



Confidence Intervals, Effect Size and Power

Chapter 8

The Dangers of NHST

- > We already discussed that statistically significant *does not* always mean practically important.
 - Be especially skeptical if they have very large sample sizes (without any of the information presented in this chapter).

Women are bad at Math.

- > While I joke about being bad at math, I *really dislike* when people tell me women are not good at math.
 - Seriously, I was on the state math team.
- > Cool presentation

Practically Useful Numbers

- > So what do we do to judge a study for the practical significance?
 - Confidence Intervals (small = good)
 - Effect size (big = good)

Confidence Intervals

- > Point estimate: summary statistic – one number as an estimate of the population
 - e.g., mean
 - Just one number rather than an aggregate...

Confidence Intervals

- > Interval estimate: based on our sample statistic, range of sample statistics we would expect if we repeatedly sampled from the same population
 - e.g., Confidence interval
 - Tells you a lot at once!

Confidence Intervals

- > Interval estimate that includes the mean we would expect for the sample statistic a certain percentage of the time were we to sample from the same population repeatedly
 - Typically set at 95 or 99%
 - (hint that matches $p < .05$, $p < .01$)

Confidence Intervals

- > The range around the mean when we add and subtract a margin of error
- > Confirms findings of hypothesis testing and adds more detail

Confidence Intervals

- > Maybe a solution to the problems of NHST?
 - Some argue we should calculate means and CIs and if they *do not* overlap, that means they are different!
 - > Only works sometimes though ...

Confidence Intervals

> The calculations are not that different than what we've been doing:

- Take a z-score where 95% of the data is in the middle of the distribution
- Translate that back to a raw score to match the mean.
- (draw)

Calculating Confidence Intervals

- > Step 1: Draw a picture of a distribution that will include confidence intervals
- > Step 2: Indicate the bounds of the CI on the drawing
- > Step 3: Determine the z statistics that fall at each line marking the middle 95%
- > Step 4: Turn the z statistic back into raw means
- > Step 5: Check that the CIs make sense

Confidence Interval for z-test

$$M_{lower} = - z(\sigma_M) + M_{sample}$$

$$M_{upper} = z(\sigma_M) + M_{sample}$$

> Think about it: Why do we need two calculations?

CI in R

```
> 42.5 + qnorm(.05/2, lower.tail = T)*12
```

```
> 42.5 + qnorm(.05/2, lower.tail = F)*12
```

- Made up numbers
- Notice the QNORM
 - > Lower tail = T gives you the lower confidence interval
 - > Lower tail = F gives you the upper confidence interval.

Example

- > Let's compare the average calories consumed by patrons of Starbucks that posted calories on their menus to the population mean number of calories consumed by patrons of Starbucks that do not post calories on their menus (Bollinger, Leslie, & Sorensen, 2010).
- > The population mean was 247 calories, and we considered 201 to be the population standard deviation. The 1000 people in the sample consumed a mean of 232 calories.

Think About It

- > Which do you think more accurately represents the data: point estimate or interval estimate?
 - Why?
- > What are the advantages and disadvantages of each method?

Effect Size: Just How Big Is the Difference?

- > Significant = Big?
- > Significant = Different?
- > Increasing sample size will make us more likely to find a statistically significant effect

Effect Size: Just How Big Is the Difference?

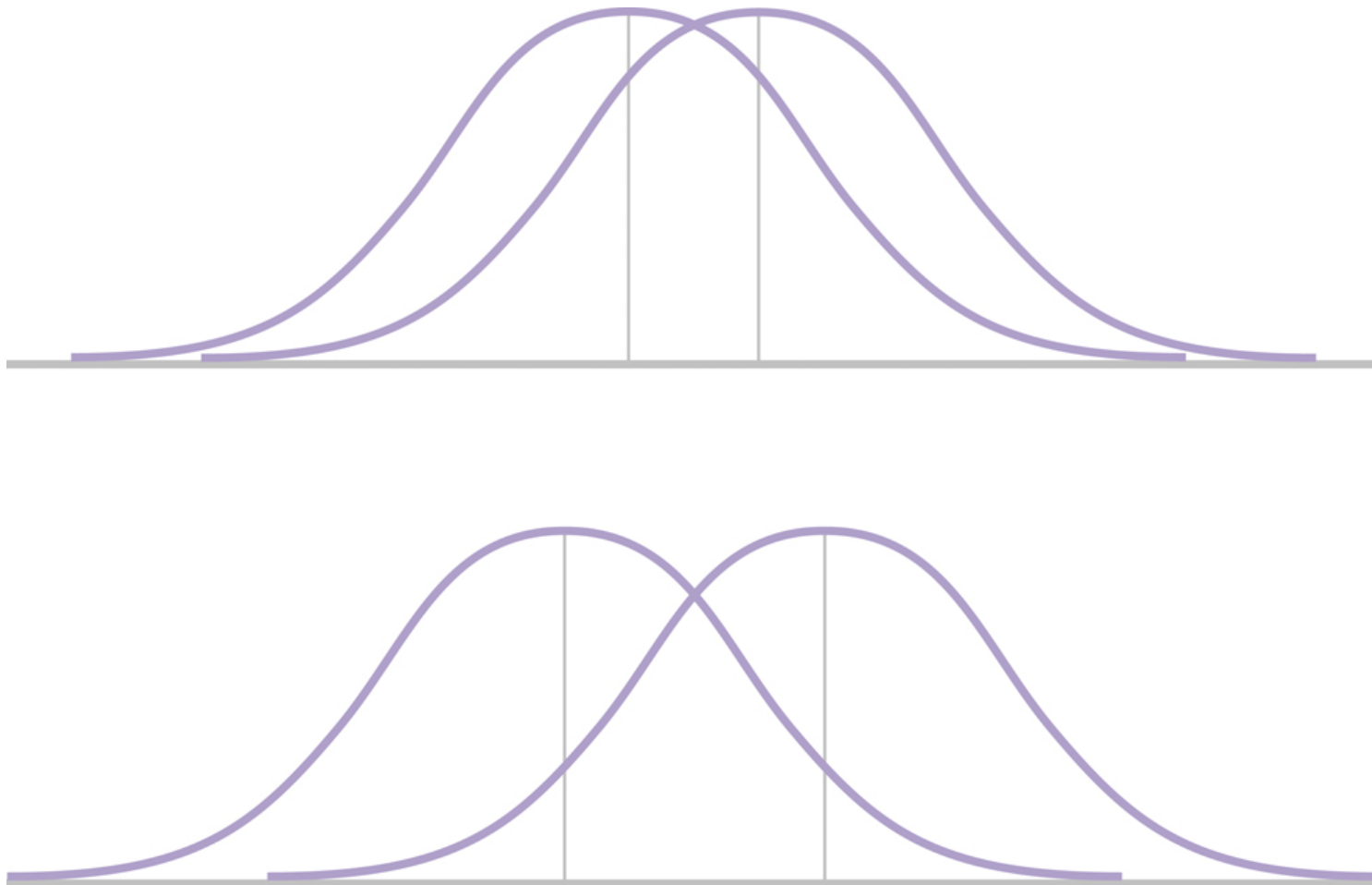
	Small Effect Size	Big Effect Size
Fail to reject null	Probably not there, sample size not issue	Important! May not have enough people in study
Reject null	Probably not there, sample size may be causing rejection (Type 1)	Important and we found it! (power).

What is effect size?

- > Size of a difference that is unaffected by sample size
 - How much two populations *do not* overlap
- > Standardization across studies

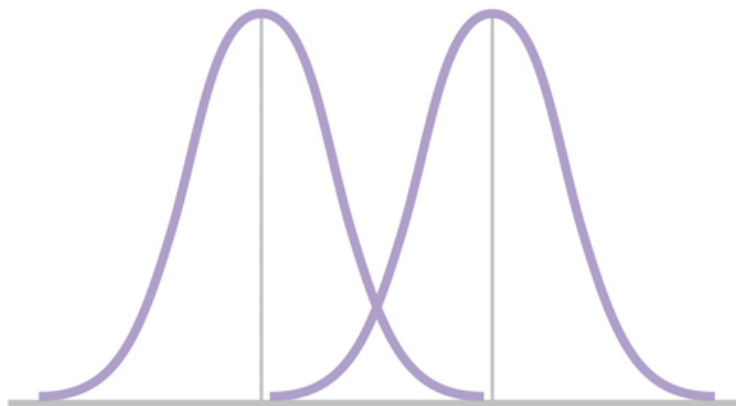
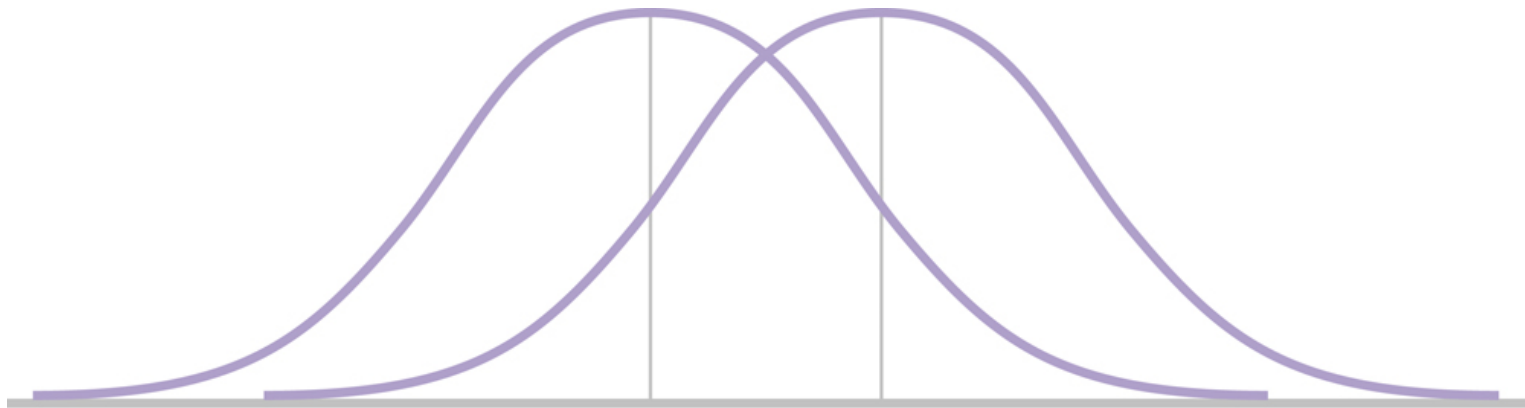
Effect Size & Mean Differences

Imagine both represent significant effects: Which effect is bigger?



Effect Size and Standard Deviation

Note the spread in the two figures:
Which effect size is bigger?



Increasing Effect Size

- > Decrease the amount of overlap between two distributions:
 1. Their means are farther apart
 2. The variation within each **population** is smaller

Calculating Effect Size

- > Cohen's d Estimates effect size
- > Assesses difference between means using standard deviation instead of standard error

$$d = \frac{M - \mu}{\sigma}$$

Example

- > Let's compare the average calories consumed by patrons of Starbucks that posted calories on their menus to the population mean number of calories consumed by patrons of Starbucks that do not post calories on their menus (Bollinger, Leslie, & Sorensen, 2010).
- > The population mean was 247 calories, and we considered 201 to be the population standard deviation. The 1000 people in the sample consumed a mean of 232 calories.

TABLE 8-1. Cohen's Conventions for Effect Sizes: d

Jacob Cohen has published guidelines (or conventions), based on the overlap between two distributions, to help researchers determine whether an effect is small, medium, or large. These numbers are not cutoffs, merely rough guidelines to aid researchers in their interpretation of results.

Effect Size	Convention	Overlap
Small	0.2	85%
Medium	0.5	67%
Large	0.8	53%

Statistical Power

- > The measure of our ability to reject the null hypothesis, given that the null is false
 - The probability that we will reject the null when we should
 - The probability that we will avoid a Type II error

Power Rangers Example

> See handout!

Calculating Power

- > Step 1: Determine the information needed to calculate power
 - Population mean (μ)
 - Standard error (N , σ needed)
 - Sample mean found (M)

Calculating Power

- > Step 2: Determine a critical z value (cut off score)
- > Calculate raw score for that z:
- > $M_{needed} = \mu_m + (\sigma_m) * z$

Calculating Power

> Step 3: Figure out the distance between the M needed and sample M you estimated

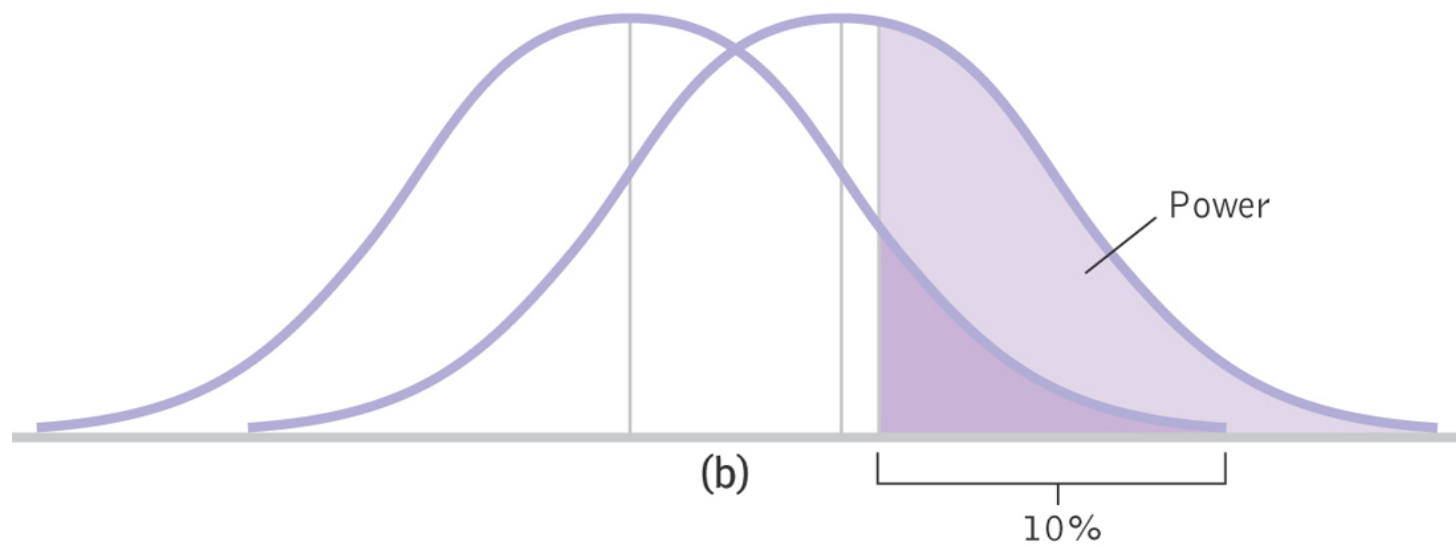
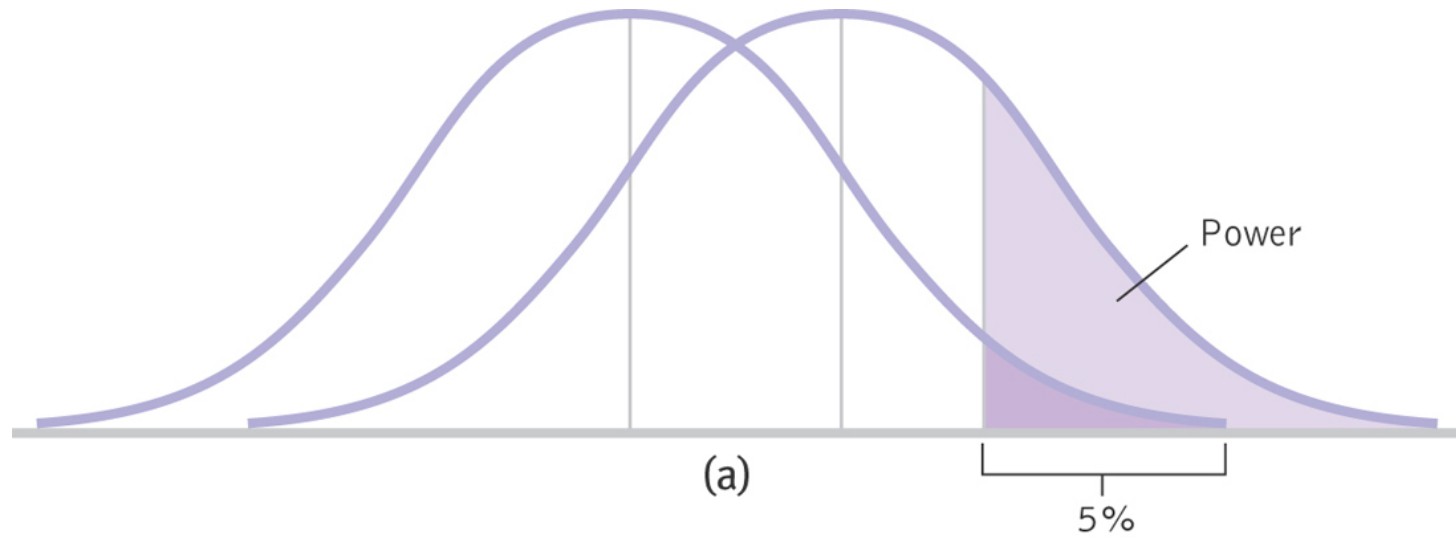
- $Z = (M_{\text{needed}} - M_{\text{sample}}) / \sigma_m$

> Then the percentage of the distribution above/below that score is power

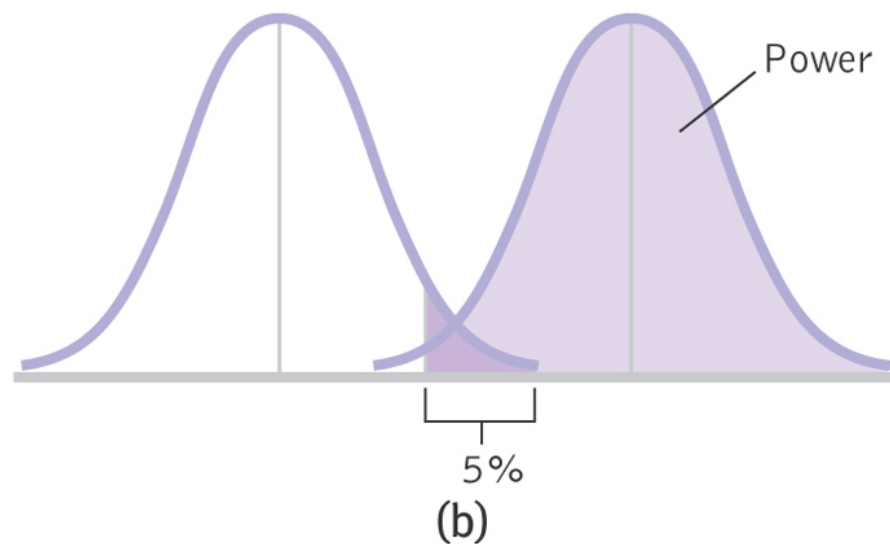
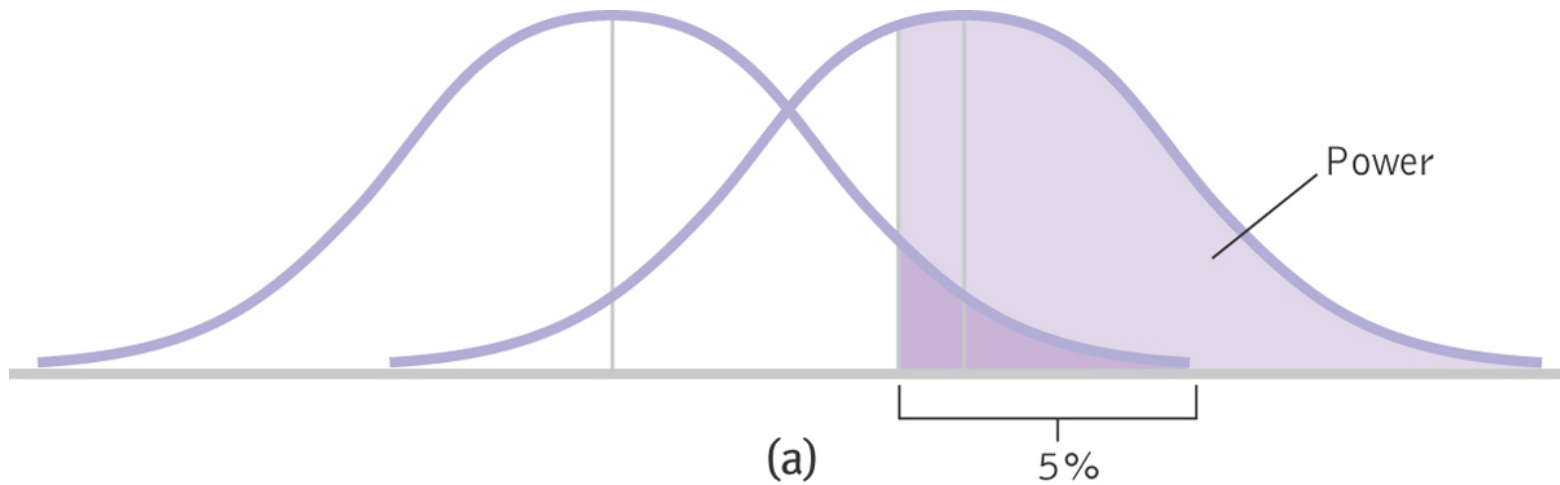
- (matches your hypothesis of higher or lower)

What increases power?

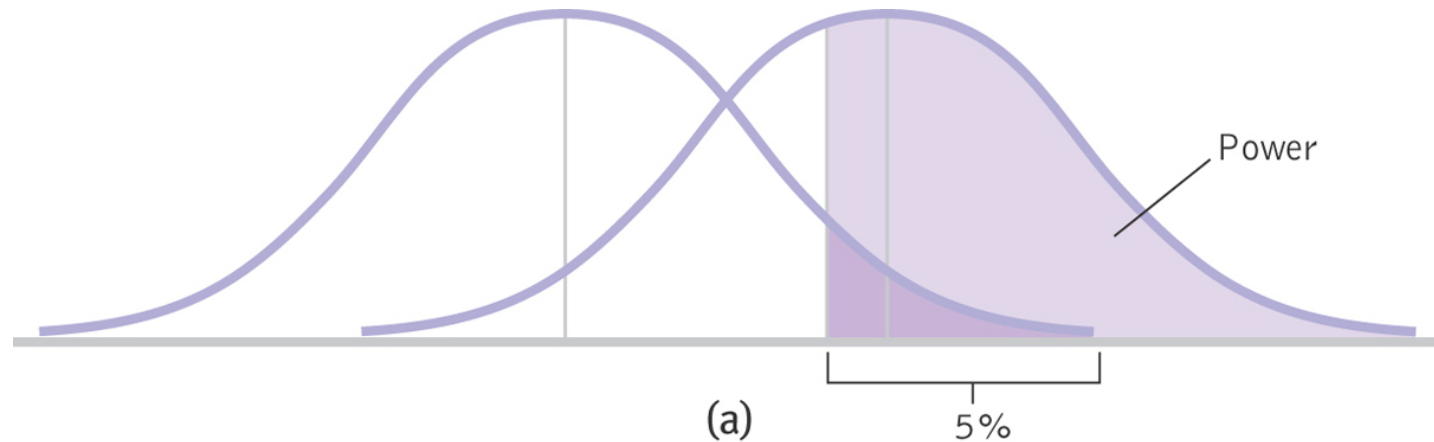
Increasing Alpha



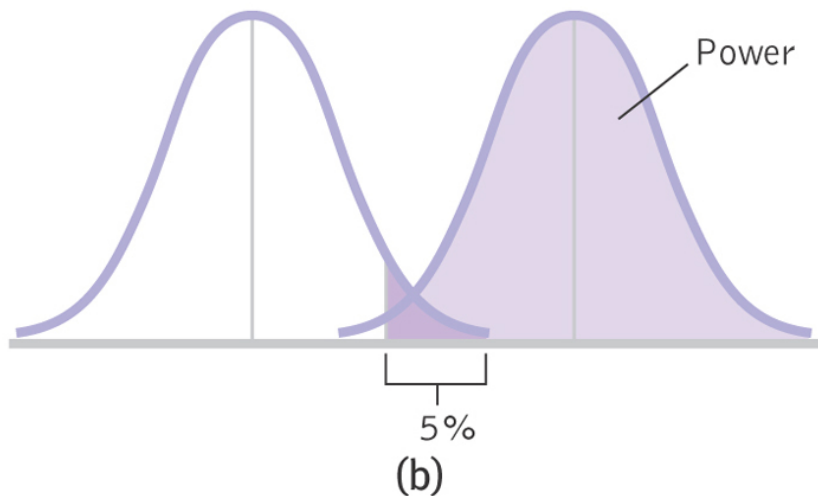
Two-Tailed Verses One-Tailed Tests



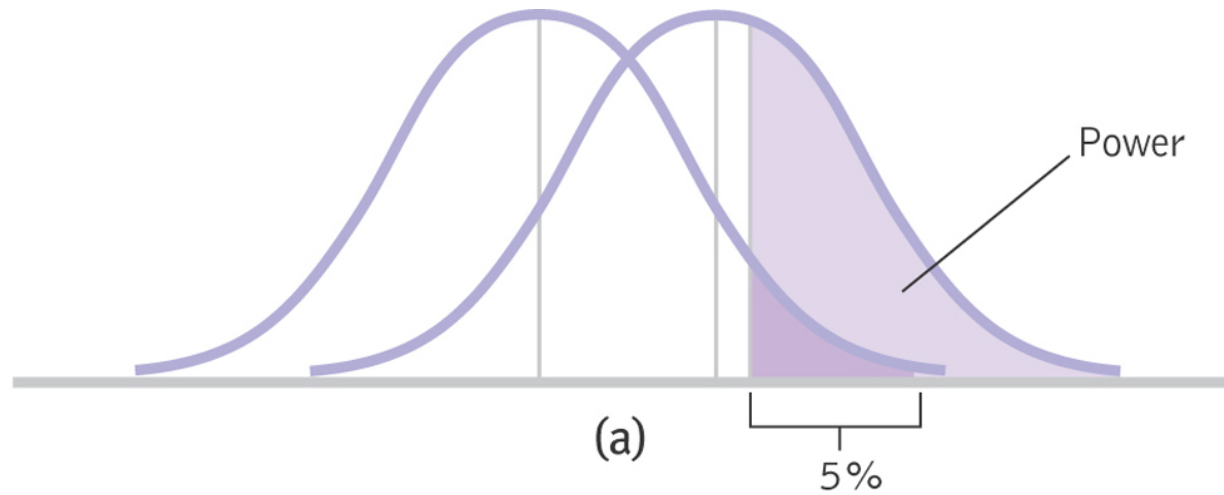
Increasing Sample Size or Decreasing Standard Deviation



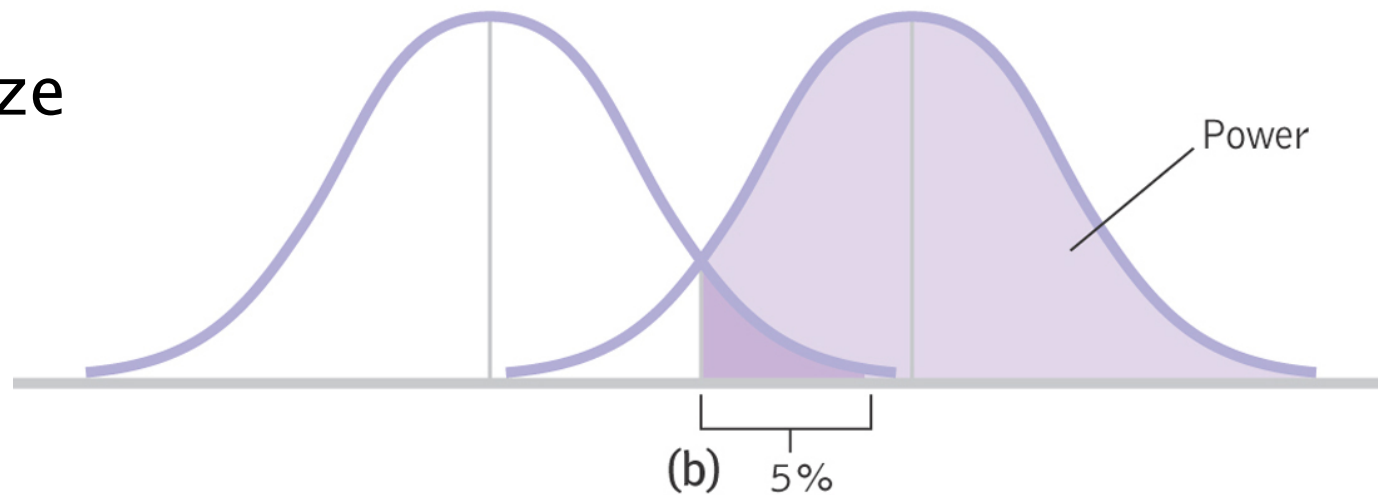
Both of
these
decrease
standard
error



Increasing the Difference between the Means



Effect size



Factors Affecting Power

- > Larger sample size increases power
- > Alpha level
 - Higher level increases power (e.g., from .05 to .10)
- > One-tailed tests have more power than two-tailed tests
- > Decrease standard deviation
- > Increase difference between the means

Think About It

- > Why would it be important to know the statistical power of your study?

Meta-Analysis

- > Meta-analysis considers many studies simultaneously.
- > Allows us to think of each individual study as just one data point in a larger study.

The Logic of Meta-Analysis

- > STEP 1: Select the topic of interest
- > STEP 2: Locate every study that has been conducted and meets the criteria
- > STEP 3: Calculate an effect size, often Cohen's *d*, *for every study*
- > STEP 4: Calculate statistics and create appropriate graphs

