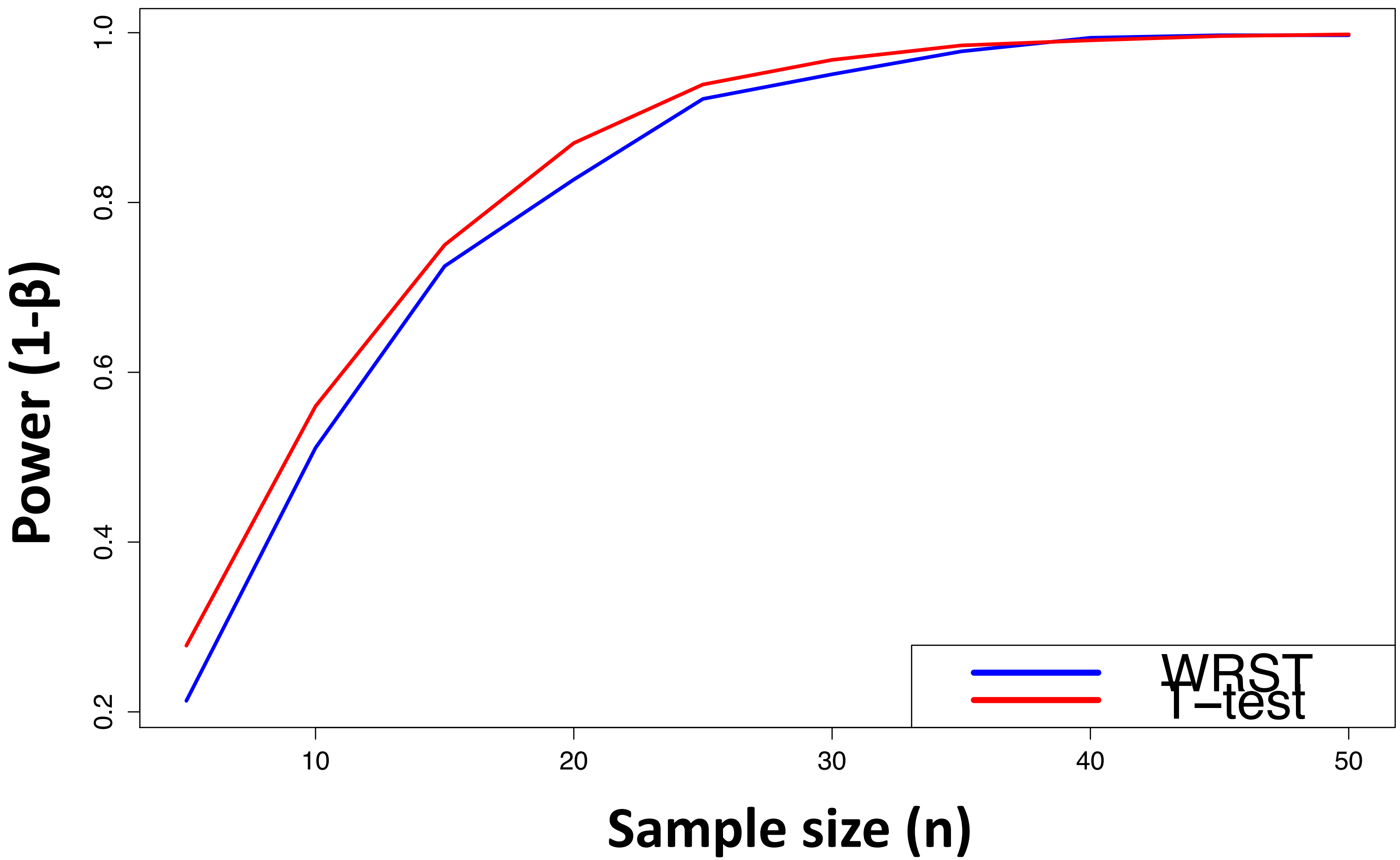
Or We Can Determine the Effect Size We Can Detect Given a Sample

- We have an 80% chance of detecting a difference of ~1 if n=10



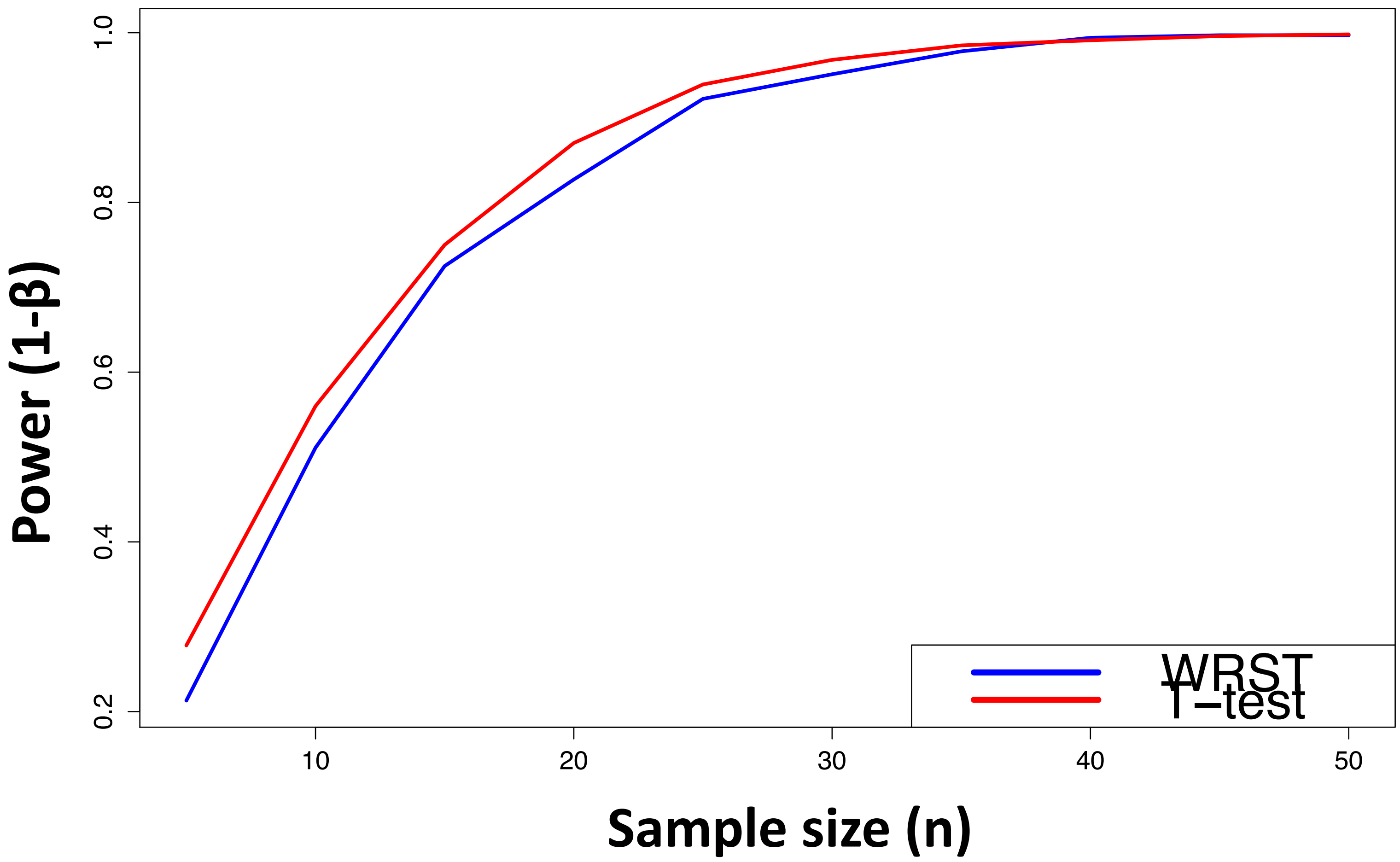
Power Analysis To Compare Tests

Power is a useful concept when deciding which test to employ for a particular task



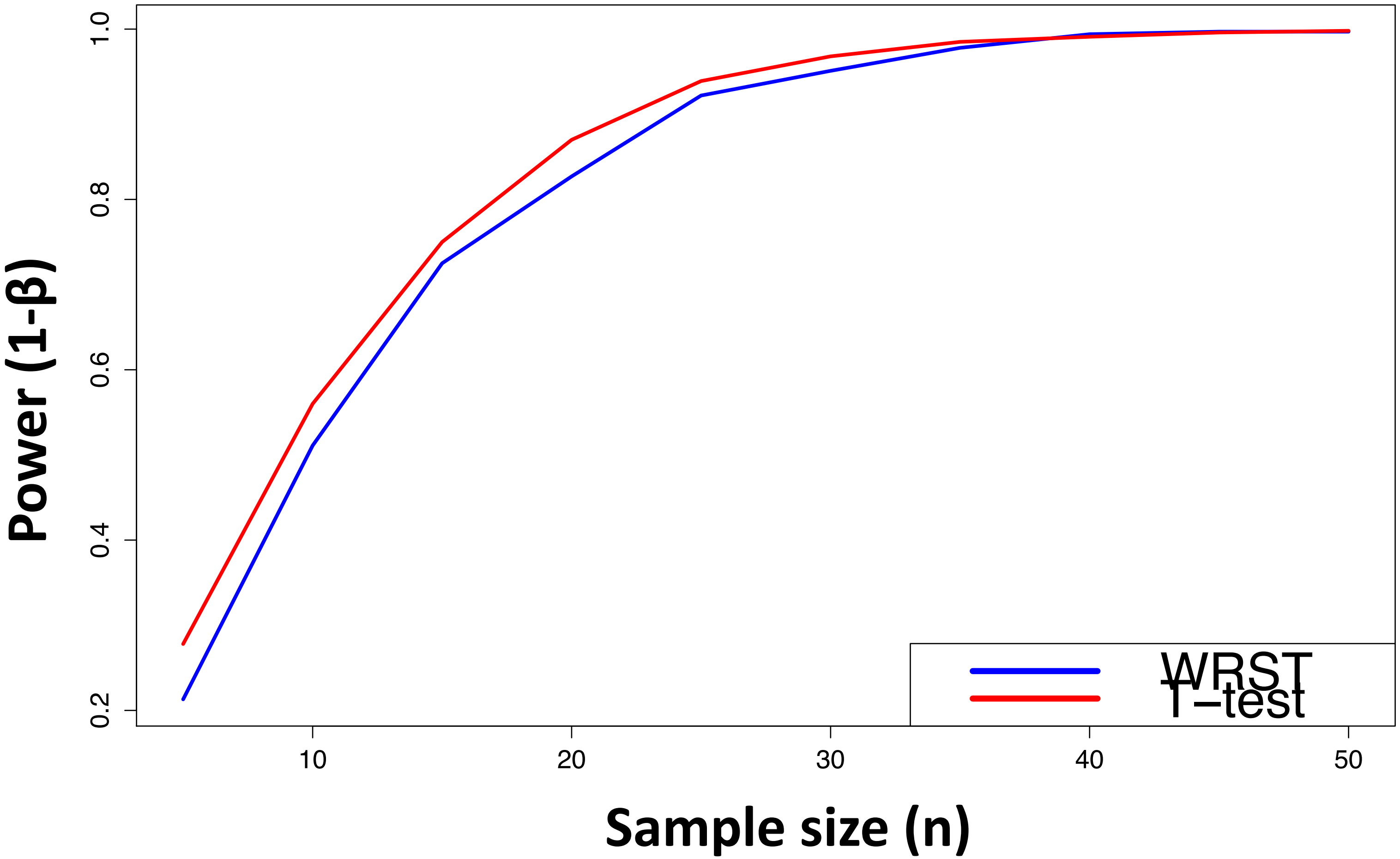
Power Analysis To Compare Tests

To do this type of analysis we calculate power for each test across a range of α, effect sizes and sample sizes



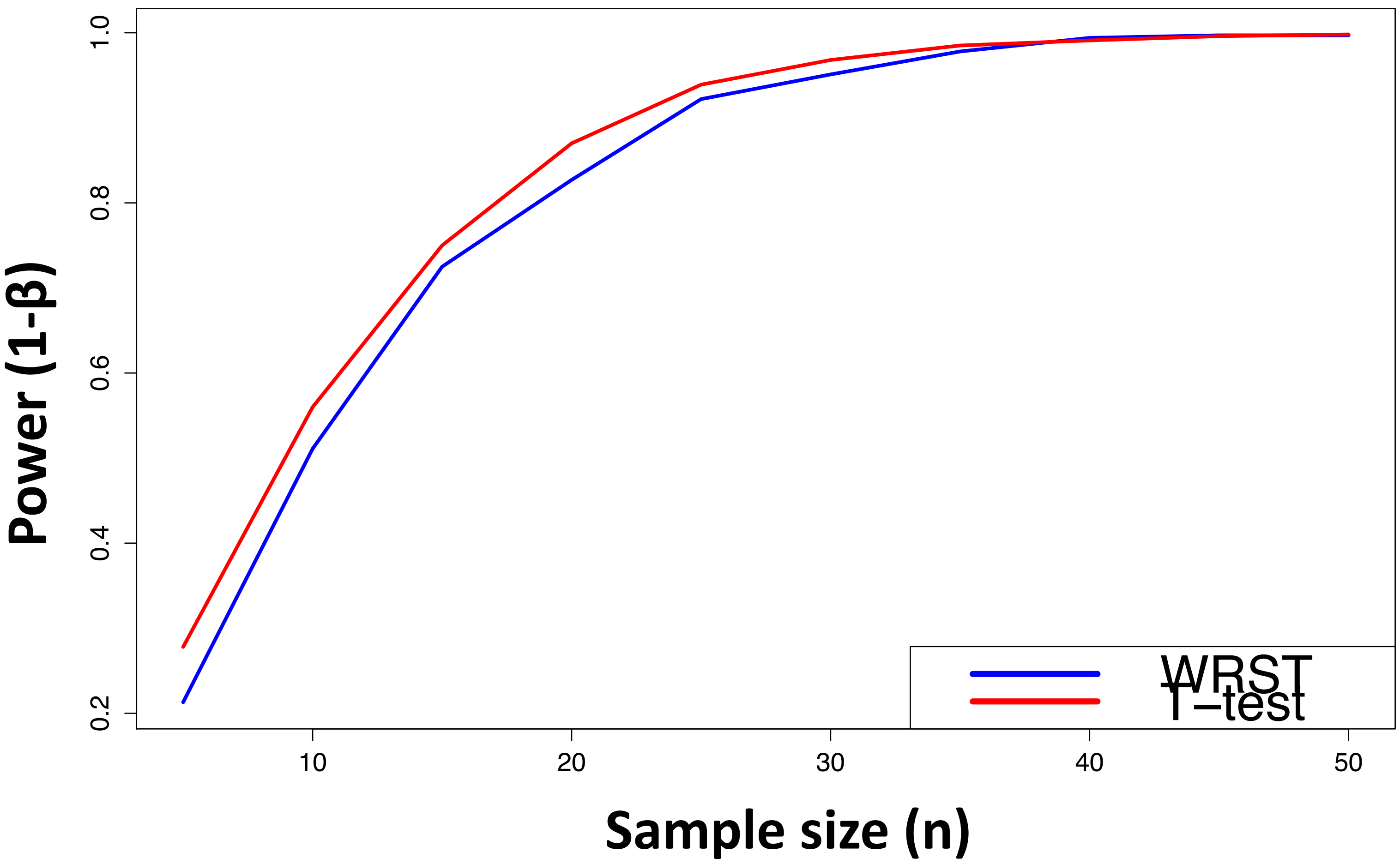
Power Analysis To Compare Tests

Here we compare power for the Mann-Whitney-Wilcoxon test vs. the t-test for samples from two normally distributed pop'ns



Power Analysis To Compare Tests

Note that the t-test is better in all cases, as promised



Why Analyze Power?

Statistical Power, Sample Size, and Their Reporting in Randomized Controlled Trials

David Moher, MSc; Corinne S. Dulberg, PhD, MPH; George A. Wells, PhD

JAMA. 1994;272(2):122-124. doi:10.1001/jama.1994.03520020048013.

Study of 383 randomized clinical trials, of which 102 had negative results

Why Analyze Power?

Statistical Power, Sample Size, and Their Reporting in Randomized Controlled Trials

David Moher, MSc; Corinne S. Dulberg, PhD, MPH; George A. Wells, PhD

JAMA. 1994;272(2):122-124. doi:10.1001/jama.1994.03520020048013. Text Size: A A A

Study of 383 randomized clinical trials, of which 102 had negative results

What fraction of these 102 negative studies had sufficient power to detect a 50% relative difference between treatment and control?

Summary

- Power depends on several factors:
  - Effect size
  - Sample variance
  - α/β
  - The test you perform

- Simply put, power analysis will tell you what effect size you can actually detect given the test you intend to perform

Thus Ends the First Half

- Hopefully this was a helpful introduction

- We covered much of what would be an a yearlong stats course in a month – don’t be afraid to go read more!

- Thanks for your attention, participation and candor

- Please feel free to send suggestions by email, or just stop by my office for a chat

R Goals

- Using the pwr package to do some simple power calculations
- Roll your own power calculations for a variety of functions
- Compare the power of the Mann-Whitney-Wilcoxon and T-tests

Reading/Resources

- http://www.ncbi.nlm.nih.gov/pmc/articles/PMC137461/pdf/cc1521.pdf
- http://www.biostathandbook.com/power.html
- http://www.statmethods.net/stats/power.html