MODULE 4: STUDY DESIGN
SAMPLE SIZE JUSTIFICATION

Bios6624: Advanced Statistical Methods and Analysis

Nichole E Carlson, PhD

SAMPLE SIZE JUSTIFICATION

- The most common reason to contact a statistician
- Sample size is contingent on design, analysis plan and outcome

- With the wrong sample size

– May not be able to make conclusions because the study is ‘underpowered’ i.e. Is result non- significant because treatment doesn’t work OR is sample size too small to detect difference

– Waste time and money because your study is larger than required to answer the question of interest –may find statistical significance for a difference that is not clinically relevant

REVIEW: STATISTICAL DECISION MAKING

		True Difference	
		Present (H_a)	Absent (H_0)
Conclusion of Statistical Test	Significant	Correct (Power)	Type I (α) error
	Not Significant	Type II (β) error	Correct

POWER IS DEPENDENT ON

- Study design
- Distribution of data
- Type-I error rate (a)
- Two-sided test (usual) or One-sided Test
- Size of difference to be detected
- Variability of data
- Sample size

LET'S LOOK AT POWER FOR ONE SAMPLE Z-TEST

- You have continuous data as your outcome.
- You are interested in testing of the mean of the data differs from zero (Null hypothesis).
- You decided a priori the Z-test is the correct statistical analysis (you have a larger sample size).
- We want to know what the power of this test is for different alternative hypotheses.

- Recall, power = p(reject the null hypothesis| under the alternative)

- So, power is the probability you reject the null when you should

POWER FOR ONE SAMPLE T-TEST

$$P\left(|Z^*| > Z_{1-\alpha/2} \middle| \mu = \mu_{alt}, \sigma_{\bar{X}}^2\right)$$

$$Z^* = \frac{(\bar{X} - 0)}{\sqrt{\sigma_{\bar{X}}^2/n}}$$

Under the alternative: $\bar{X} \sim N(\mu_{alt}, \sigma_{\bar{X}}^2\ /n)$

$$P\left(\bar{X} > \sqrt{\sigma_{\bar{X}}^2/n} * Z_{1-\alpha/2} \middle| \mu = \mu_{alt}, \sigma_{\bar{X}}^2\right) + P\left(\bar{X} < -\sqrt{\sigma_{\bar{X}}^2/n} * Z_{1-\alpha/2} \middle| \mu = \mu_{alt}, \sigma_{\bar{X}}^2\right)$$

$$P\left(\bar{X} - \mu_{alt} > \sqrt{\frac{\sigma_{\bar{X}}^2}{n}} * Z_{1-\alpha/2} - \mu_{alt} \middle| \mu = \mu_{alt}, \sigma_{\bar{X}}^2\right) + P\left(\bar{X} - \mu_{alt} < -\sqrt{\frac{\sigma_{\bar{X}}^2}{n}} * Z_{1-\alpha/2} - \mu_{alt} \middle| \mu = \mu_{alt}, \sigma_{\bar{X}}^2\right)$$

$$P\left(\frac{\bar{X} - \mu_{alt}}{\sqrt{\frac{\sigma_{\bar{X}}^2}{n}}} > Z_{1-\alpha/2} - \frac{\mu_{alt}}{\sqrt{\frac{\sigma_{\bar{X}}^2}{n}}} \middle| \mu = \mu_{alt}, \sigma_{\bar{X}}^2\right) + P\left(\frac{\bar{X} - \mu_{alt}}{\sqrt{\frac{\sigma_{\bar{X}}^2}{n}}} > -Z_{1-\alpha/2} - \frac{\mu_{alt}}{\sqrt{\frac{\sigma_{\bar{X}}^2}{n}}} \middle| \mu = \mu_{alt}, \sigma_{\bar{X}}^2\right)$$

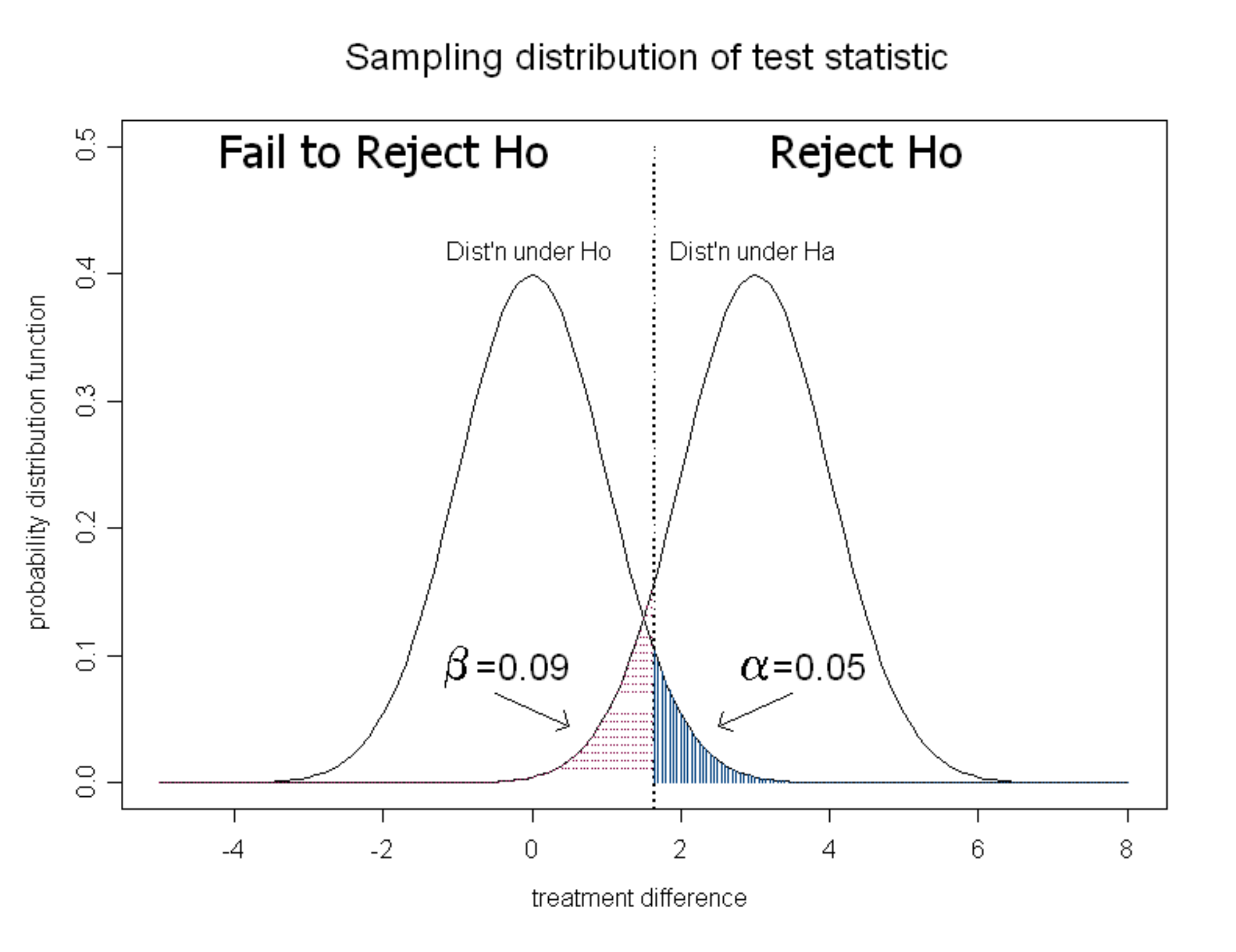
Which are normal probabilities we can compute.

PICTURE OF POWER FOR SAMPLING DISTRIBUTIONS UNDER H_0 AND H_A

ASSUMES NORMAL DISTRIBUTION OR LARGE SAMPLE, ONE TAIL TEST OF MEAN

H0: N(0,1)
HA: N(3,1)

Beta=1-power
= Type II error



Power: Pr(reject null hypothesis | alternative hypothesis is true)

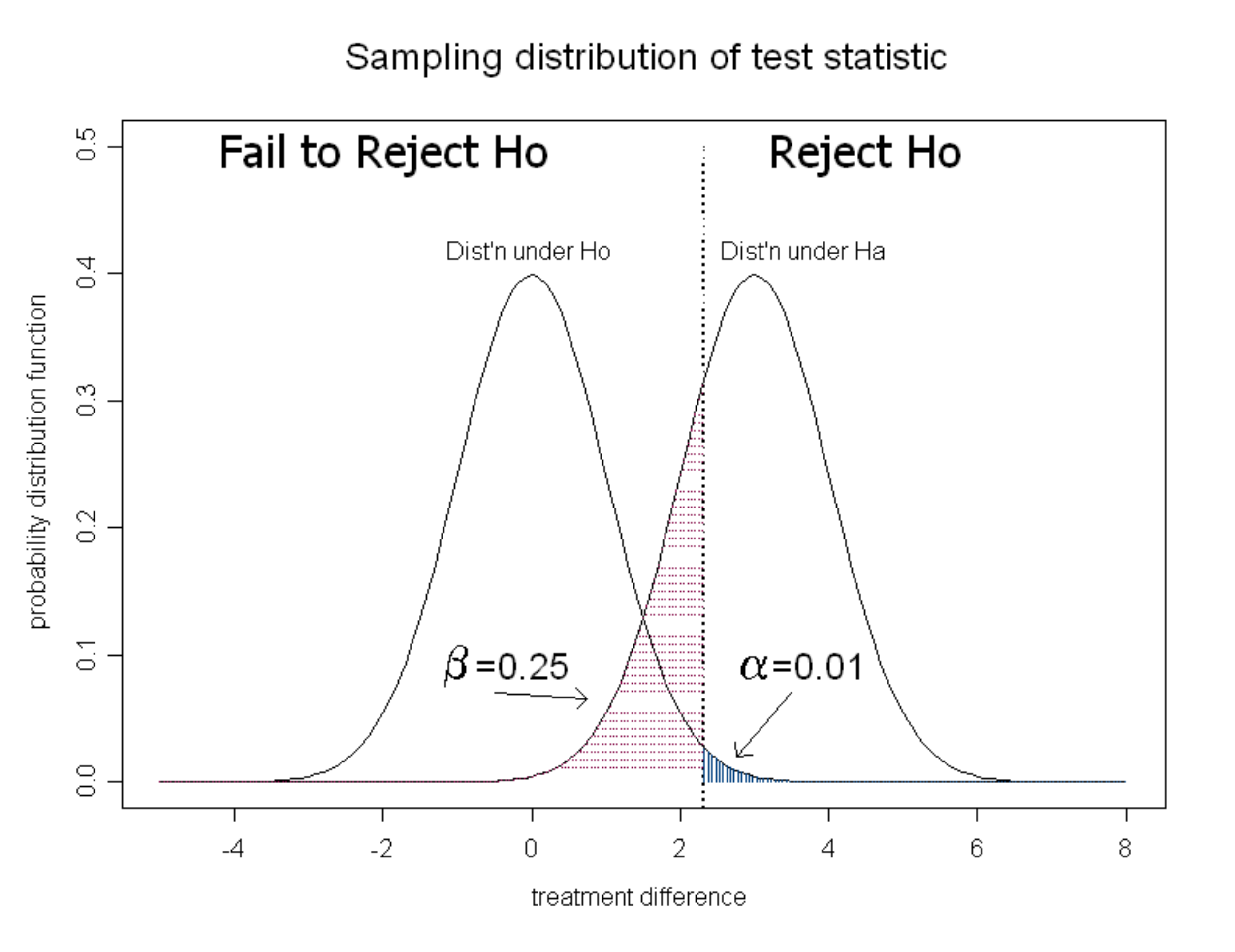
POWER INCREASES AS:

- Type-I error rate increases
  - Often fix a at 0.05 (other levels may be appropriate)

- Difference between H_0 and H_A increases (represented by a mean farther from zero in our example)

- Standard error decreases
  - Increase sample size
  - Decrease between subject variability

TRADE-OFF BETWEEN α AND β



POWER INCREASES AS:

- Type-I error rate increases

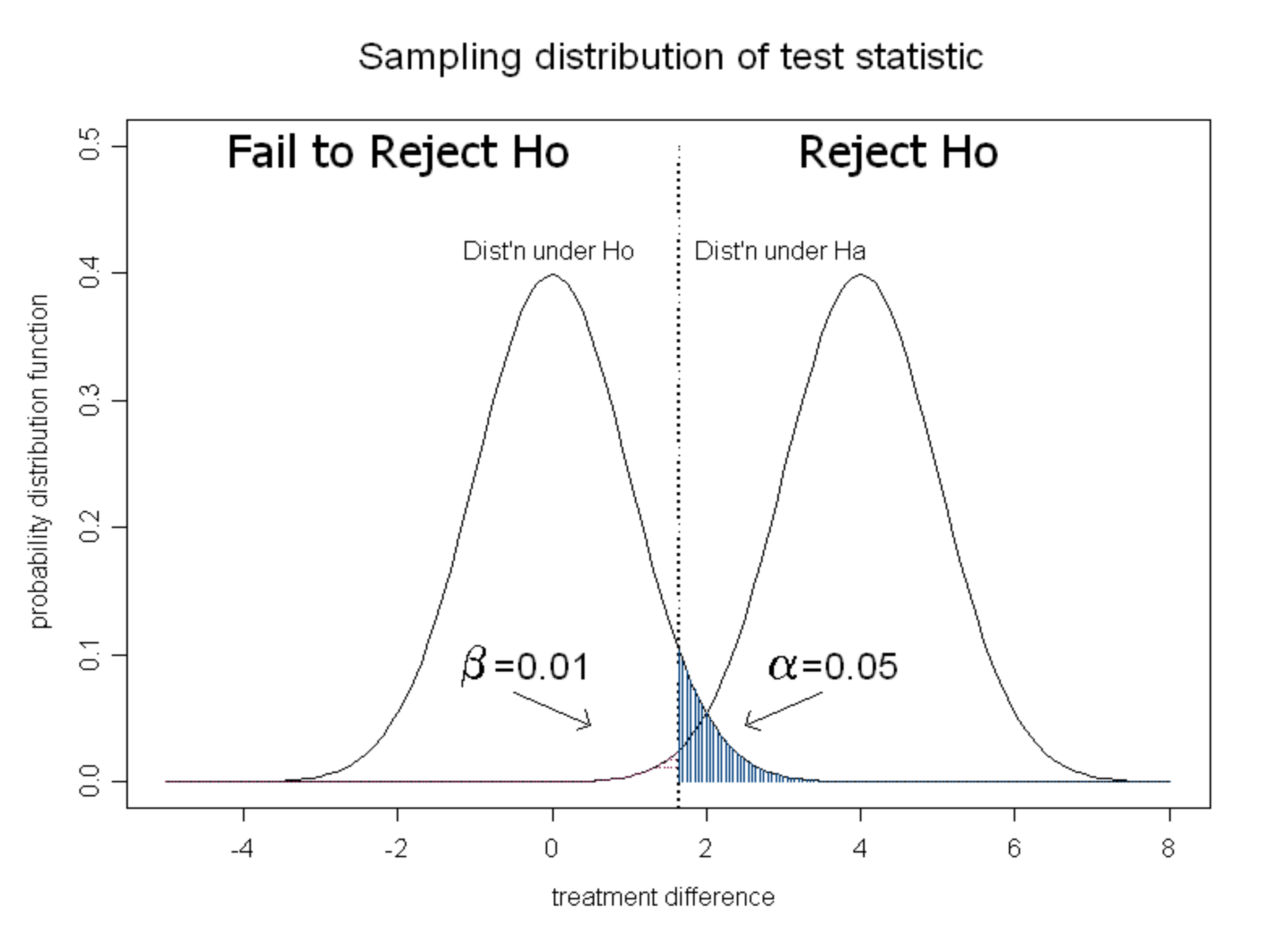
- Difference between H_0 and H_A increases (represented by a mean farther from zero)

- Standard error decreases
  - Increase sample size
  - Decrease between subject variability
  - Decrease variability of assay (within subject)

POWER INCREASES AS DIFFERENCE BETWEEN H_0 AND H_A INCREASES

H0: mu=0
HA: mu=4 here

Much smaller red region.



POWER INCREASES AS:

- Type-I error rate increases
  - Usually fix a at 0.05

- Difference between H_0 and H_A increases

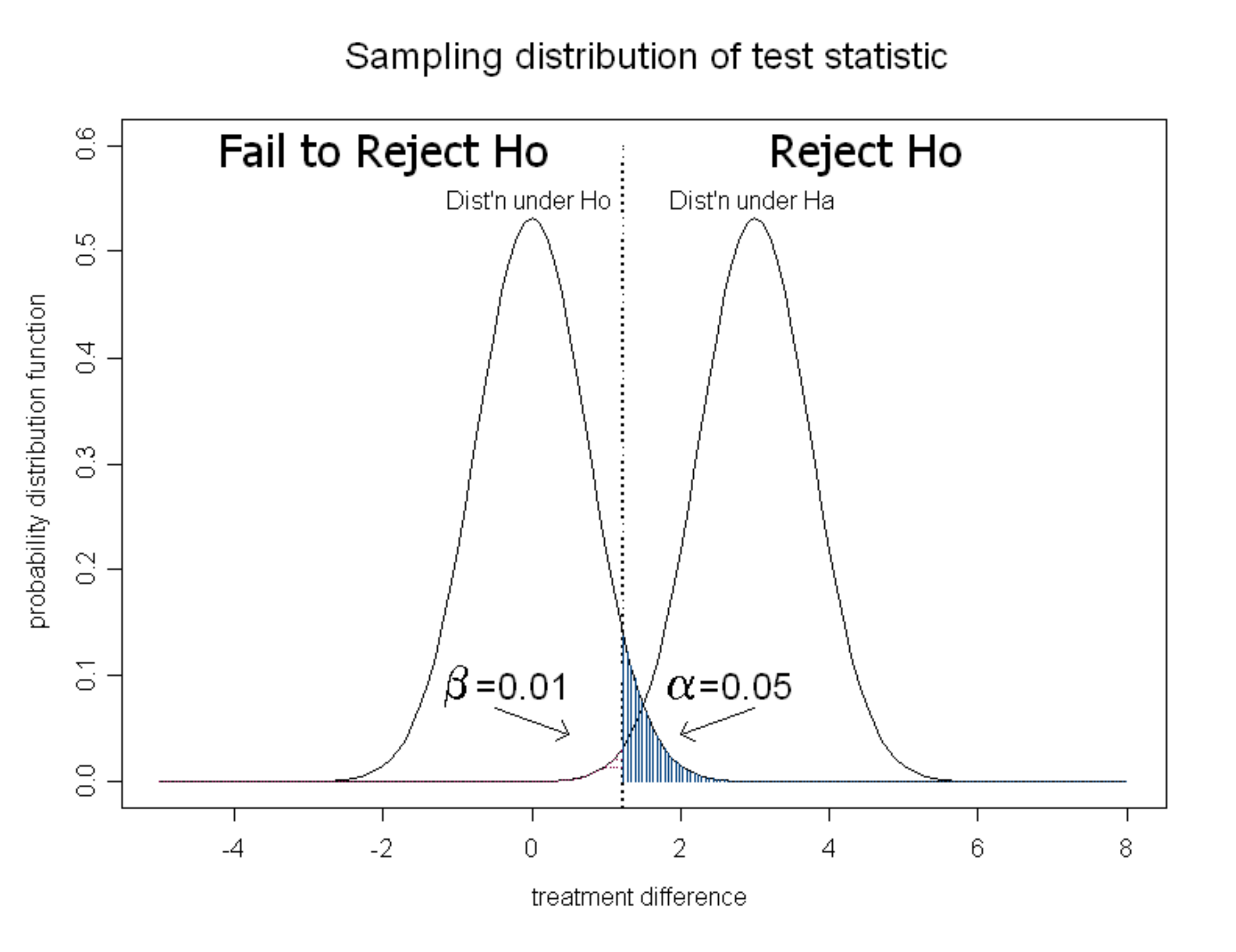
- Standard error decreases
  - Increase sample size
  - Decrease between subject variability
  - Decrease variability of assay (within subject)

POWER INCREASES AS STANDARD ERROR DECREASES

H0: mu=0,
SE=0.75

HA: mu=3,
SE=0.75

Notice the red region in much smaller.



GENERALIZATIONS

- Can generalize these calculations to:
- Two sample t-test.
- Paired t-test (useful for pre-post studies)
- Difference of differences (useful for pre-post studies on two groups)
- Difference of proportions (useful for 0/1 variables on two groups)
- Regression coefficients (these are t-tests)
- Correlations (this is a z-test)

- Many other things available in software.

Sample size ALWAYS requires the investigator to make some assumptions

– What difference do you expect between the groups?
– How much variability do we expect in the measurements?
– What is a clinically relevant difference

- Statistician cannot give you these estimates (unless you provide preliminary data)

- It is the responsibility of clinical investigator to define these parameters (and the statistician to ask for them)

POWER ANALYSIS (SAMPLE SIZE/SIZE OF DIFFERENCE/POWER)

- First determine:
  - Hypothesis, study design, and analysis plan
  - Obtain from the investigator, variability in the outcome measure (usually a standard deviation)
  - Obtain from the investigator, clinically meaningful difference.
  - Acceptable Type I error
- Then solve for one of the following (after setting the values of the other two):
  - Sample Size
  - Difference to be detected
  - Power
- Write-up in grant should clearly describe ALL of the above

POWER

The power analysis section should include treatment of (at least) the following topics:

- Assumptions you made to estimate power
  - Simplifying assumptions (e.g. most sophisticated model will control for covariates, but you use a t-test to estimate power for primary test)
  - Details of the parameters you assumed (e.g. estimate of effect, alpha-level, sample size)
  - Where you got estimates of effect and variability in measures
- Power calculations
  - Estimate of power to detect effect or range of effect sizes for which you have power at X% based on your sample size
  - Cite what program you used (if you used one) or how you estimated power
  - Summary of power calculations if you have several

A FEW THINGS TO REMEMBER

- It is just an estimated sample size or power.
- Want to balance statistical and practical considerations and justify your choices.
- Choose sample size and power so that statistical significance will coincide with clinical significance.
- The sample size is calculated for the number of subjects who finish the study, not the number that start the study.
- The power section of your grant should be consistent with the analysis section of the grant.
  - Often make simplifying assumptions to estimate power.
  - Those assumptions and their presumed impact on power/sample size estimates should be very clearly described.

THANKS TO THESE INDIVIDUALS FOR SOME SLIDES AND/OR INFORMATION FOR THE POWER SECTION

- Tasha Fingerlin, PhD
- Sam MaWhinney, ScD
- NJC Short course notes (1998)
  - Becki Bucher Bartelson, PhD
  - Lynn Ackerson, PhD