Research Design, Fall 2021

09: statistical power, sample size, simulations

Scott Spencer | Columbia University

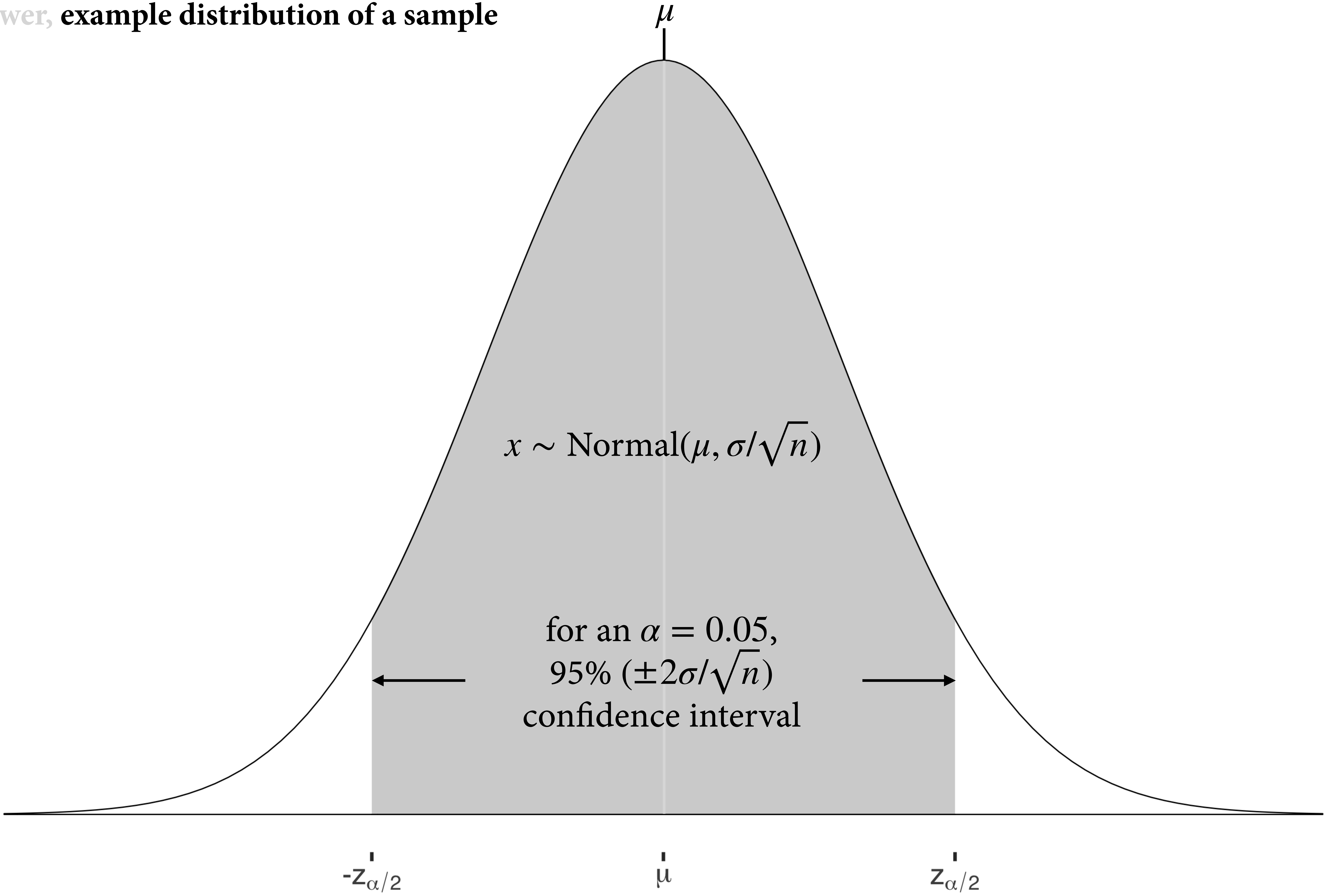

Research Review 1
Homework 2
(research in retail supply chains)
Homework 3
(user behaviors in digital apps)
Homework 4
(customer service for airlines)
Final Exam
SEPT 21
OCT 5
OCT 26
NOV 9
DEC 1
OCT 19
NOV 2
NOV 30
DEC 7
YOU
ARE
HERE
Homework 1
(academic integrity)
Group project
(checkpoint)
Research review 2
Homework 5
(work and life balance)
Group project
(final submission,
presentation)

probability of identifying statistical significance, given an effect — ***statistical power***

STATISTICAL POWER | the probability, *before* a study is performed, that a particular comparison will achieve "statistical significance" at some predetermined level (typically a p-value below 0.05), given some assumed true effect size.

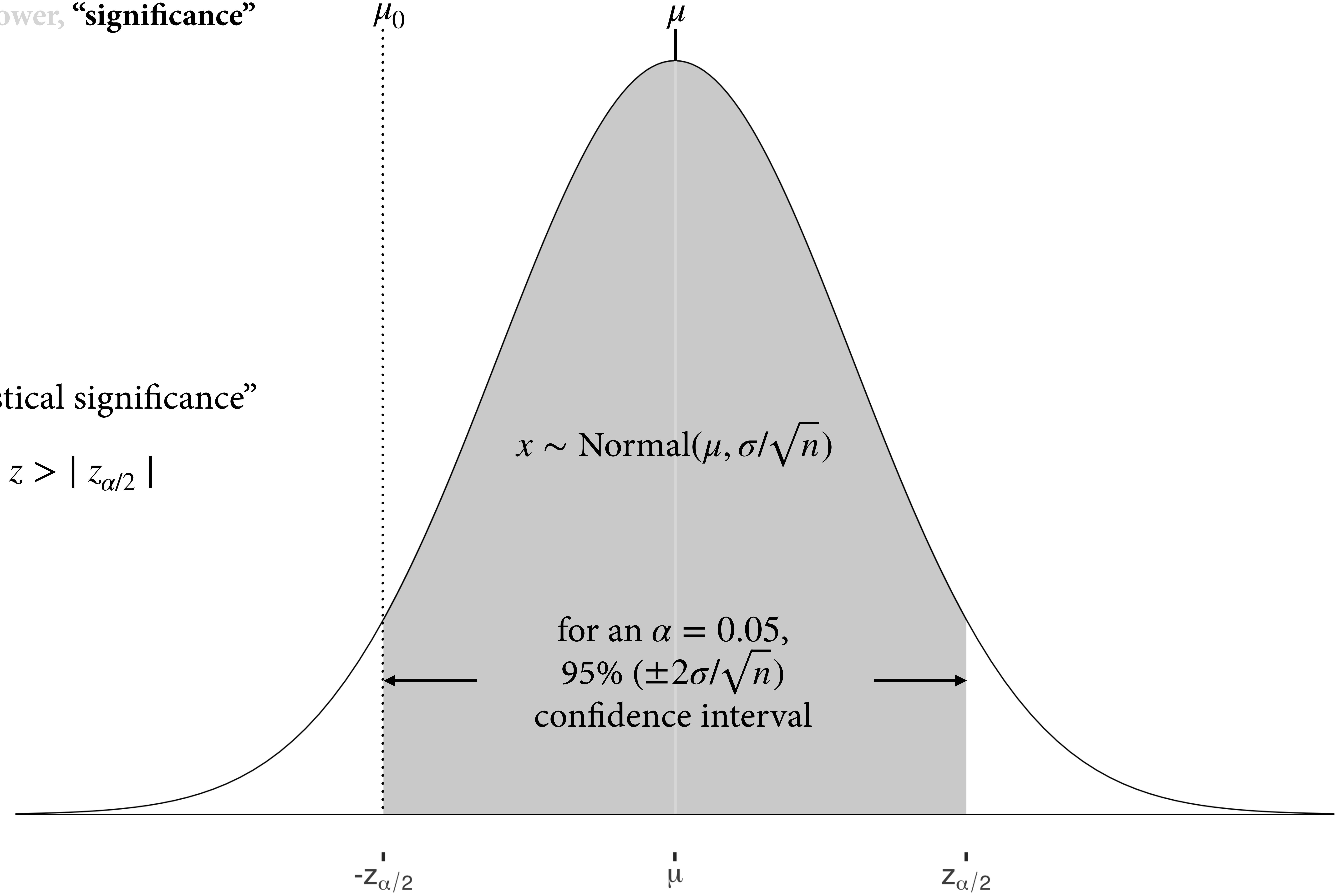
NOTE | a typical threshold for statistical power $1 - \beta$ is 0.8 but — as with choosing a level of confidence α — choice of β should inform good decisions.

 example distribution of a sample



"statistical significance"

$z > | z_{\alpha/2} |$



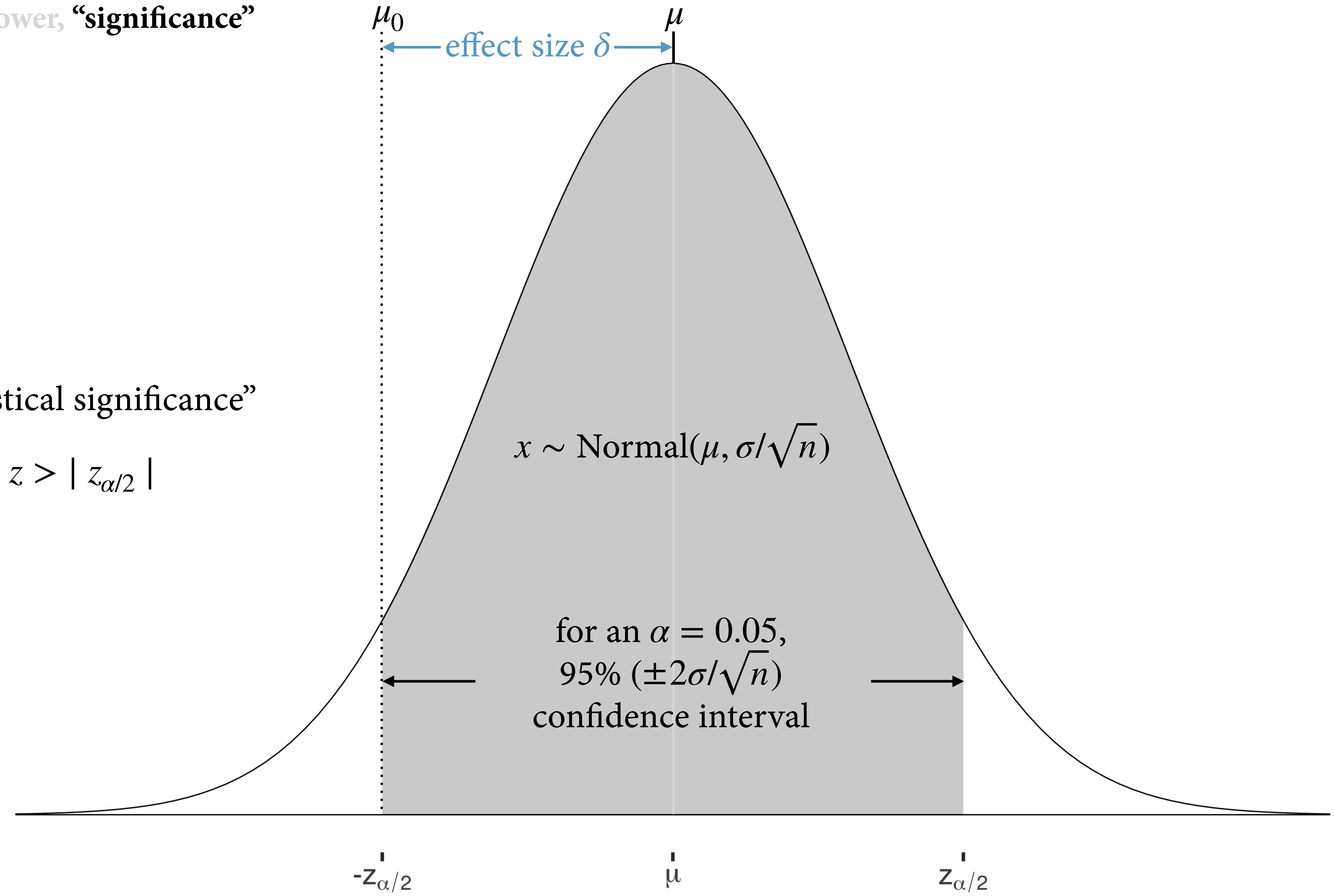
statistical power, “significance tests” focus on limit finding significance ($p < \alpha$) when no difference exists ($\mu_0 = \mu$)

	$p \geq \alpha$	$p < \alpha$
$H_0 : \mu_0 = \mu$	correct conclusion	false positive (Type I error)
$H_a : \mu_0 \neq \mu$	false negative (Type II error)	correct conclusion

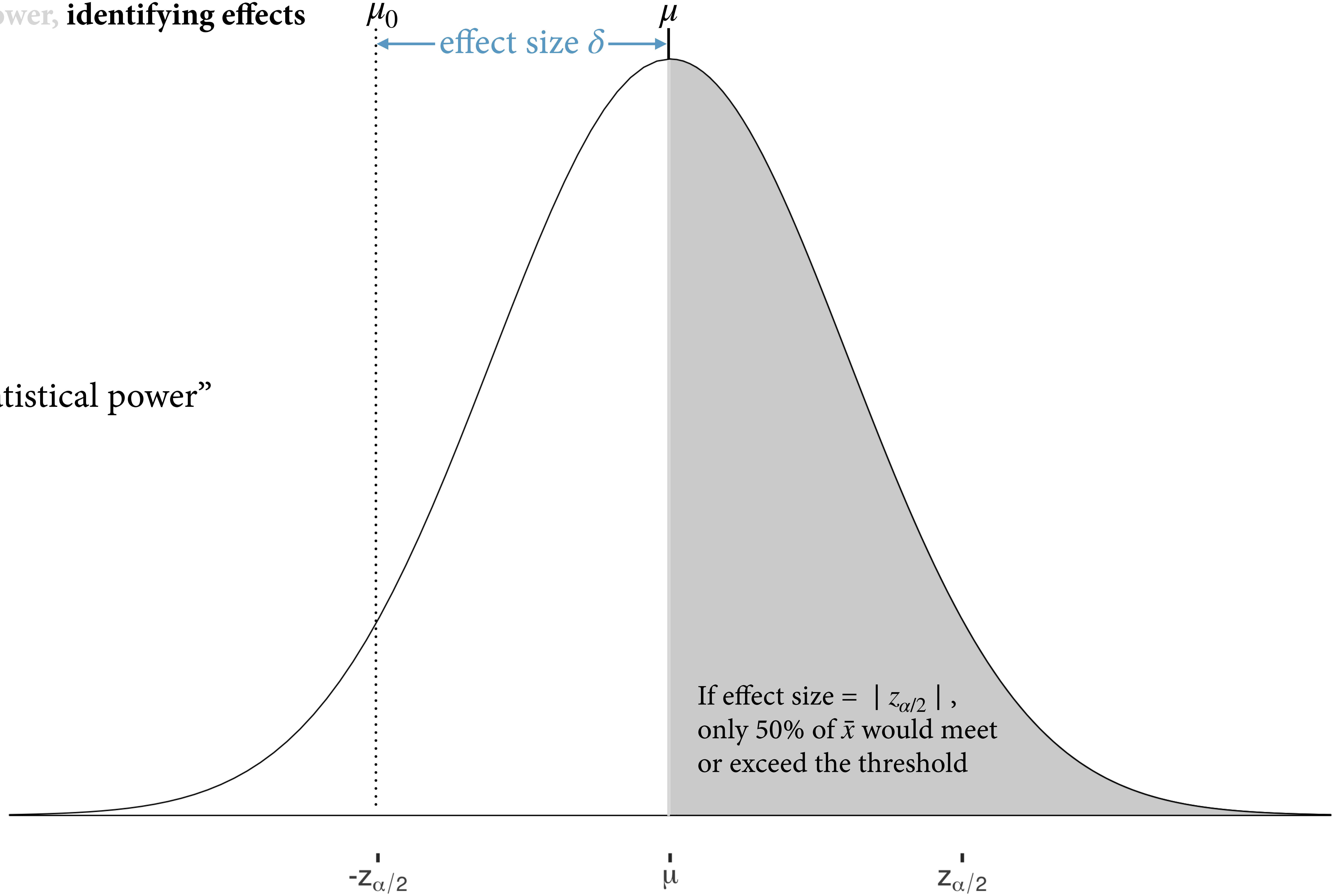
/ https://ssp3nc3r.github.io scott.spencer@columbia.edu

“statistical significance”

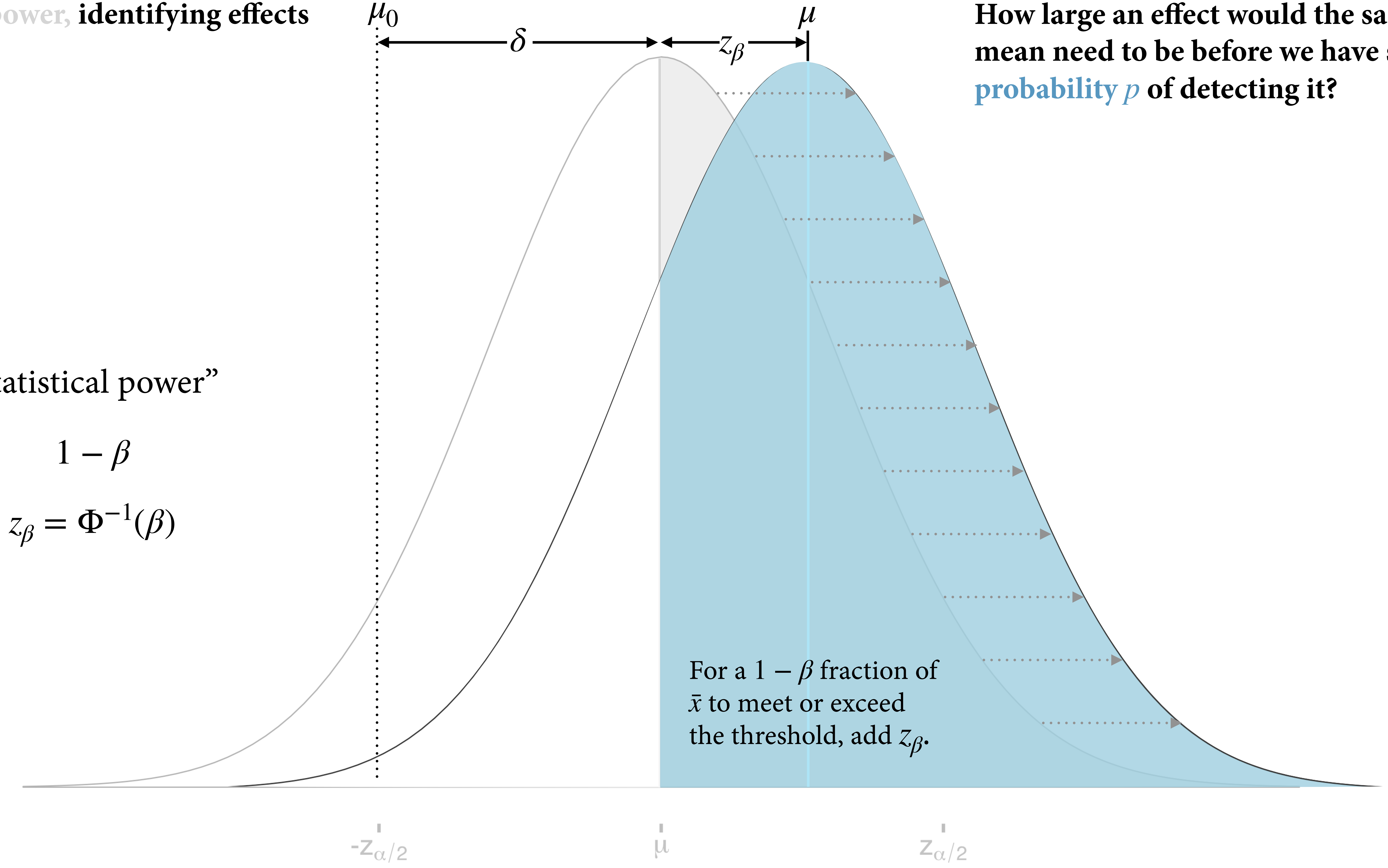
$z > | z_{\alpha/2} |$



“statistical power”



How large an effect would the sample mean need to be before we have some probability p of detecting it?



"statistical power"

$$1-\beta$$

$$z_\beta = \Phi^{-1}(\beta)$$

For a $1-\beta$ fraction of $\bar{x}$ to meet or exceed the threshold, add z_β.

Scenario one — detection? | test with sample size, presumed effect size, distribution, …

Consider a hypothesis to test

$$H_0 : \bar{x} = \mu_0 \ , \ H_a : \bar{x} > \mu_0$$

Choose a sample size

$$n$$

Choose an appropriate test statistic and reference distribution (probability model)

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > |\, z_\alpha \,| \ \text{ compared against } F_\Phi$$

Choose a meaningful effect size and variation to test

$$\bar{x} > \frac{\sigma}{\sqrt{n}} \,|\, z_\alpha \,| + \mu_0$$

Calculate the probability p of identifying an underlying effect

$$p(\bar{x} > \frac{\sigma}{\sqrt{n}} \,|\, z_\alpha \,| + \mu_0) = F_\Phi(\frac{\sigma}{\sqrt{n}} \,|\, z_\alpha \,| + \mu_0)$$

Scenario two — ***sample size*** estimated to achieve a specified probability of finding a real difference

For a given $\mu - \mu_0$, α, and β: $\mu_0 + |\, Z_{\alpha/2} \,| \dfrac{s}{\sqrt{n}} = \mu - |\, Z_\beta \,| \dfrac{s}{\sqrt{n}}$

rearrange: $$(\,|\, z_{\alpha/2} \,| + |\, z_\beta \,|\,) \cdot \frac{s}{\sqrt{n}} = \mu - \mu_0$$

solve for n: $$n = \left[\frac{(\,|\, z_{\alpha/2} \,| + |\, z_\beta \,|\,) \cdot s}{\mu - \mu_0} \right]^2$$

For a given $\mu - \mu_0$, α, and β: $\mu_0 + |Z_{\alpha/2}| \frac{s}{\sqrt{n}} = \mu - |Z_\beta| \frac{s}{\sqrt{n}}$

Let $\mu = 0.6$, $\mu_0 = 0.5$, $\alpha = 0.05$, and $\beta = 0.2$.

rearrange: $(|z_{\alpha/2}| + |z_\beta|) \cdot \frac{s}{\sqrt{n}} = \mu - \mu_0$

$$(1.96 + 0.84)\frac{\sqrt{0.6(1-0.6)}}{\sqrt{n}} = 0.6 - 0.5$$

solve for n: $$n = \left[\frac{(|z_{\alpha/2}| + |z_\beta|) \cdot s}{\mu - \mu_0}\right]^2$$

$$n = \left[\frac{(1.96 + 0.84) \cdot 0.49}{0.1}\right]^2 = 196$$

simulations

 estimating sample size — a toy example, estimate sample size for a proportion, simulating experiments

Let $\mu = 0.6$, $\mu_0 = 0.5$, $\alpha = 0.05$, and try $\beta = \{0.2\ , 0.5\}$

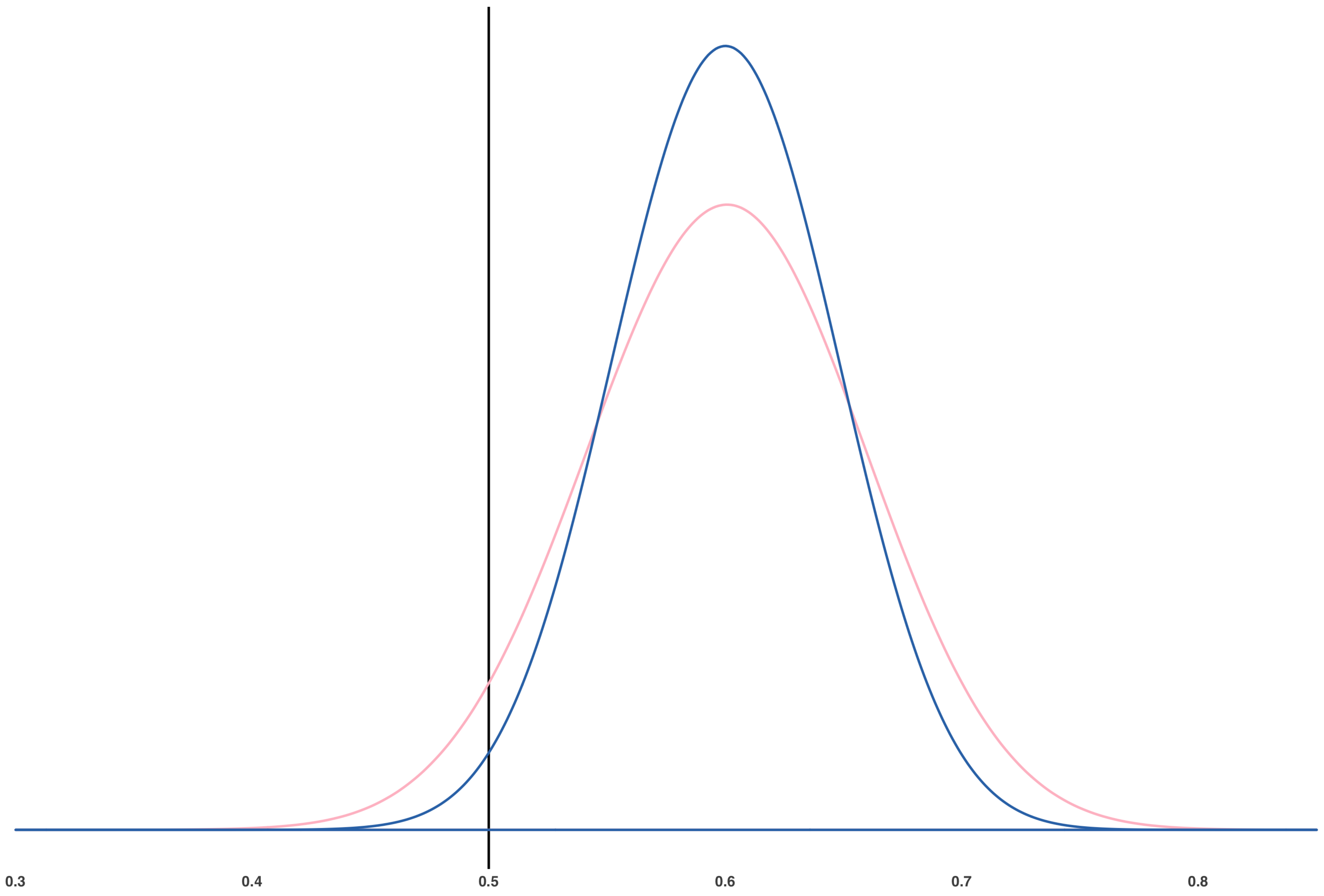
Using $n_{\beta_{0.5}} = 96$ and $n_{\beta_{0.2}} = 196$, simulate experiments.

```
p0    <- 0.5
p     <- 0.6
alpha <- 0.05 / 2

z_alpha_2 <- qnorm(p = alpha, mean = 0, sd = 1, lower.tail = F)

# get n for 80 percent power
beta    <- 0.2
z_beta  <- qnorm(p = 1 - beta, mean = 0, sd = 1, lower.tail = T)
n_pwr80 <- ( (z_alpha_2 + z_beta) * sqrt( p0 * (1 - p0) ) / (p - p0) ) ^ 2

# get n for 50 percent power
beta    <- 0.5
z_beta  <- qnorm(p = 1 - beta, mean = 0, sd = 1, lower.tail = T)
n_pwr50 <- ( (z_alpha_2 + z_beta) * sqrt( p0 * (1 - p0) ) / (p - p0) ) ^ 2
```

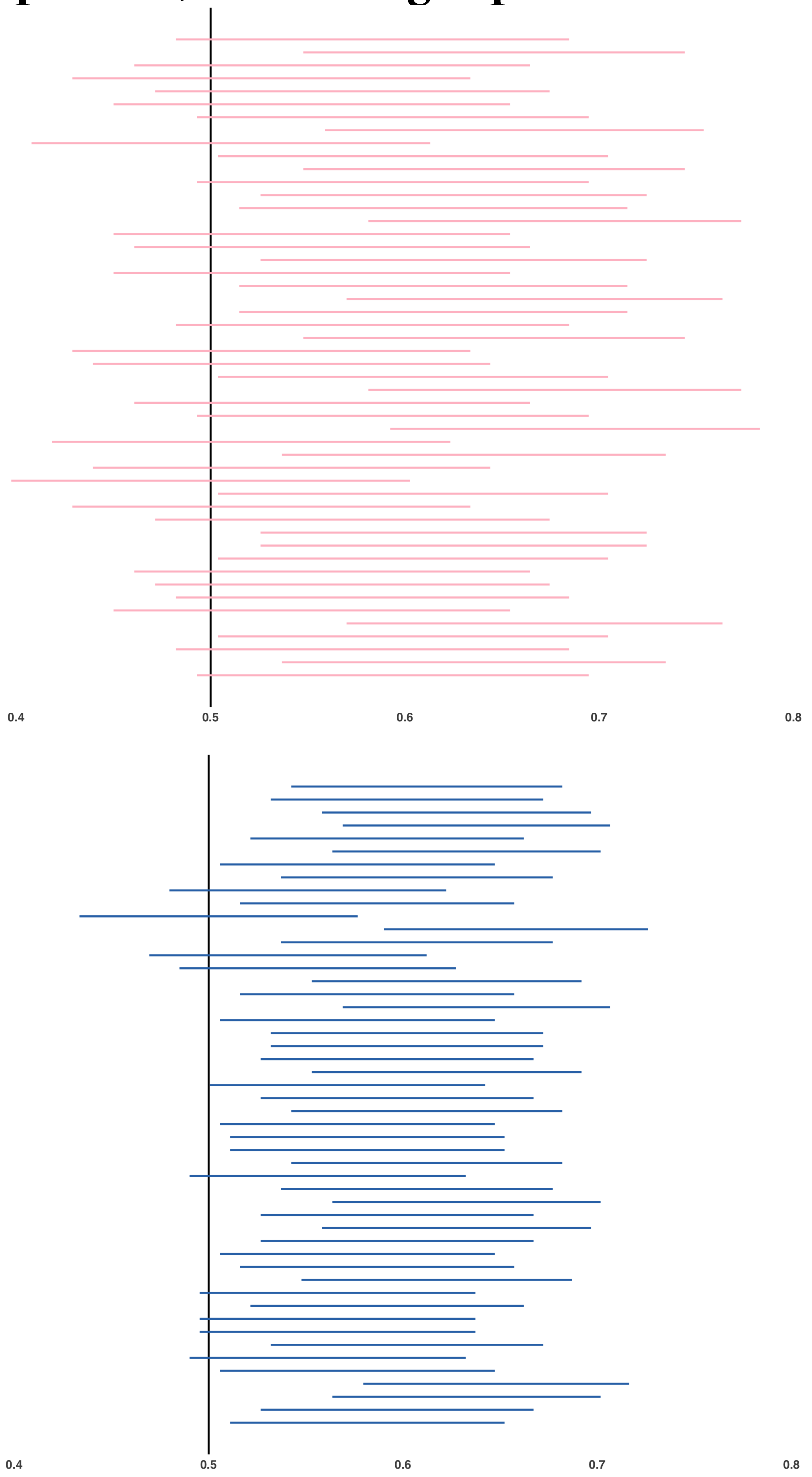
```
# simulate experiments
survey <- function(n, p) {
  x = ( rbinom(n = n, size = 1, prob = p) )
  x_bar = mean(x)
  se = sd(x) / sqrt(n)
  c(x_bar, se)
}


set.seed(1)
p_hat_96  <- replicate( 1e5, survey(n = n_pwr50, p) )

set.seed(1)
p_hat_196 <- replicate( 1e5, survey(n = n_pwr80, p) )
```

$\hat{p}$, with $n = 96$

$\hat{p}$, with $n = 196$



$$\hat{p} \pm 2 \cdot s/\sqrt{n}$$



application in your group projects

resources

References

Cox, D. R., and N. Reid. "Precision and power, Section 8.1.2." In *The Theory of the Design of Experiments*. Monographs on Statistics and Applied Probability 86. Boca Raton: Chapman & Hall/CRC, 2000.

Gelman, Andrew, and John Carlin. "*Beyond Power Calculations*." Perspectives on Psychological Science 9, no. 6 (November 2014): 641–51.

Gelman, Andrew, Jennifer Hill, and Aki Ventari. "Design and sample size decisions, Chp. 16." In *Regression and Other Stories*. S.l.: Cambridge University Press, 2020.

Kruschke, John K, and Torrin M Liddell. "*The Bayesian New Statistics: Hypothesis Testing, Estimation, Meta-Analysis, and Power Analysis from a Bayesian Perspective*." Psychonomic Bulletin & Review 25, no. 1 (March 2018): 1–29.