Powering Up
Your Lean Six Sigma Projects

MoreSteam.com®

Smita Skrivanek
March 26, 2015

MBB Webcast Sponsor: MoreSteam.com

- Founded in 2000
- Trained 470,000+ Lean Six Sigma professionals
- Served over 2,000 corporate customers (including 50+% of the F500)
- First firm to offer the complete Black Belt curriculum online
- Courses reviewed and approved by ASQ and PMI
- Academic Partnerships with Ohio State University, Notre Dame, Cal Poly and George Washington University

Select Customers:

Today's Program

- Welcome
- Introduction of MBB Webcast Series
  - Ellen Milnes, MoreSteam.com
- Speaker:
  - Smita Skrivanek, MoreSteam.com
- Open Discussion and Questions

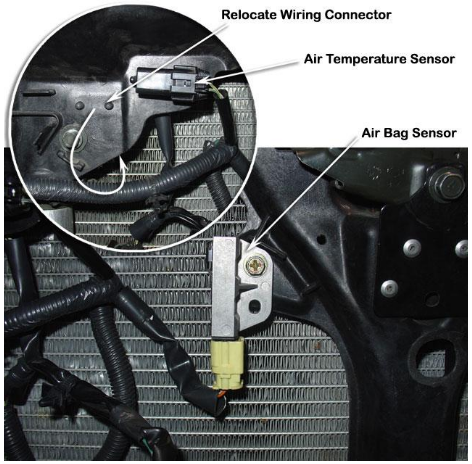
About Our Presenter

Smita Skrivanek

EngineRoom Product Manager, MoreSteam.com

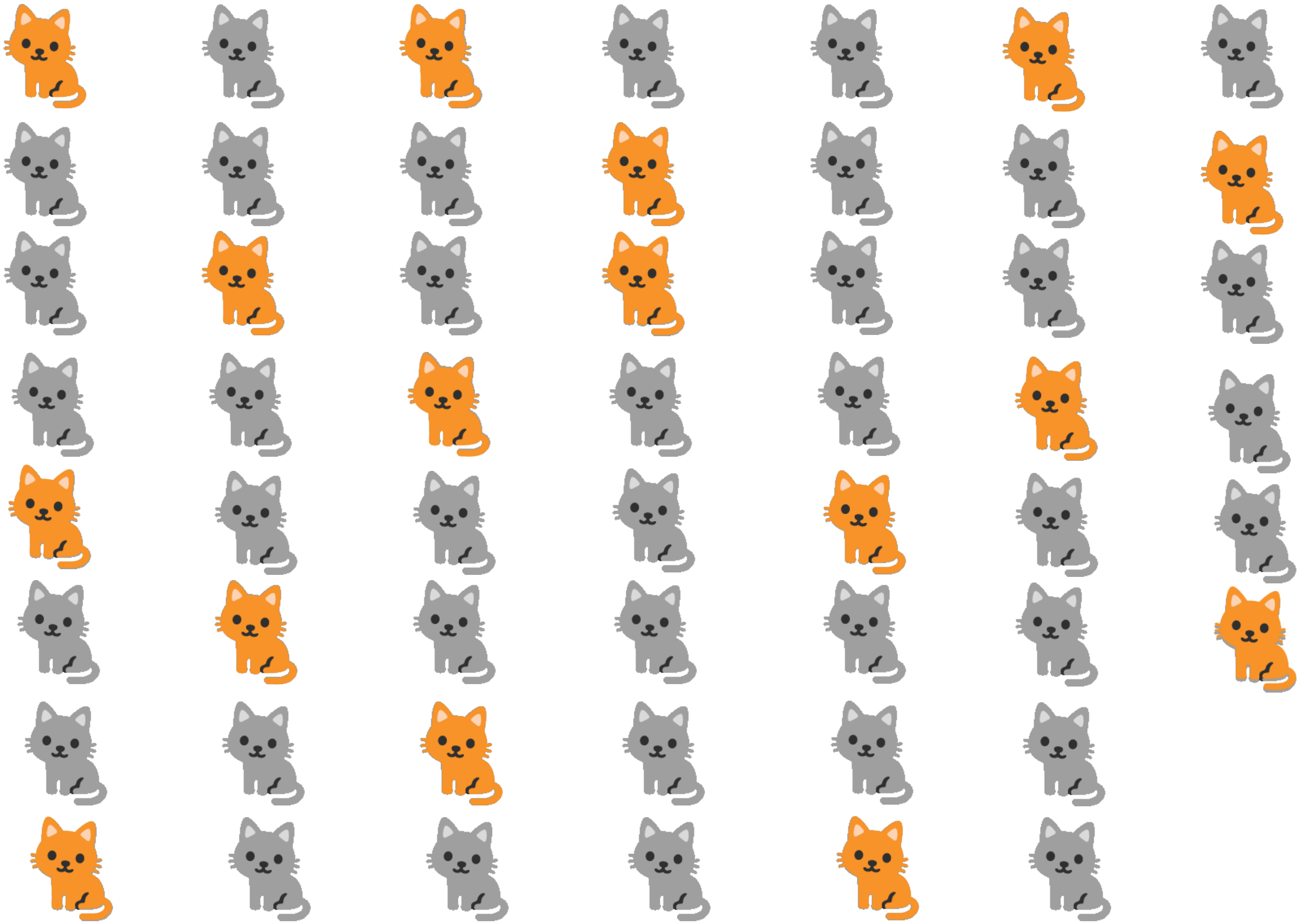
- *Heads research & development for EngineRoom® software*
- *Develops content & software functions, reviews projects, and assists students with questions on advanced statistics*
- *Masters in Applied Statistics from The Ohio State University and an MBA from Indiana University Kelley School of Business*

An Example

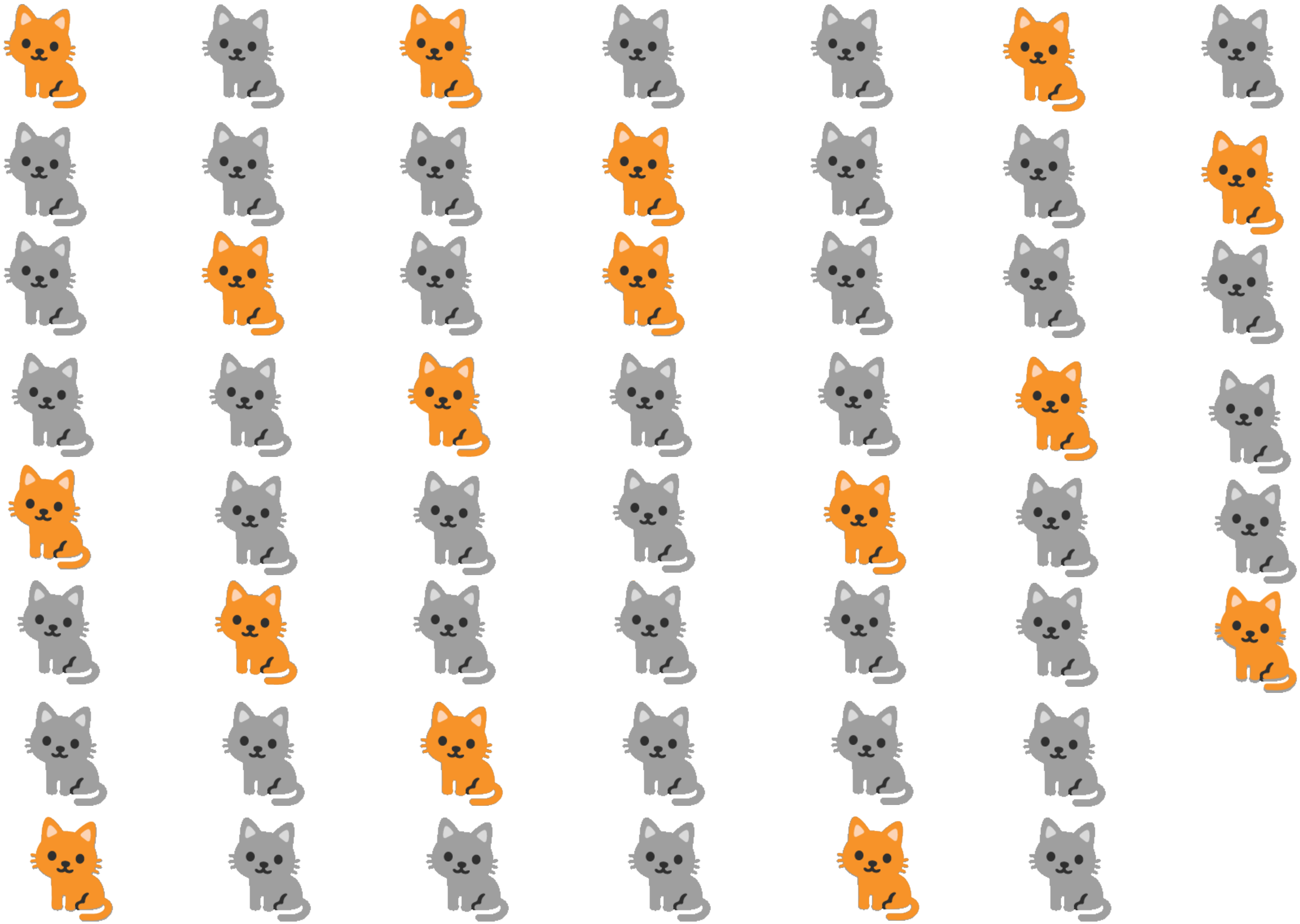


- Air bag sensors
    - ‘Zero defects’ initiative
- Produce parts at less than 300 defects per million.
    - H_0: defect rate = 0.0003
    - H_1: defect rate < 0.0003
- Power corresponding to sample size = 100?
- Sample size to achieve 80% power?

What sample size to detect the effect?

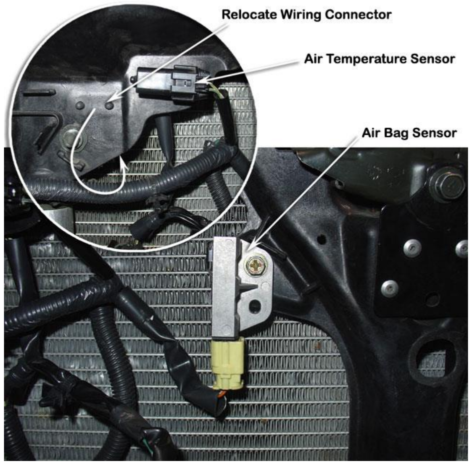


What sample size to detect the effect?



An Example

- Air bag sensors
  - ‘Zero defects’ initiative
- Produce parts at less than 300 defects per million.
  - H_0: defect rate = 0.0003
  - H_1: defect rate < 0.0003
- Power corresponding to sample size = 100?
  - For effect size 0.0002: power = 0.025 (disc.) 0.063 (cont.)
  - For effect size 0.0001: power = 0.004 (disc.) 0.056 (cont.)
- Sample size to achieve 80% power?
  - For effect size 0.0002: n ≈ 46,356
  - For effect size 0.0001: n ≈ 185,422



What's in store…

- What is power (for a study)?
- Why is it important?
- What factors affect power?
- What are the consequences of having too little or too much power?
- Some real world examples

What is statistical power?

The probability that a test will detect the effect size of interest, if that effect exists.

Types of Errors

Effect Exists?		Test Says: No	Test Says: Yes
Reality	No	√ (Specificity)	X Type I error
	Yes	X Type II error	√ (Sensitivity)

P{Type I error} = P{Test = Yes | Reality = No}
= α

P{Type II error} = P{Test = No | Reality = Yes}
= β

Power = P{Test = Yes | Reality = Yes}
= 1 - P{Test = No | Reality = Yes}
= 1 – β

Types of Errors

Effect Exists?		Test Says: No	Test Says: Yes
Reality	No	√ (Specificity)	Type I error
	Yes	Type II error	√ (Sensitivity)

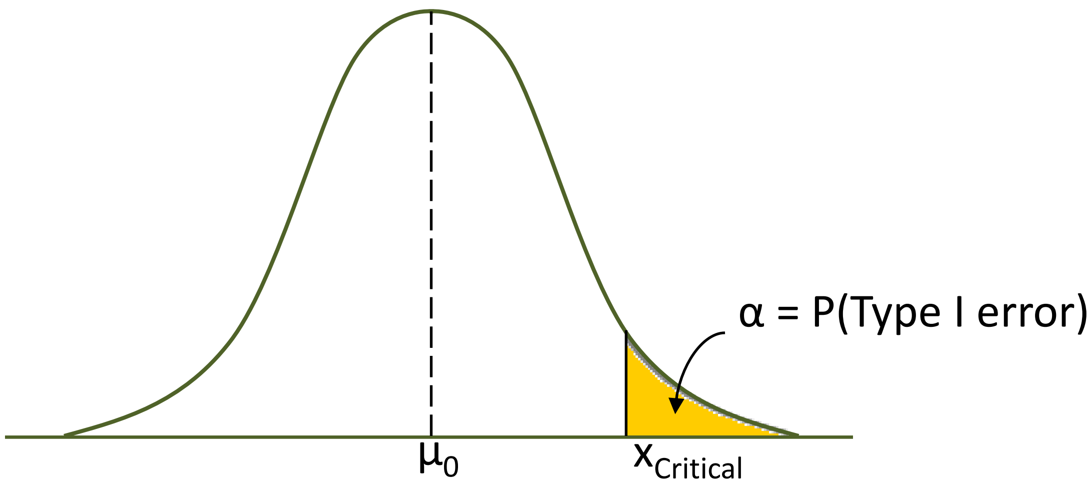
The goal:

Minimize the two types of error

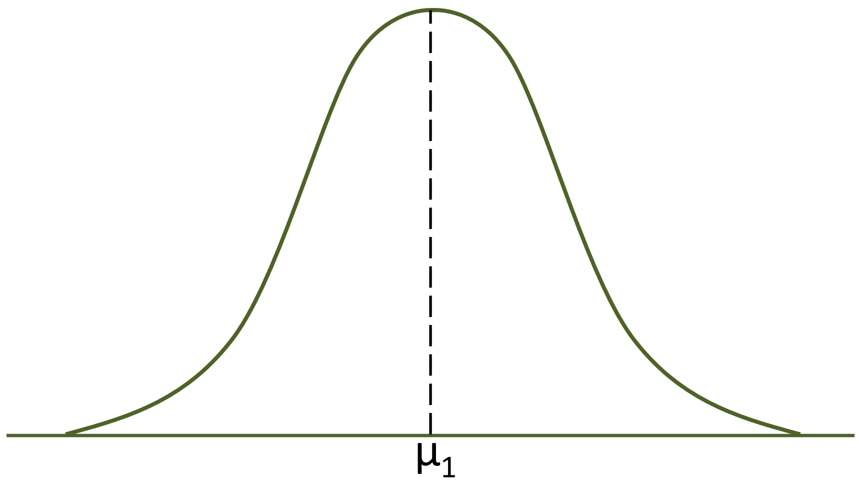
Maximize power

Power of a test – an illustration

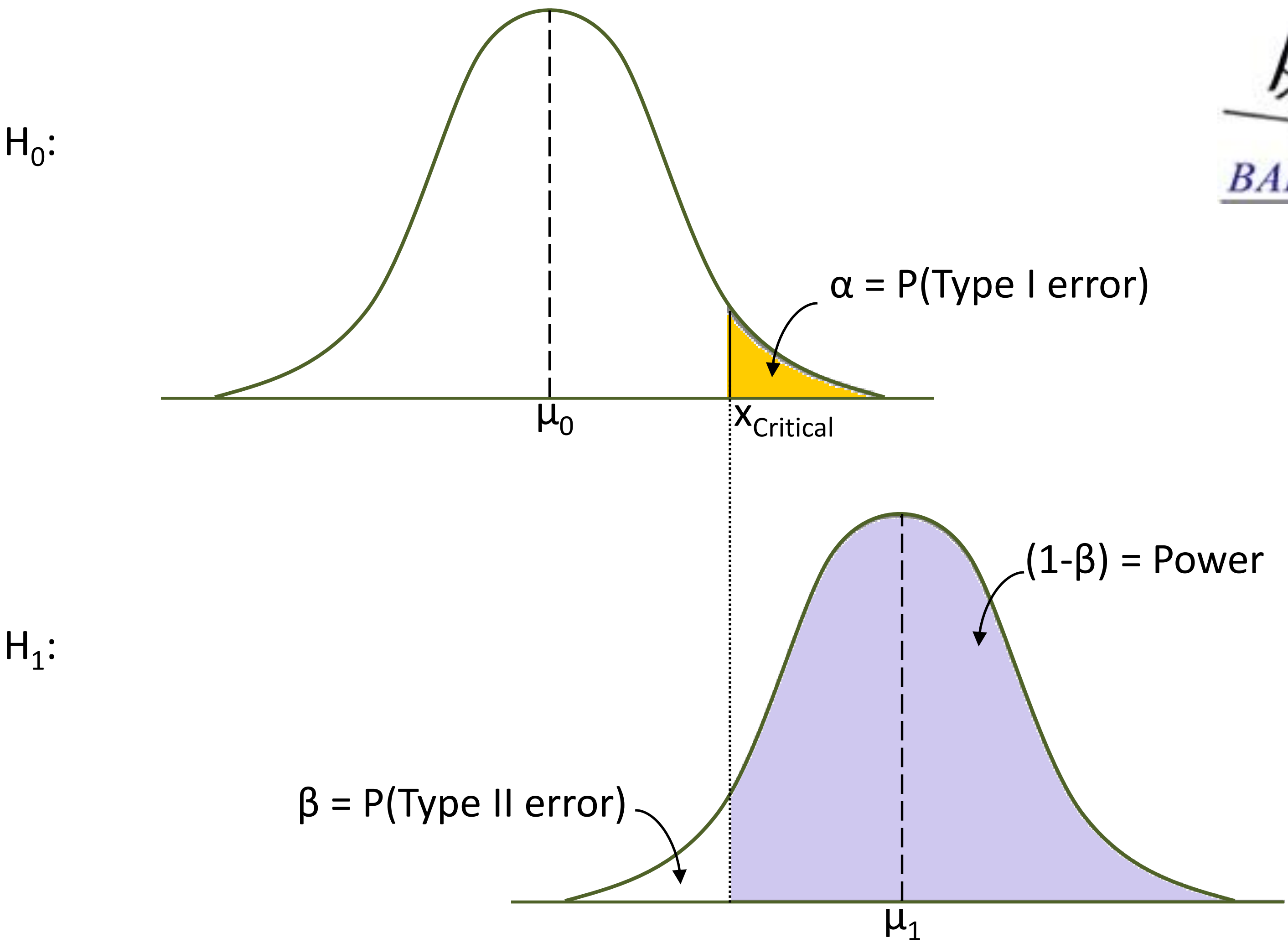
H_0:



H_1:



Power of a test – an illustration



Power Considerations

- Statistical method/test
  - Continuous vs. Discrete characteristic
    - » Continuous is better when possible
  - Parametric vs. Non-parametric tests
    - » Parametric is better <u>as long as assumptions are valid</u>
  - Type of Design
    - » One Factor At a Time (OFAT) vs Factorial study

Type of Design: OFAT vs. Factorial Design

Consider a simple experiment: baking cookies

Two factors: Temperature (350, 450), Time (30, 50)

OFAT

Varied	Run	Factors		Best Result
		Temp	Time	
Temp	1	350	50	Temp = 350
	2	450	50	
Time	3	450	30	Time = 30
	4	450	50	

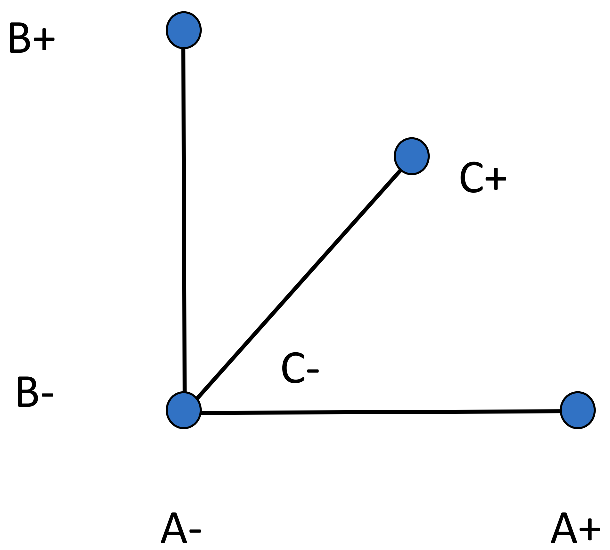
Note: Temp 350 was never run at Time 30!

Factorial Experiment

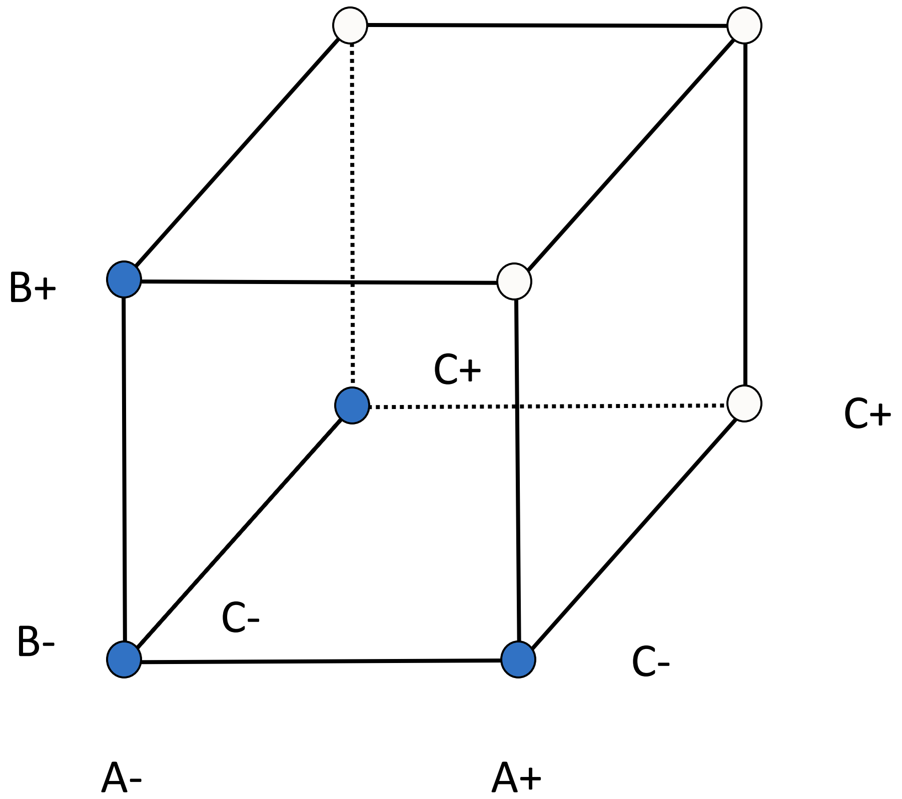
Varied	Run	Factors		Best Result
		Temp	Time	
Temp	1	350	30	Temp = 450 at Time = 30
	2	450	30	
Time	3	350	50	Temp = 350 at Time = 50
	4	450	50	

3-Factor Design

OFAT Testing

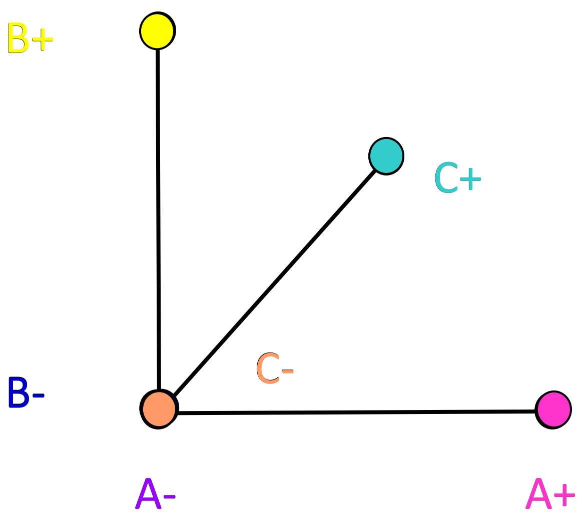


Factorial Testing

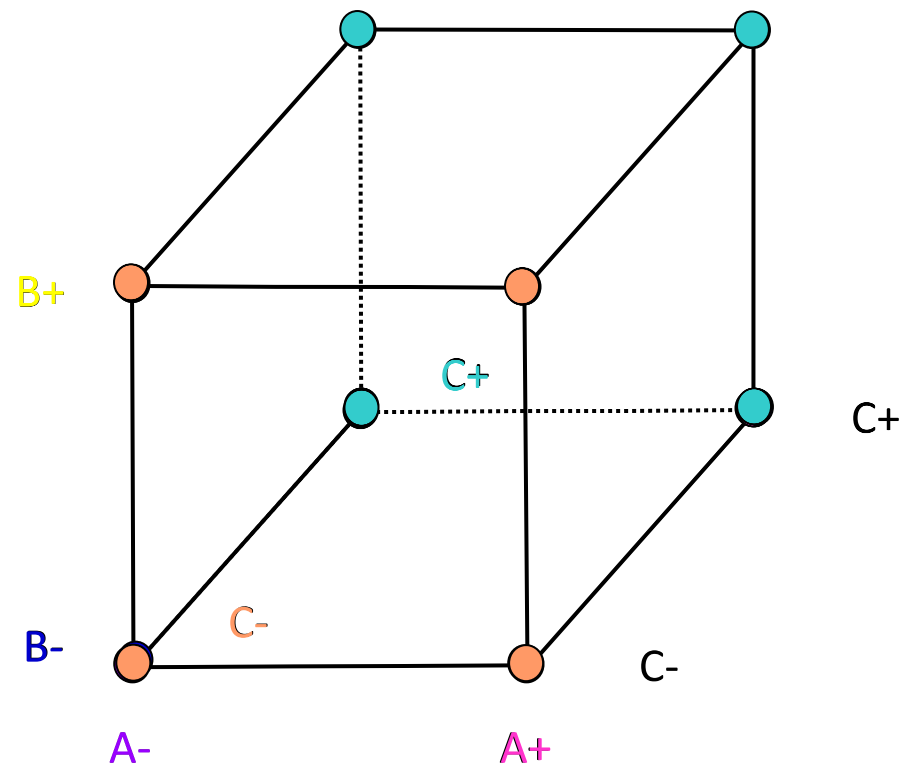


3-Factor Design

OFAT Testing



Factorial Testing



Relative Efficiency (FT/OFAT) = (1/8)/(1/16) = 2

The lowdown…

<u>OFAT</u>

- Sequential experimentation
- Only main effects can be estimated
- Large component of guesswork and luck involved!

<u>Factorial Experiments</u>

- All factors varied together
- Interaction effects can be detected and estimated
- More precise estimates
- More efficient design

Power Considerations

- Statistical method/test
  - Continuous vs. Discrete characteristic
    - » Continuous is better when possible
  - Parametric vs. Non-parametric tests
    - » Parametric is better as long as assumptions are valid
  - Type of Design
    - » OFAT vs Factorial study
    - » Replication (improves precision)

Replication

- **Objective**: Bond strength is measured indirectly by the contact angle. The goal is to maximize the Contact Angle (Y).

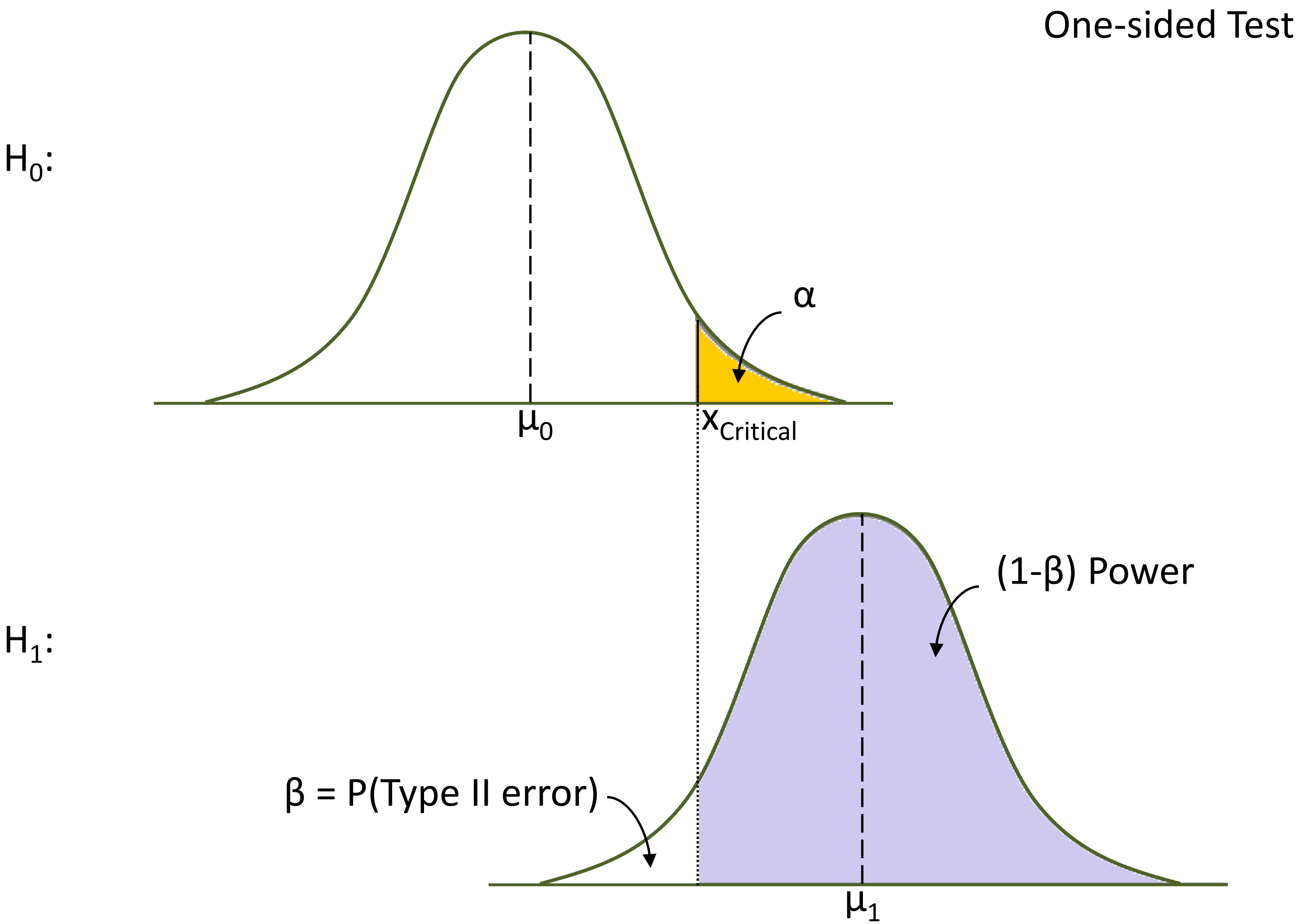
Std	Run	Location	P Pr	Flow Rate	RF Pr	Exp time	Contact angle
3	5	0	300	20	300	60	11.3
1	6	0	150	20	590	60	10.3
7	7	0	300	120	300	5	45.9
5	9	0	150	120	590	5	29.1
9	3	0.5	225	70	445	32.5	14.8
2	1	1	150	20	300	5	48.7
4	2	1	300	20	590	5	41.4
8	4	1	300	120	590	60	13.5
6	8	1	150	120	300	60	11.3

- 2^{5-2} factorial = res III design (main effects aliased with 2-factor interactions!), No replication, Single center point
- Power = 62% to detect an effect of 2 standard deviations.
- Curvature would have to be 3.5 sigma to be detectable.

Power Considerations

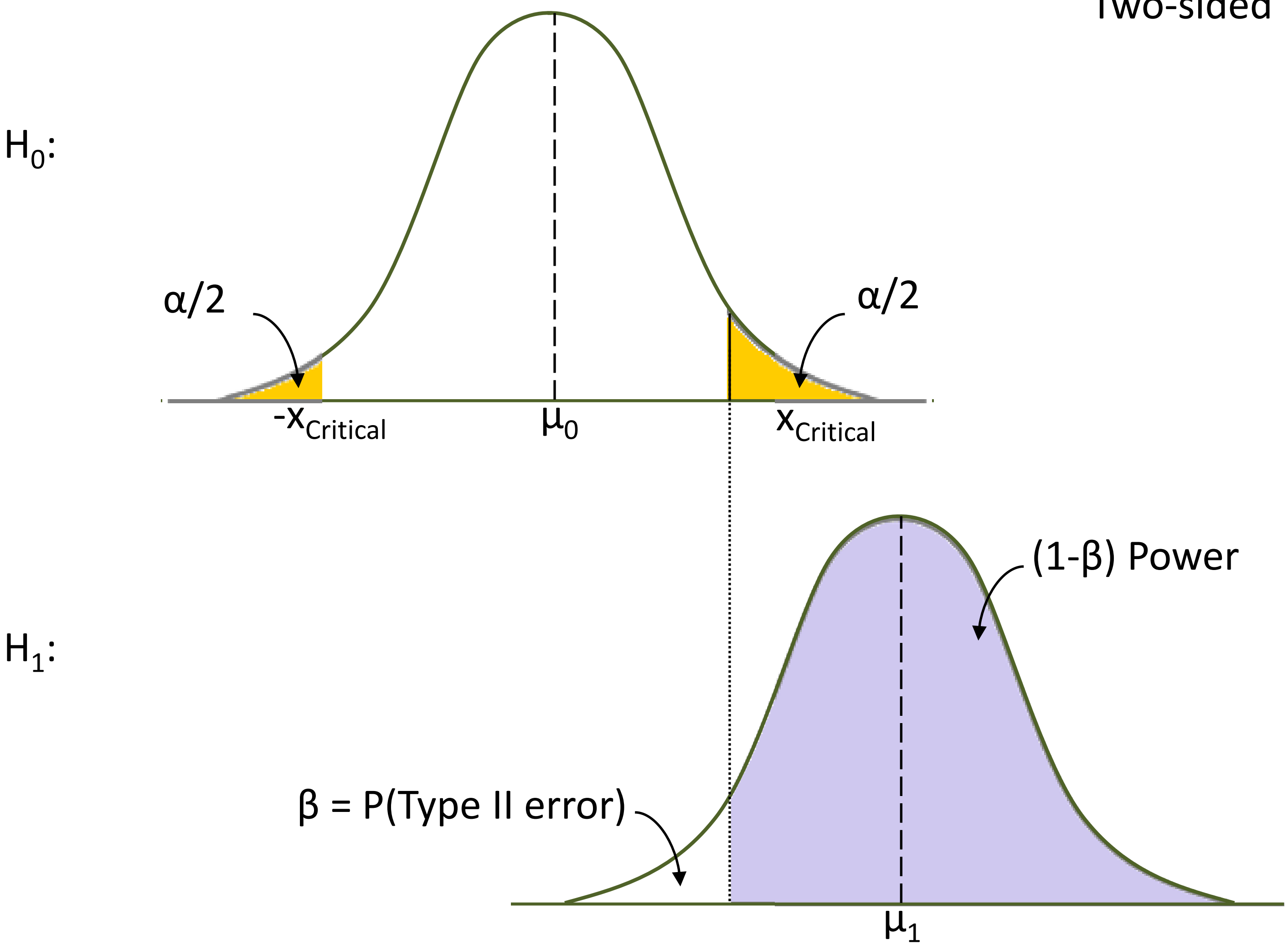
- Statistical method/test
  - Continuous vs. Discrete characteristic
    - » Continuous is better when possible
  - Parametric vs. Non-parametric tests
    - » Parametric is better as long as assumptions are valid
  - Type of Design
    - » OFAT vs Factorial study
    - » Replication (improves precision)
    - » Covariates, Blocks (reduces error variation)
    - » Repeated measurements (more precise estimates)
- Hypothesis to be tested
  - Directionality: Two-sided vs. One-sided

Directionality



Directionality

Two-sided Test



Power Considerations

- Statistical method/test
  - Continuous vs. Discrete characteristic
    - » Continuous is better when possible
  - Parametric vs. Non-parametric tests
    - » Parametric is better as long as assumptions are valid
  - Type of Design
    - » OFAT vs Factorial study
    - » Replication (improves precision)
    - » Covariates, Blocks (reduces error variation)
    - » Repeated measurements (more precise estimates)
- Hypothesis to be tested
  - Directionality: Two-sided vs. One-sided
    - » Use One-sided test if justifiable

Power Parameters

Factors that affect power:

1. Effect Size
2. Specified alpha (α) level
3. Sample size
4. Population variance/s (known/estimated)

Power Parameters

1. Effect Size: The difference in the response that is of practical/scientific importance

2. Specified alpha (α) level

Power Parameters

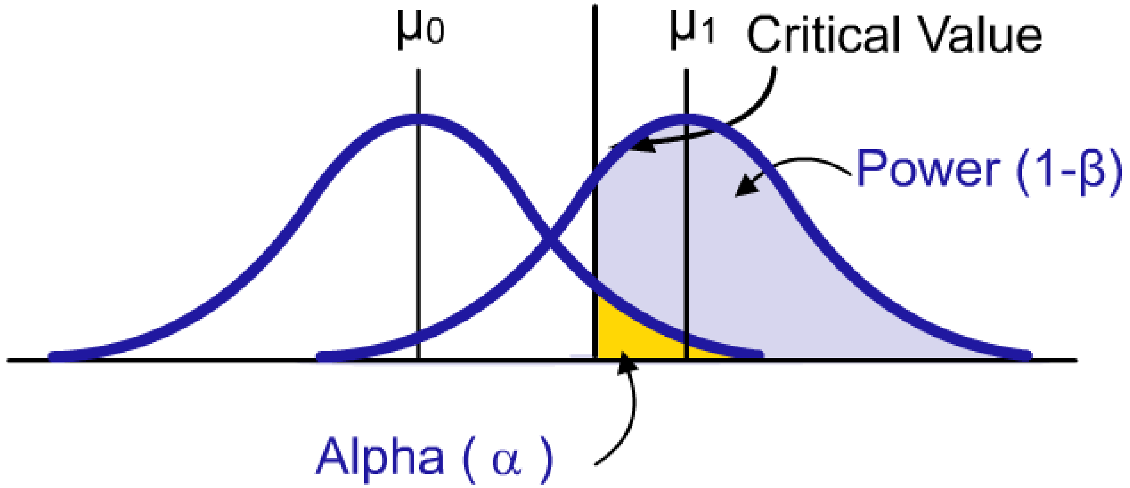
3. Desired sample size

Power Parameters

4. Known/Estimated population variance/s

Underpowered Studies

- May miss a real effect – wasted resources
- Contradictory results from repetitions
- ‘No evidence of effect’ misinterpreted as ‘Evidence of no effect’
- Variance of effect estimates (and therefore, any ‘found’ effects) are inflated



Underpowered Studies

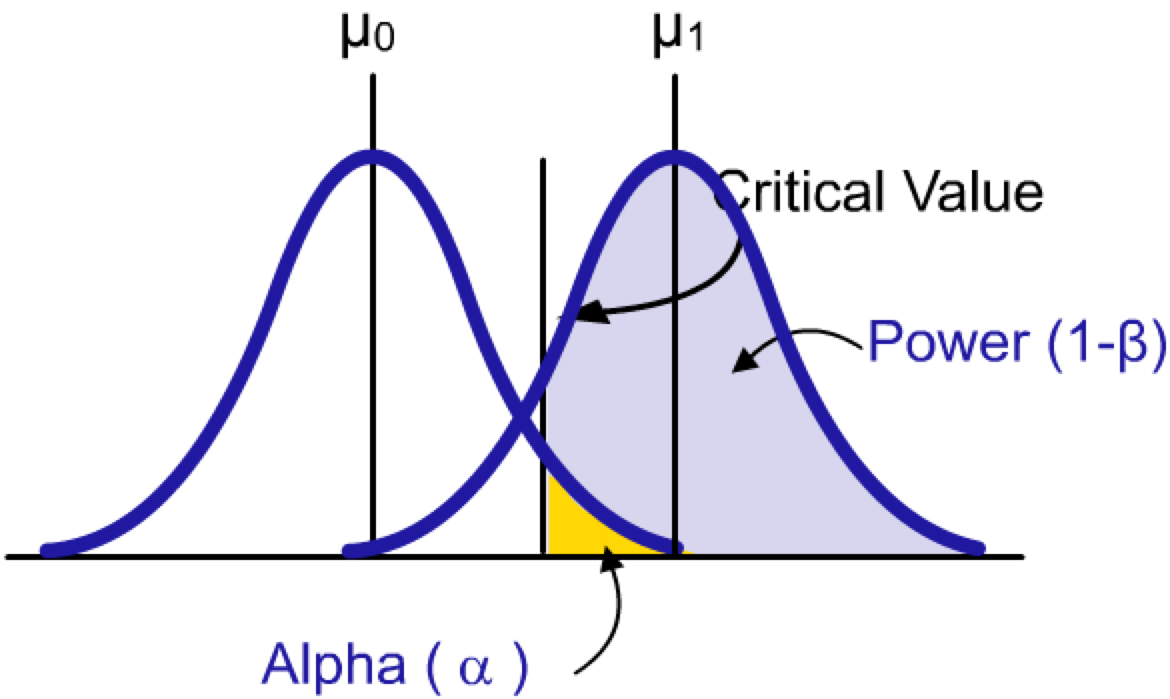
- May miss a real effect – wasted resources
- Contradictory results from repetitions
- ‘No evidence of effect’ misinterpreted as ‘Evidence of no effect’
- Variance of effect estimates (and therefore, any ‘found’ effects) are inflated

Another Example

- Glass manufacturer
    - replace raw material believed to cause acid rain
    - Coefficient of Thermal Expansion (CTE)
    - Want to see No difference in CTE between the two materials
- Sample size = 100 found a significant effect of CTE = 0.0019 (alpha = 0.01)
- Conclusion: Difference is statistically significant, but not of practical importance.
- Larger conclusion: a lot of effort and expense could have been avoided by doing a power analysis before conducting the study!

Overpowered Studies

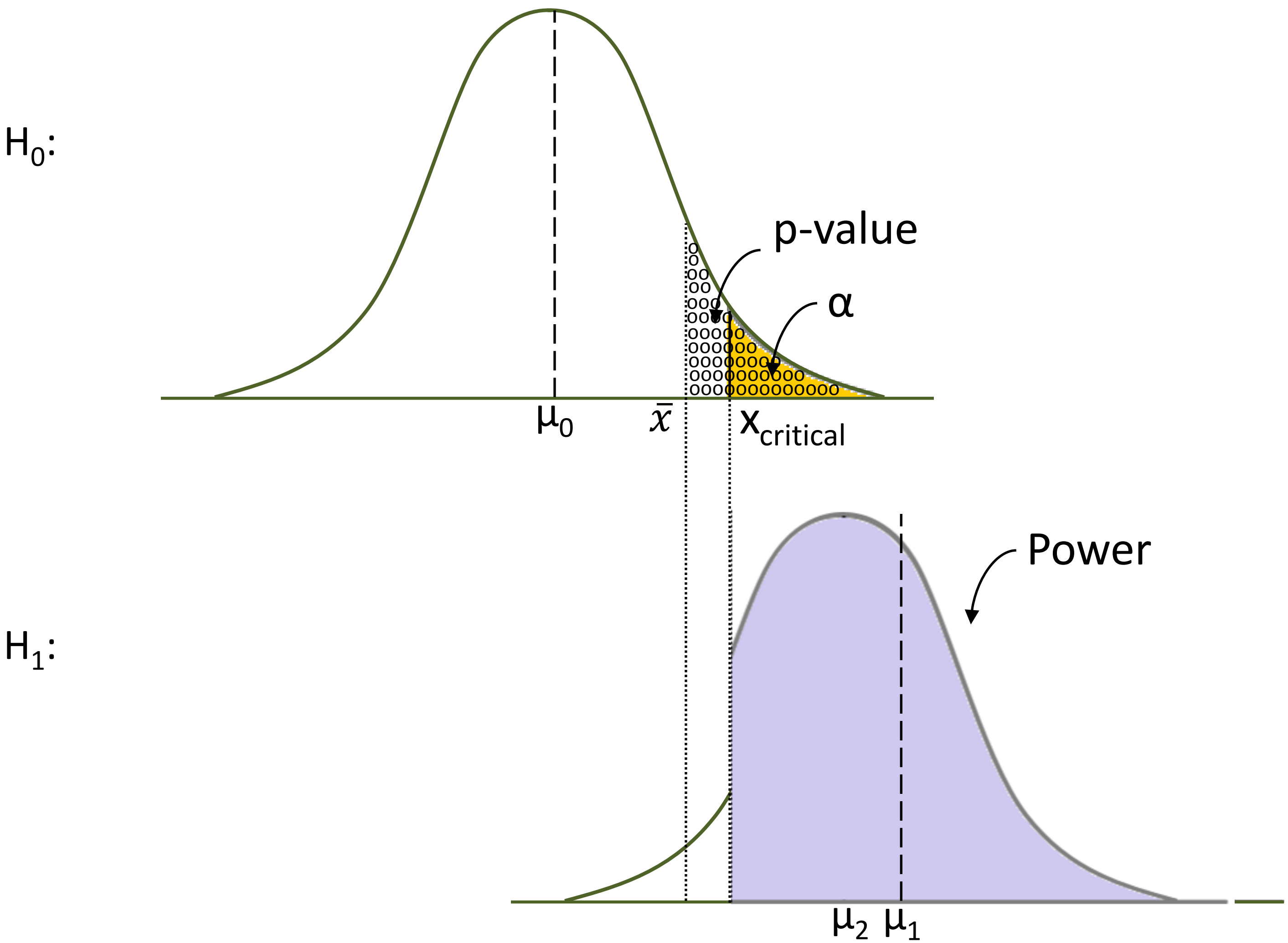
- Found effect may not be meaningful
- Ethically and economically questionable
- Waste of resources!



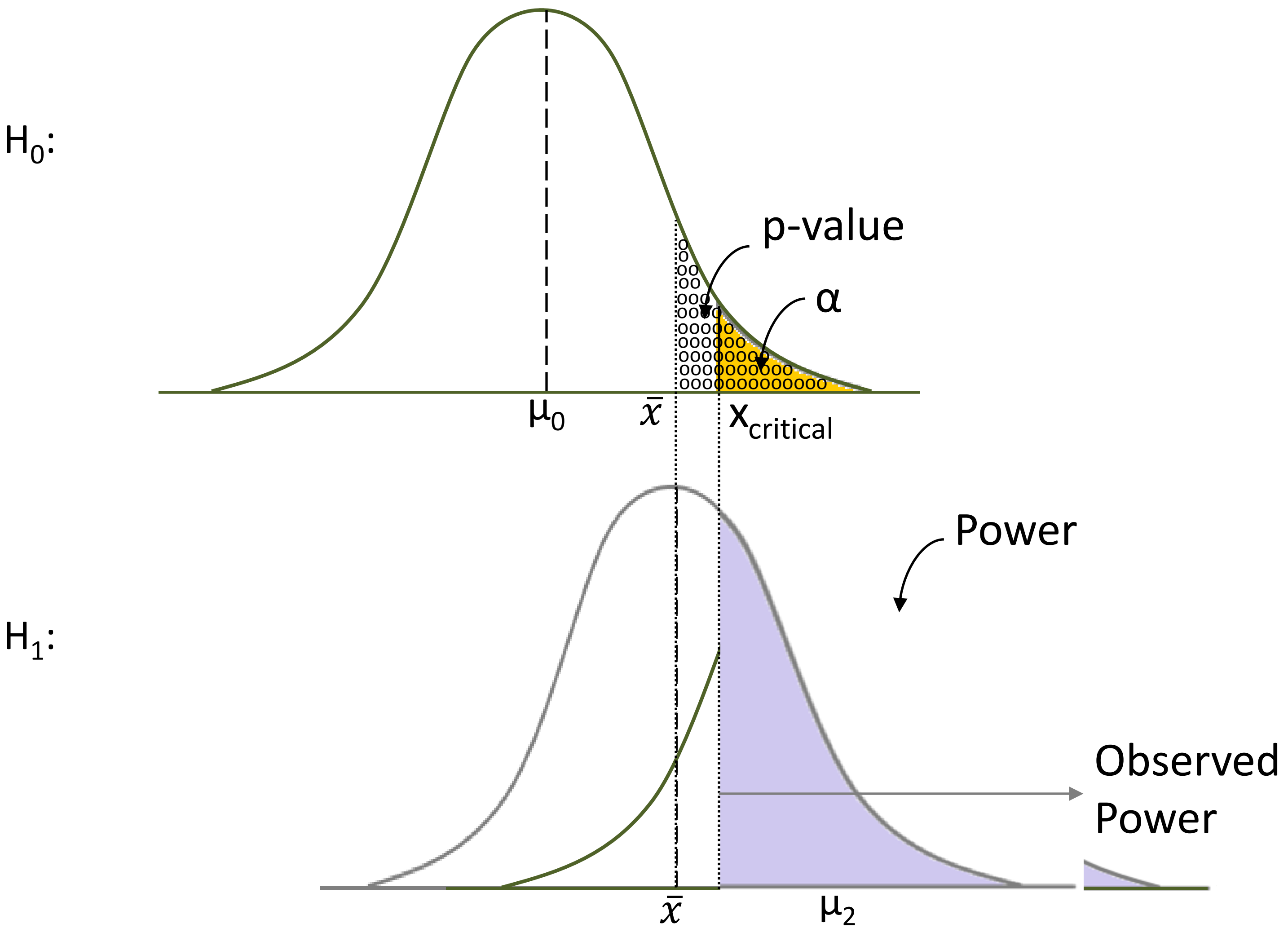
Overpowered Studies

- Found effect may not be meaningful
- Ethically and economically questionable
- Waste of resources!

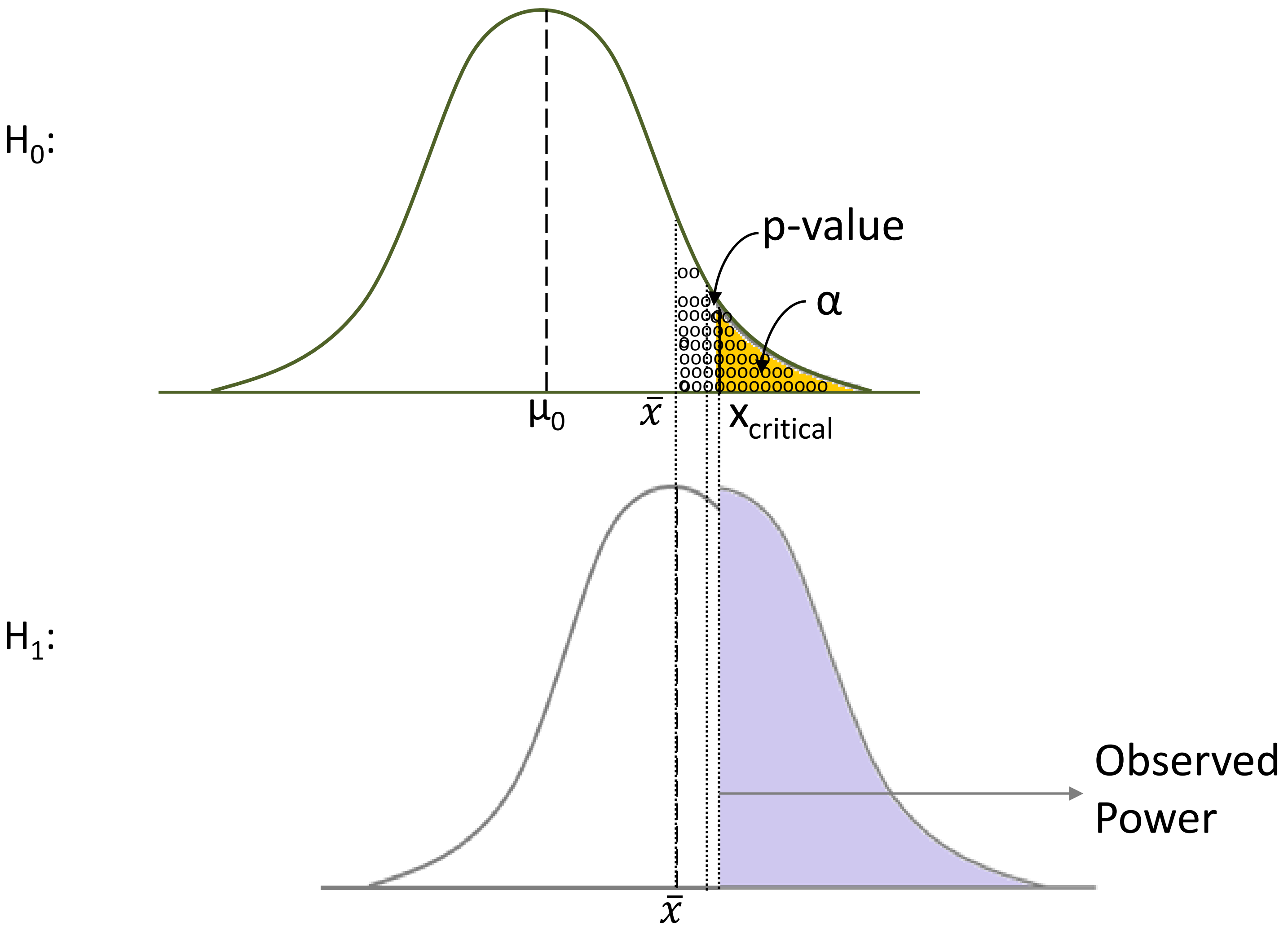
Post-hoc or 'Observed' Power



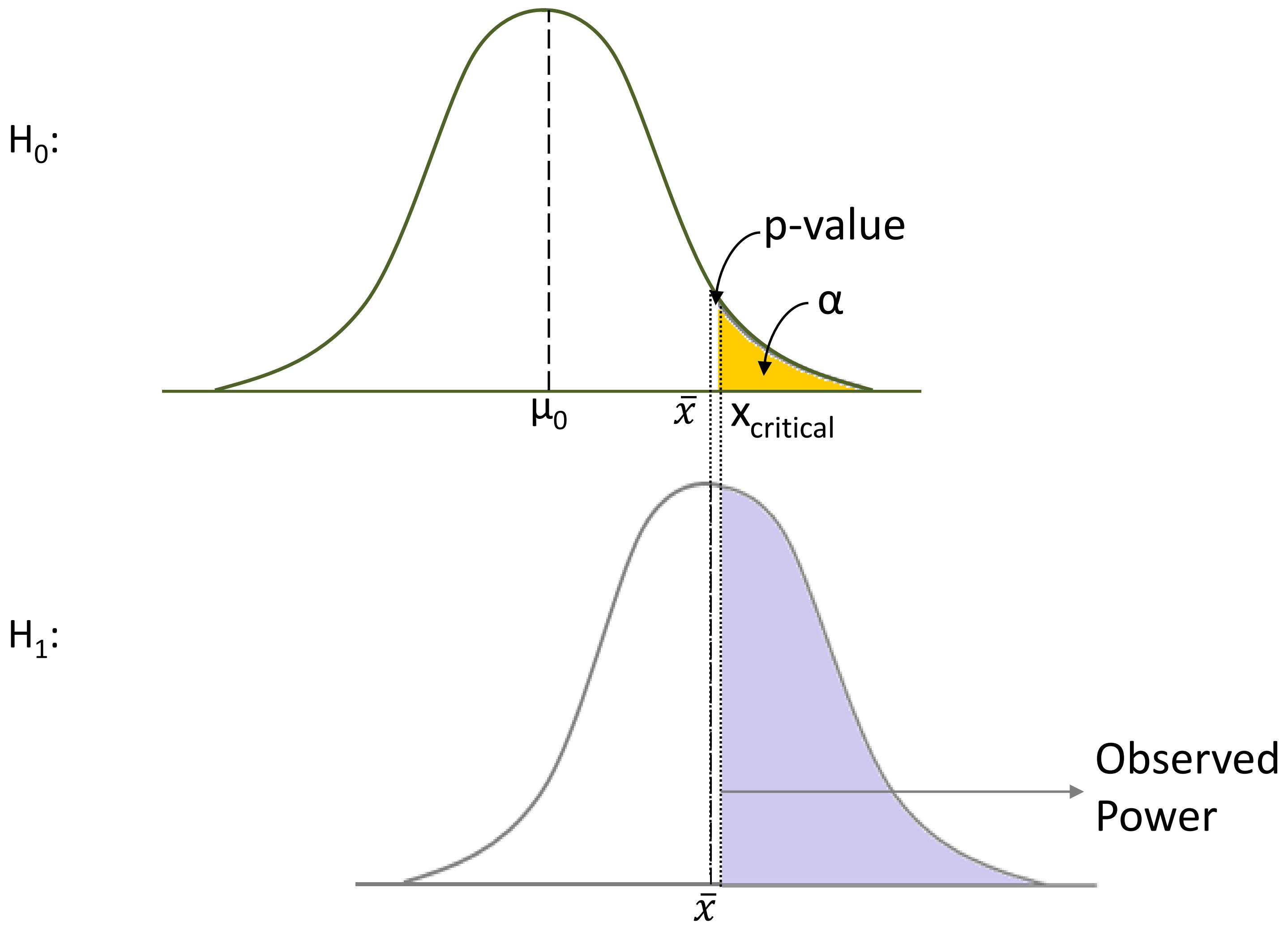
Post-hoc or 'Observed' Power



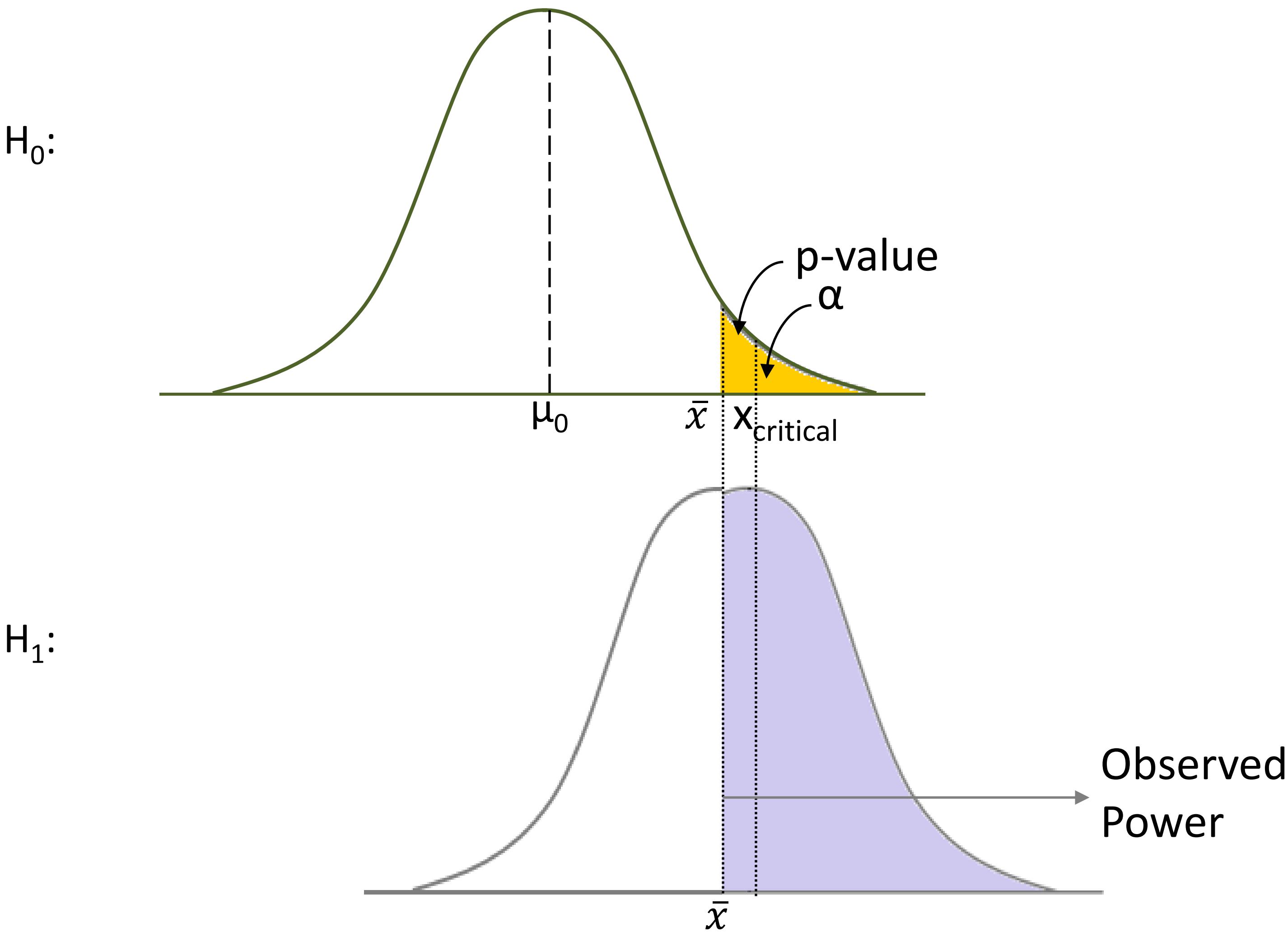
Post-hoc or 'Observed' Power



Post-hoc or 'Observed' Power



Post-hoc or 'Observed' Power



Post-hoc or ‘Observed’ Power

Conclusion??

- ‘Observed’ power is a BAD idea!!!
- Post-hoc power investigation for other alternative parameter values may help in future planning
- Confidence intervals are a much better way to understand the reliability of the results.

Strategies for maximizing power

- Use a continuous outcome with a parametric test when possible
- Use a factorial design and replicate
- Use a one-sided test if defensible
- Include covariates, blocks
- Use a paired/repeated measurements design
- Set the desired effect size as large as possible while still being important from a practical standpoint
- Increase alpha, sample size

Questions

Smita Skrivanek

How have you handled ….

Have you ever encountered ….

Would you explain more how you've approached ….

Visit us at MoreSteam.com

- MBB training/certification offered in partnership with Fisher College of Business at The Ohio State University
- Full suite of online Lean Six Sigma courses, data analysis, project management, and simulation tools
- Free resources, e.g. tutorials, articles, on-demand webcasts

Thank you for joining us

Questions? Comments about today's program?

Smita Skrivanek, MoreSteam.com
sskrivanek@moresteam.com

Ellen Milnes, MoreSteam.com
emilnes@moresteam.com

Join us next month:

Wed., April 29th – Gene Rogers, SteelPointe

Archived presentations and other materials:
http://www.moresteam.com/presentations/

Credits and Citations

- Thomas Scripps (Scripps and Associates), Doug Evans (Ohio State University), Mark Anderson (StatEase), David Seibert (Nationwide), - with thanks for generously sharing their knowledge and expertise.
- Ellen Milnes (MoreSteam.com) – with thanks for her technical assistance.
- Introduction to Power Analysis. UCLA: Statistical Consulting Group. http://www.ats.ucla.edu/stat/seminars/Intro_power/
- Some Practical Guidelines for Effective Sample-Size Determination
  - Russell V. Lenth
- The Abuse of Power
  - John M. Hoenig, Dennis M. Heisey
- Simple Facts about P-Values
  - Craig Blocker et. al.
- Post Hoc Power, Observed Power, …Achieved Power: Sorting out appropriate uses of statistical power analysis
  - Daniel J. O'Keefe