JIPP Lecture #4

Statistics Review

- Slides are at socialneuro.com\chula
- Follow me on Twitter: @socialneuro
- My email address: e.vanman@uq.edu.au

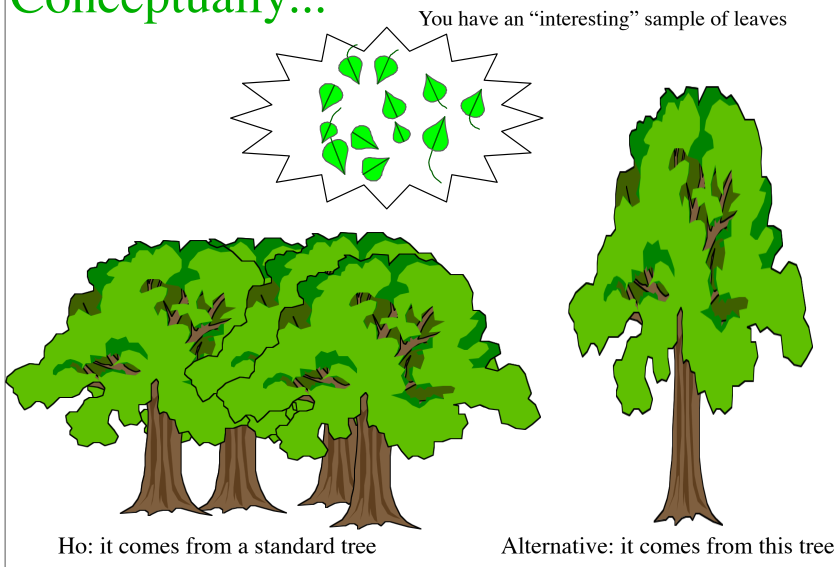
This Week:

- Hypothesis Testing with the *z*-test
  - The Null and Alternative Hypotheses
  - Errors in Hypothesis Testing
  - Increasing Power

Hypothesis Testing and Sampling

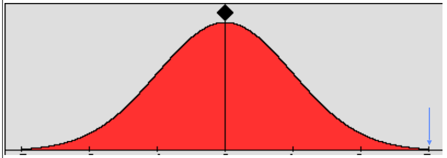
- Hypothesis Testing relies on the concepts of *sampling*, *populations*, and *probability*.
- The basic logic: When we use a sample to make inferences about population parameters, we want to know whether our sample comes from a population that is best described by the null hypothesis.
- The null hypothesis is a statement which we usually are attempting to *reject* in our research.

Conceptually...



An Example of Hypothesis Testing with Statistics

A group of Norwegian sailors has an average pulse of 98 beats per minute. Is this *significant* (that is, are they not healthy)?

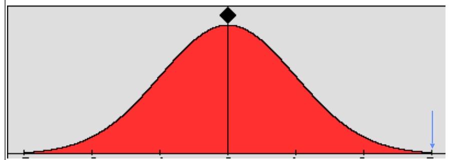


Null Hypothesis: μ=73

A "healthy" population has a mean pulse of 73 bpm, with an SD of 8. So, 98 beats per minute would be 3.125 standard deviations above the mean.

An Example of Hypothesis Testing with Statistics

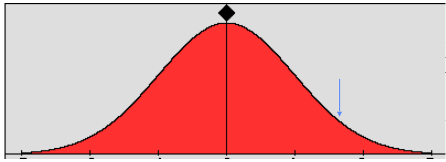
A group of Norwegian sailors has an average pulse of 98 beats per minute. Is this *significant* (that is, are they not healthy)?



Null Hypothesis: μ=73

A "healthy" population has a mean pulse of 73 bpm, with an SD of 8. So, 98 beats per minute would be 3.125 standard deviations above the mean.

--or--



Alternative Hypothesis: $\mu \neq 73$

An "unhealthy" population that has a higher mean pulse, with an SD of 8. So, 98 beats per minute would be closer to the mean.

Five Steps of Hypothesis Testing

1. Make some statements about the population parameters: a null hypothesis and an alternative (research) hypothesis.
2. Determine the population parameters assuming the null hypothesis is true.
3. Determine a "cut-off" point in the population at which the null hypothesis should be rejected.
4. Determine the probability of your sample statistic assuming the null hypothesis is true.
5. If that probability exceeds the "cut-off" point, reject the null hypothesis. If it doesn't, retain it.

An Example: Hypothesizing about a Single Sample Score

- A certain depression questionnaire can have scores that range from 1 (*not at all depressed*) to 50 (*extremely depressed*). When given to a population of "normal" adults, μ=20 and σ = 5.
- Joe, an air traffic controller, completes the questionnaire, and gets a score of 42. Is this "normal?" Should he be considered at possible risk for depression?

Step #1. State the Null Hypothesis and Research Hypothesis

Example: Joe's depression score is 42 on a scale that is 1-50.

Ho: Joe comes from a population with μ=20.

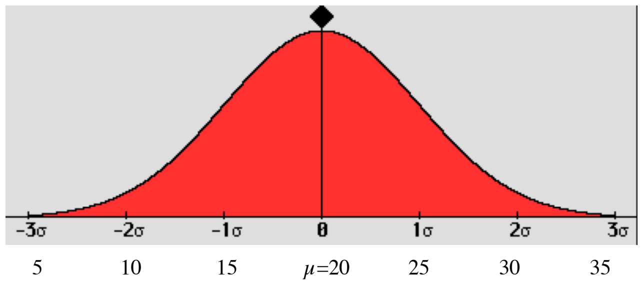
Ha: Joe comes from a population with $\mu \neq$20.

- NOTICE: the alternative (research) hypothesis is the logical complement to the null hypothesis (i.e., it should cover all instances not covered by Ho).

Step #2. Determine the Population's parameters

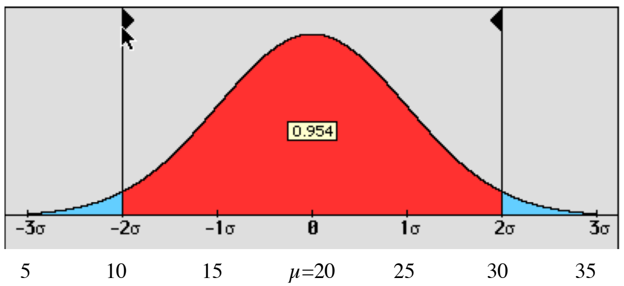
Example: Joe's depression score is 42 on a scale that is 1-50.

- If Ho is true, μ= 20, and σ = 5



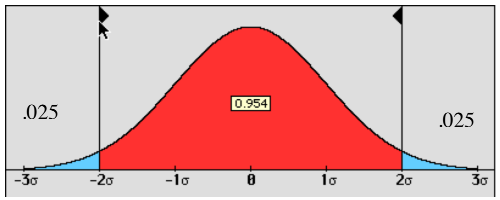
Step #3. Determine a "cut-off" statistic at which to reject Ho.

- We need to choose a "level of risk." Typically, researchers use p = .05 as the cut-off level.
- We then need to find the cut-off point, where a given parameter occurs less than .05 of the time, assuming Ho is true.



Step #3 (cont.) A Slight Detour: One-tailed or Two-tailed?

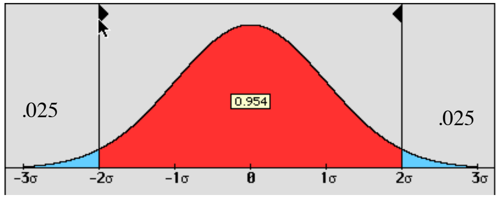
- We have the option of how to "divvy up" the .05 area:



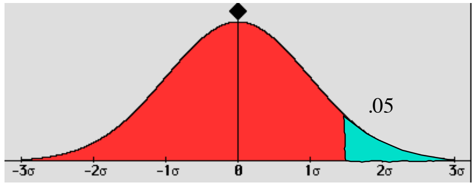
"Two-tailed"

Step #3 (cont.) A Slight Detour: One-tailed or Two-tailed?

- We have the option of how to "divvy up" the .05 area:



OR...

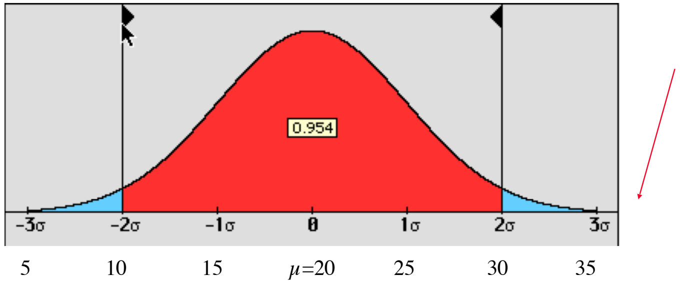


"One-tailed"

Step #4. Determining the probability of your sample score if Ho is true.

Example: Joe's depression score is 42 on a scale that is 1-50.

- If Ho is true, $\mu=20$, $\sigma=5$.
- Joe's score is z= (X-Mean)/SD = (42-20)/5 = 4.4



Step #5. Reject or Retain Ho, depending on the probability of your sample statistic

- Because Joe's Z-score exceeds the "cut-off," reject Ho, and adopt the alternative hypothesis.
- Joe therefore comes from a population where $\mu \neq 20$.
- We can say "Joe's score significantly differs from the normal population of depression scores, $p<.05$"

Choosing between H_0 and H_1

In choosing between the two hypotheses, we assume H_0 to be true.

↓

We test statistically the probability of obtaining our observed results if H_0 is actually true

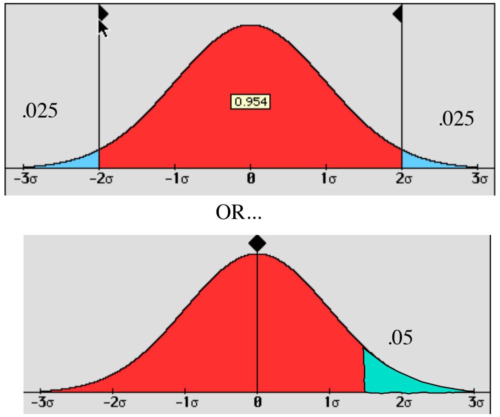
↓

If the probability is low	If the probability is high
Reject H_0	Retain H_0
↓	↓
Infer observed difference is genuine	Infer observed difference is due to chance

How Unlikely does our observation have to be?

- Conventionally
  - $p<0.05$
  - 5%
  - 1 in 20
- More conservative criteria
  - $p<0.01$
  - 1%
  - 1 in 100

Returning to the issue of One-tailed vs. Two-tailed



General Guidelines for Deciding Between One-Tailed versus Two-Tailed Tests

- If your research hypothesis is nondirectional ("there is a difference, but I don't know the direction of the difference"), use a two-tailed test.
- If your research hypothesis is directional ("A is greater than B"), then you can use a one-tailed test, but keep in mind that
  - You're increasing the chance to reject Ho
  - And if you choose the *wrong* direction, you won't be able to reject Ho.
- So, it's most common for people pick the two-tailed test even when they have a directional hypothesis.

Hypothesis Testing of <u>Means</u> When σ is known.

- One-sample *z*-test

Five Steps of Hypothesis Testing

1. Make some statements about the population parameters: a null hypothesis and an alternative (research) hypothesis.
2. Determine the population parameters assuming the null hypothesis is true.
3. Determine a "cut-off" point in the population at which the null hypothesis should be rejected.
4. Determine the probability of your sample statistic assuming the null hypothesis is true.
5. If that probability exceeds the "cut-off" point, reject the null hypothesis. If it doesn't, retain it.

The *z*-test. An Example

- Example: Kirsty and two other teachers are teaching a total of 100 fifth grade students at Oakwood Primary. After the students are given an IQ test, she finds that the sample mean is 110.
- For the IQ test, μ=100 and σ=15.
- Are Kirsty's students significantly "different" from average?

Step 1. Statements about parameters.

- Research Ha: The students at Oakwood come from a population where $\mu \neq 100$.
- Ho: μ=100.

Step 2. Assuming Ho, describe the population.

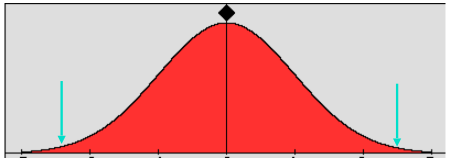
- In this case, we'll work with the *distribution of sample means* for samples of 100 students each.

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\bar{X}} = \frac{15}{\sqrt{100}} = \frac{15}{10} = 1.5$$

$\mu = 100$ which is also the mean of the distribution of sample means

Step 3. Determine a “cut-off”



Null Hypothesis: μ=100
for a distribution of means for N=100

This is a bidirectional hypothesis, so we need to determine the Z-scores that represent the “critical region” where we would reject. Let’s assume the usual 5% significance level.
From Table B.1, we know that Z=1.96 is the point where .025 of the scores in the tail beyond. If our sample statistic exceeds that point (in either direction), we’ll reject Ho.

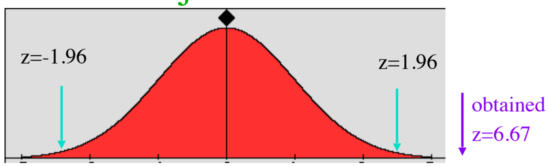
Step 4. Determine the probability of sample statistic assuming Ho is true.

- We need to determine the z-score for our sample mean assuming Ho is true.

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

$$Z = \frac{110 - 100}{1.5} = 6.67$$

Step 5. Compare the results of steps 3 & 4 to determine whether to reject or retain Ho.



Null Hypothesis: μ=100
for a distribution of means for N=100

- Our Z, 6.67, is far greater than the upper cut-off score. So this means the chance of getting a Z value of 6.67 is extremely rare if Ho is true.
- So, we reject Ho.
- Kirsty’s class has a significantly higher mean IQ score, z=6.67, p<.05

Now, More Generally Speaking...

- What about any errors that we could make “just by chance?”

Errors in Hypothesis Testing

The Researcher concludes...

But reality is...

	Reject	Retain
Reject	GOOD! (1-β)	Type II Error (β)
Retain	Type I Error (α)	GOOD

Errors in Hypothesis Testing

The Researcher concludes...

But reality is...

	Reject	Retain
Reject	GOOD! (1-β)	Type II Error (β)
Retain	Type I Error (α)	GOOD

Type I Error

- The probability of rejecting Ho when Ho is true.
- Known as alpha (α)
- Same as significance level!
- When we say p<.05, we're stating that the Type I error is less than 5%.

Errors in Hypothesis Testing

The Researcher concludes...

But reality is...

	Reject	Retain
Reject	GOOD! (1-β)	Type II Error (β)
Retain	Type I Error (α)	GOOD

Type II Error

- The probability of retaining Ho when Ho is false.
- Known as Beta or β

The Relationship of Type I and Type II Errors

- Generally, as you decrease α, you increase β
- Generally, as you increase α, you decrease β
- The choice of an alpha of .05 or .01 is based on a compromise of the two types of Error

Not Making an Error

The Researcher concludes...

But reality is...

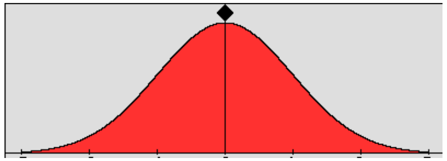
	Reject	Retain
Reject	GOOD! (1-β)	Type II Error (β)
Retain	Type I Error (α)	GOOD

Power

- Power is the probability that a researcher will (correctly) reject Ho when Ho is false

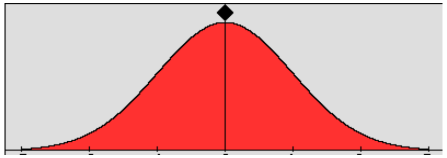
A “Visual” Look at Power

Distribution of Sample Means for Population when Ho is rejected (REALITY)



μ=200

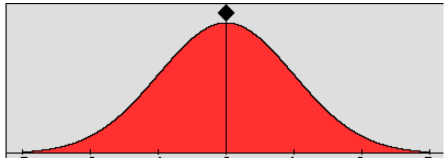
Distribution of Sample Means for Population if Ho were True



μ=100

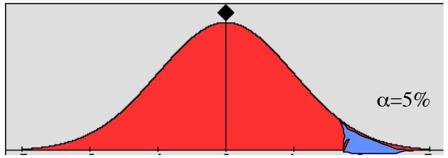
A “Visual” Look at Power

Distribution of Sample Means for Population when Ho is rejected



μ=200

Distribution of Sample Means for Population when Ho is True

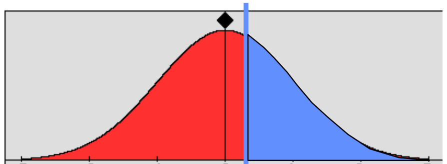


μ=100

(a one-tailed test was used)

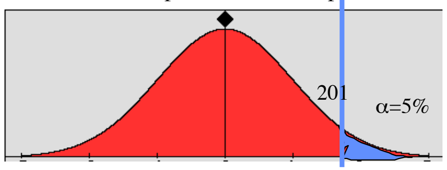
A “Visual” Look at Power

Distribution of Sample Means for Population when Ho is rejected



μ=200

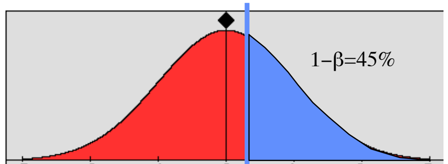
Distribution of Sample Means for Population when Ho is True



μ=100

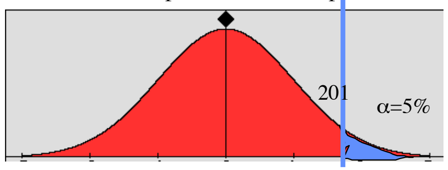
A “Visual” Look at Power

Distribution of Sample Means for Population when Ho is rejected



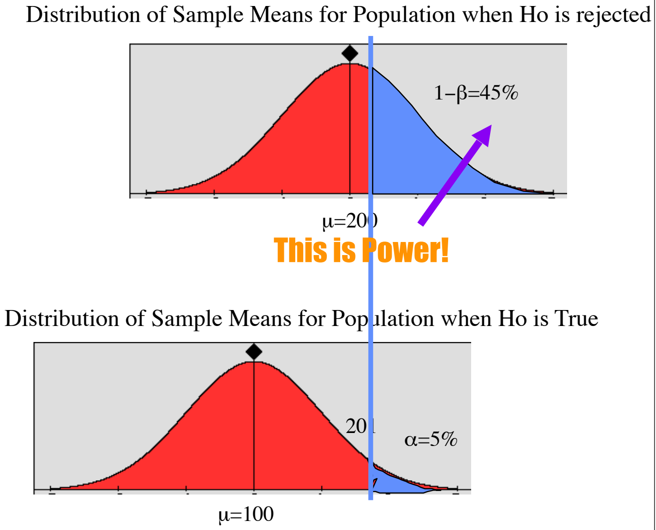
μ=200

Distribution of Sample Means for Population when Ho is True

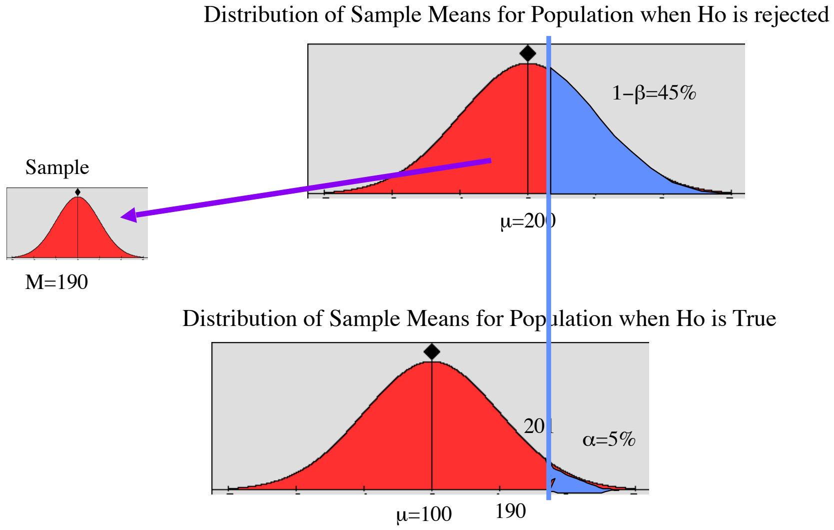


μ=100

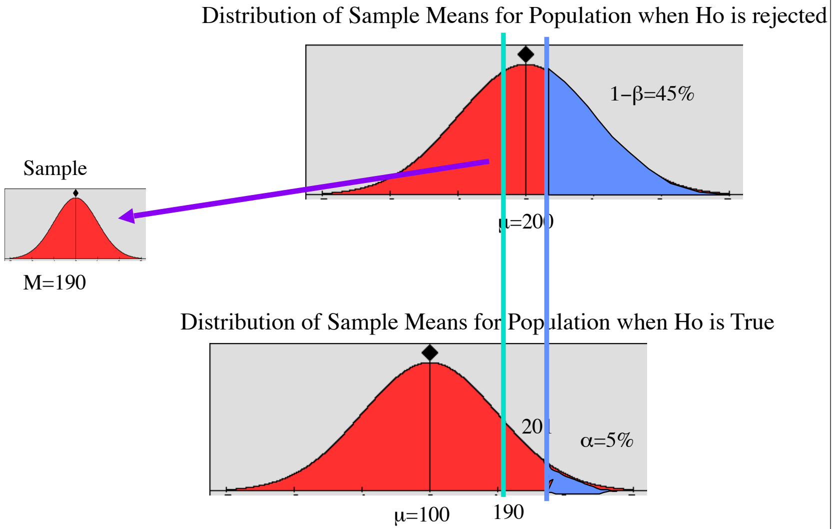
A “Visual” Look at Power



A “Visual” Look at Power



A “Visual” Look at Power

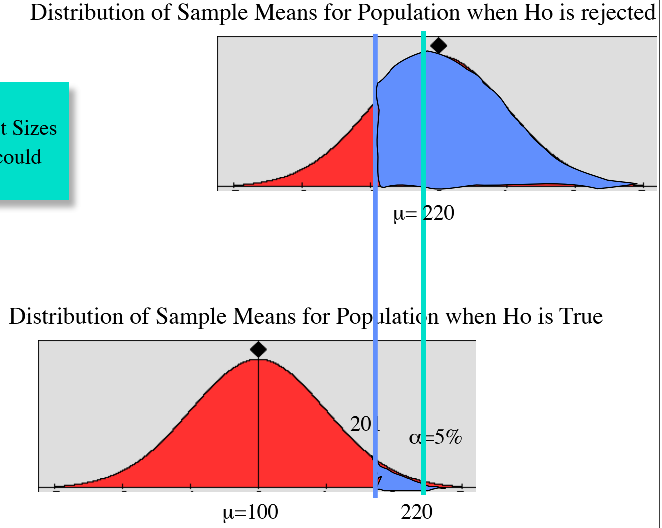


How do we maximize Power?

- Increase the effect size
- Increase the sample size
- Increase alpha
- Use one-tailed tests

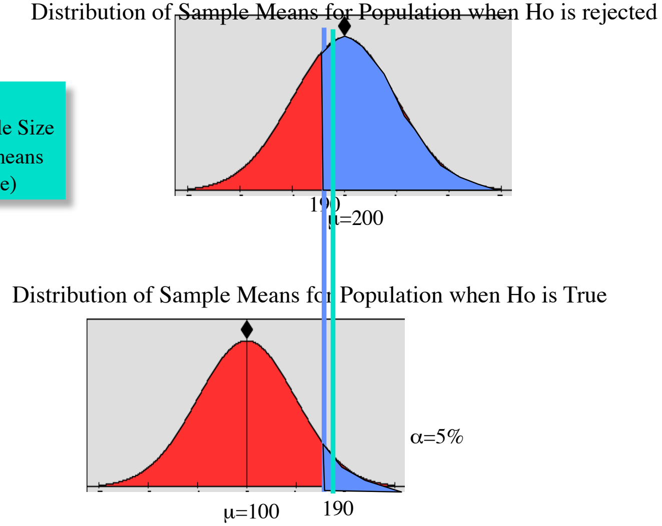
A “Visual” Look at Increasing Power
Way #1

Increase Power
By Getting big Effect Sizes
(e.g., our pop mean could
be much higher)



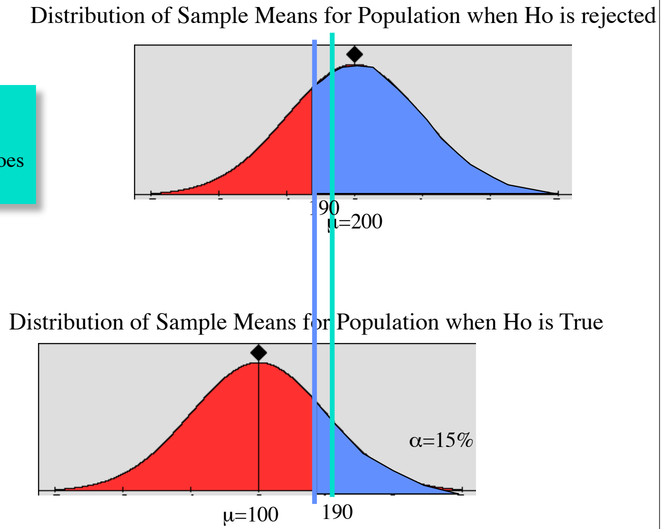
A “Visual” Look at Increasing Power
Way #2

Increase Power
By Increasing Sample Size
(the distribution of means
becomes less variable)



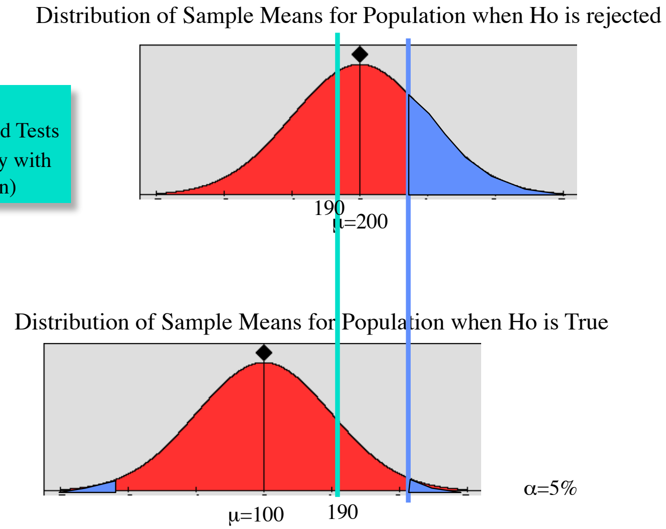
A “Visual” Look at Increasing Power
Way #3

Increase Power
By Increasing Alpha
(as α increases, so does
power)



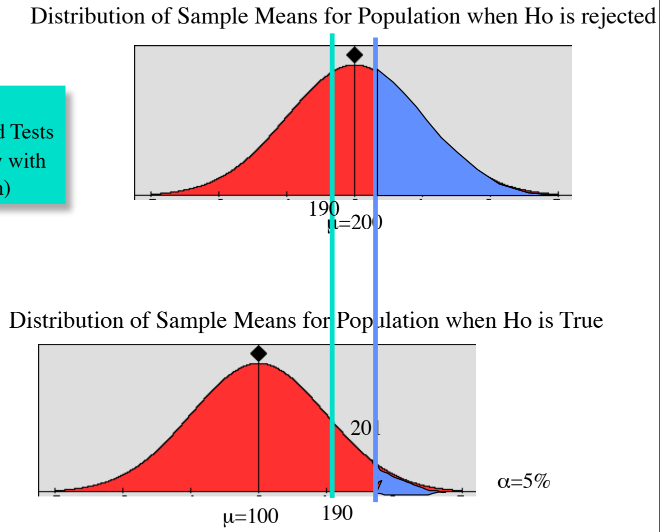
A “Visual” Look at Increasing Power
Way #4

Increase Power
By Using One-Tailed Tests
(i.e., we could get by with
a lower sample mean)



A "Visual" Look at Increasing Power
Way #4

Increase Power
By Using One-Tailed Tests
(i.e., we could get by with
a lower sample mean)



Calculating Power

- It is possible to calculate power.
- However, it's sometimes difficult to compute power by hand, so we usually rely on power tables.

How do we maximize Power?

- Increase the effect size
- Increase the sample size
- Increase alpha
- Use one-tailed tests

Choosing the right statistical test

Four major research questions which determine which statistical methods to use:

- Do groups differ?
- What is the degree of relationship among variables?
- What is the structural relationship among variables?
- Can we predict group membership?

First question: Do groups differ?

- *z*-test
- *t*-test
  - single sample
  - independent means
  - dependent means
- ANOVA
  - Oneway
  - Factorial
  - Repeated measures (within-subjects)
  - Mixed

First question: Do groups differ? (cont.)

- Analysis of covariance (ANCOVA)
  - Goal:estimate the effect of some "unwanted" variable (a *covariate*), and remove the variance associated with it from the analysis of the effects in which you are interested
- Chi-square test
- MANOVA and MANCOVA
  - multivariate analyses of variance and covariance
  - Use when there are multiple dependent variables that measure the same thing
  - The analysis creates a linear combination of the variables as a sort of "single DV" to test in the analysis

Second Question: What is the degree of relationship among variables?

- Pearson *r* (bivariate)
- Rank order tests: e.g., Spearman rho
- Point biserial *r*: when one DV is dichotomous and the other is continuous
- Partial *r*: the association between two variables considering the influence of other variables
  - Strategy is to "partial out" or "control for" the effects of another variable

Second Question: What is the degree of relationship among variables? (cont.)

- When multiple DVs are used to predict to another DV…
  - Multiple regression
  - To assess effects of different combinations of DVs separately (perhaps to assess a covariate's effect), we use sequential regression methods:
    - Hierarchical (a theory or hypothesis determines the order)
    - Stepwise (the computer program chooses the order based on which variables have the greatest effects)
    - The regression steps are evaluated by examining how much variance (R^2) on the outcome variable they predict

Third Question: What is the structural relationship among variables?

- **Factor Analysis**--used with a large number of variables

Imagine you ask 30 questions about how people deal with emotional experiences in their everyday life.

1. When I get angry, I feel like I'm going to explode.
2. When I'm sad, I cry.
3. I don't get emotional when I hear something really good or really bad.
4. I talk to other people when I'm upset.
5. People can't tell when I'm upset.

Third Question: What is the structural relationship among variables? (cont.)

- A factor analysis assesses how much different variables are correlated with another and not correlated with each other, considering the effects of all on each other simultaneously.
- Some items are found to clump together as "factors"
- This method was developed by people interested in assessing the nature of intelligence-- assuming that there could be many factors that underlie intelligence: spatial ability, verbal, etc.

Third Question: What is the structural relationship among variables? (cont.)

- Causal Modeling also involves testing individuals on many variables. However, beyond just assuming that variables are correlated, this type of analysis tries to determine whether a variable causes effects on another variable
- Remember the adage, "correlation does not imply causality?"
- Structural equation modeling is the most common type

Fourth Question: Can we predict group membership?

- Some choices: **Logit analysis**, **Discriminant functions**, and **Logistic Regression**
- Example:
  - Want to know whether or not someone might become a criminal (Yes or No)
  - Could look at possible predictors: family income, education, biological factors (e.g., resting heart rate), etc.
  - These predictors could be simultaneously combined and analyzed to predict the outcome (much like a multiple regression) in a logistic regression

Eric Vanman
School of Psychology
University of Queensland

twitter: @socialneuro
UQ Social Neuroscience Lab: www.socialneuro.com