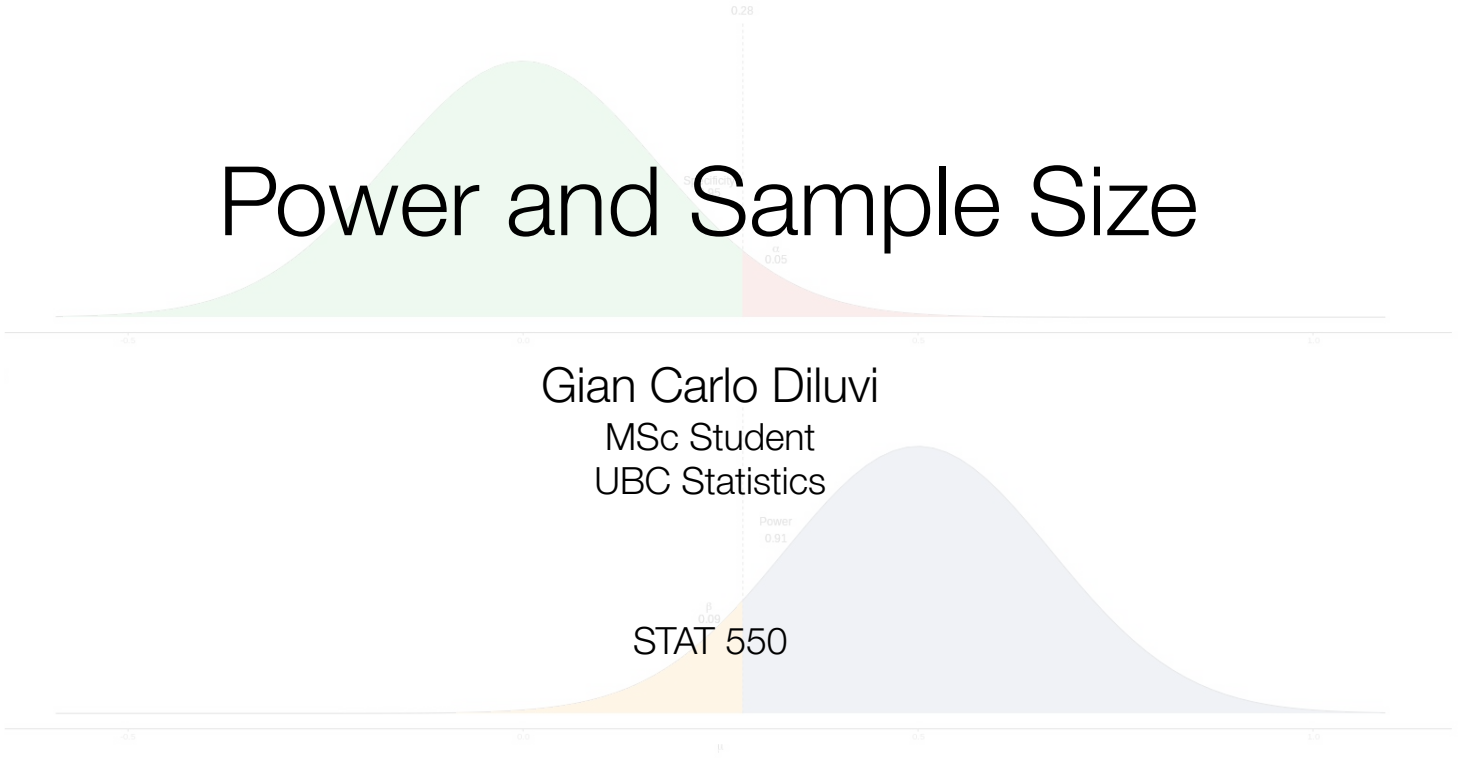
Power and Sample Size

Gian Carlo Diluvi
MSc Student
UBC Statistics

STAT 550



Hypothesis testing

statistical hypothesis testing provides a sound framework to decide whether observed data are unreasonable under some baseline hypothesis H_0

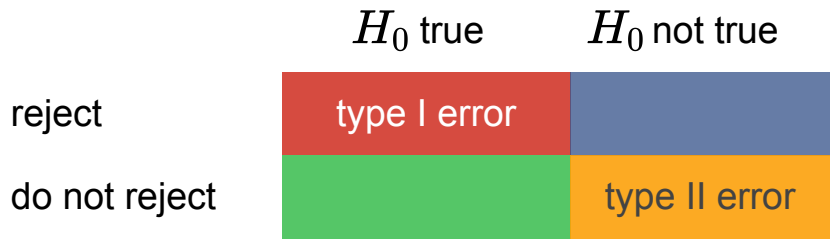
H_0 vs H_1 ← can have a huge impact!

- accused in a trial is **innocent**
- new drug has **same effect** as current treatment
- proportion of families living below poverty threshold is **small**

- accused in a trial is **guilty**
- new drug has a **larger effect** as current treatment
- proportion of families living below poverty threshold is **not small**

Uncertainty is everywhere

we do not know if H_0 is true...



$\mathrm{P}(\blacksquare) = \alpha$ $\quad$ $\mathrm{P}(\blacksquare) =$ power

$\mathrm{P}(\blacksquare) =$ specificity $\quad$ $\mathrm{P}(\blacksquare) = \beta$

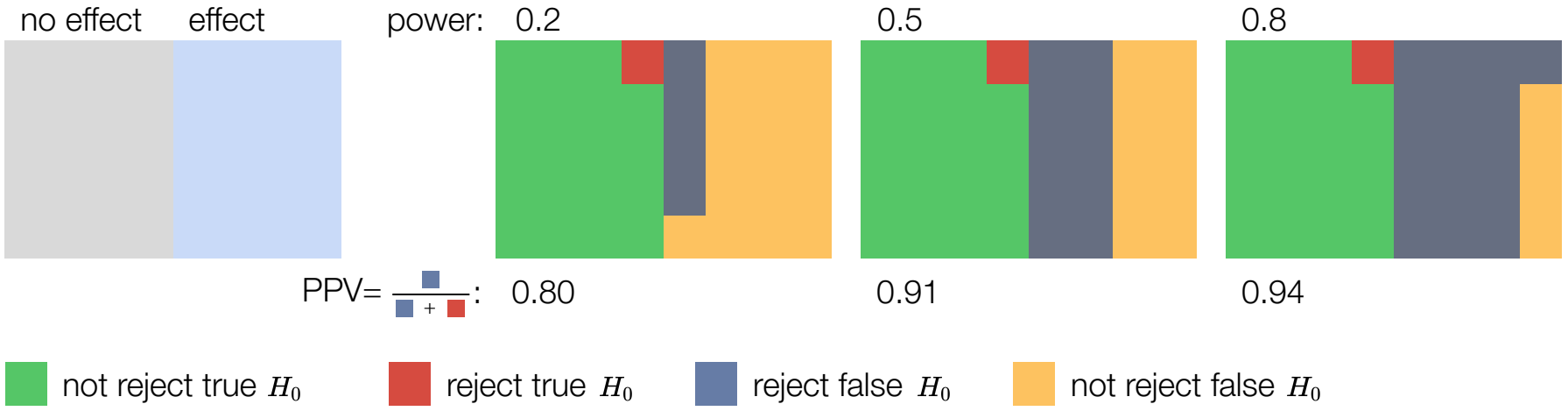
this talk: why power is important, and how it is related to sample size

Low power studies are a waste of resources

power can be as low as 0.2 in fields like neuroscience [Button et al 2013]

this can lead to a high rate of incorrect rejections (low positive predictive value, PPV)

50% effect:

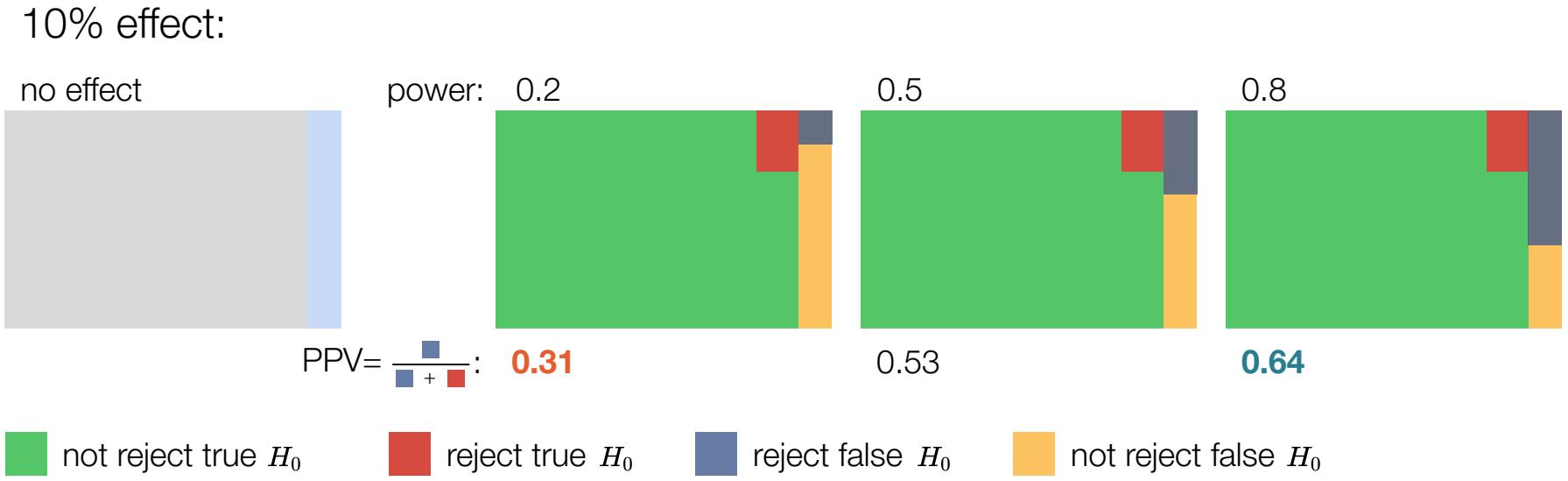


taken from [Krzywinski and Altman 13]

Low power studies are a waste of resources

in studies with multiple outcomes often only around 10% of outcomes have an effect

this can lead to over ⅔ of rejections being wrong with low power, and ⅓ with high

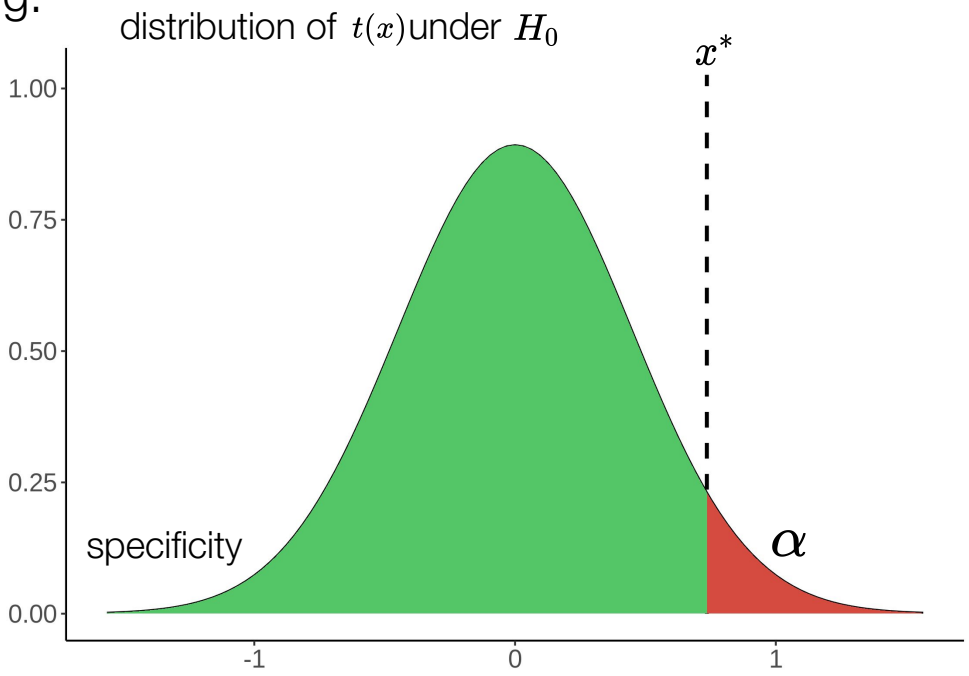


taken from [Krzywinski and Altman 13]

Too much focus is given to specificity

common procedure for hypothesis testing:

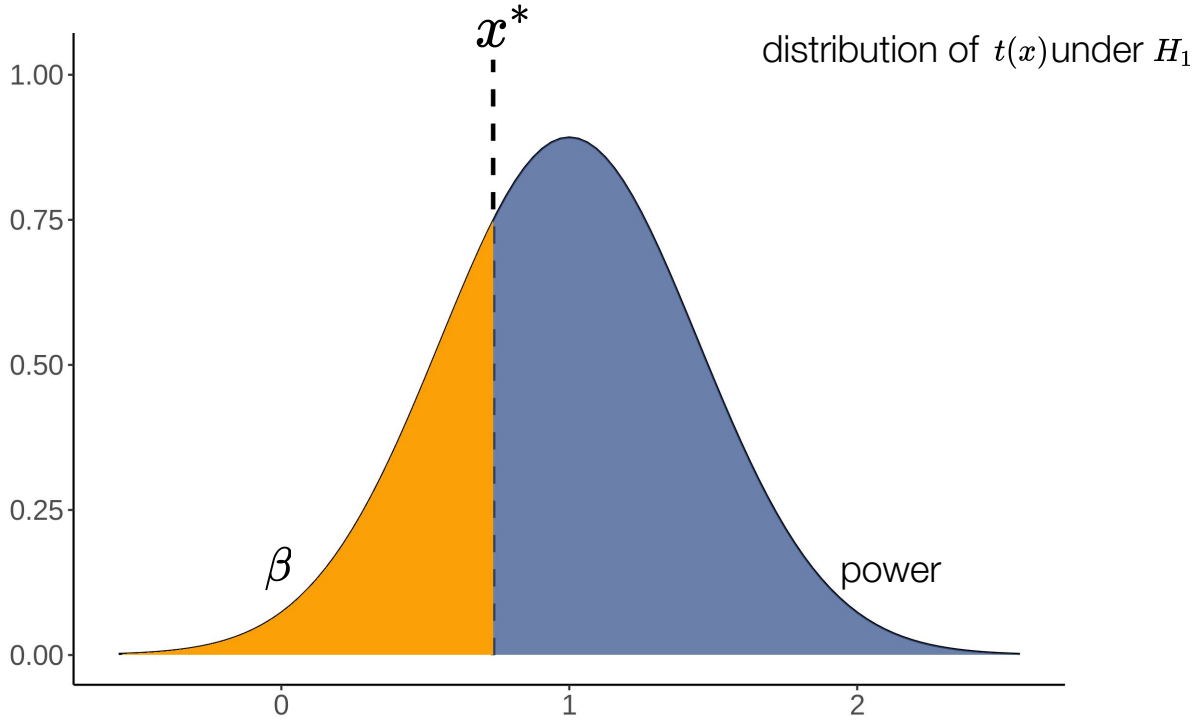
1. fix significance level α
2. compute critical value x^*
3. calculate the test statistic $t(x)$
4. if $t(x) > x^*$, reject H_0



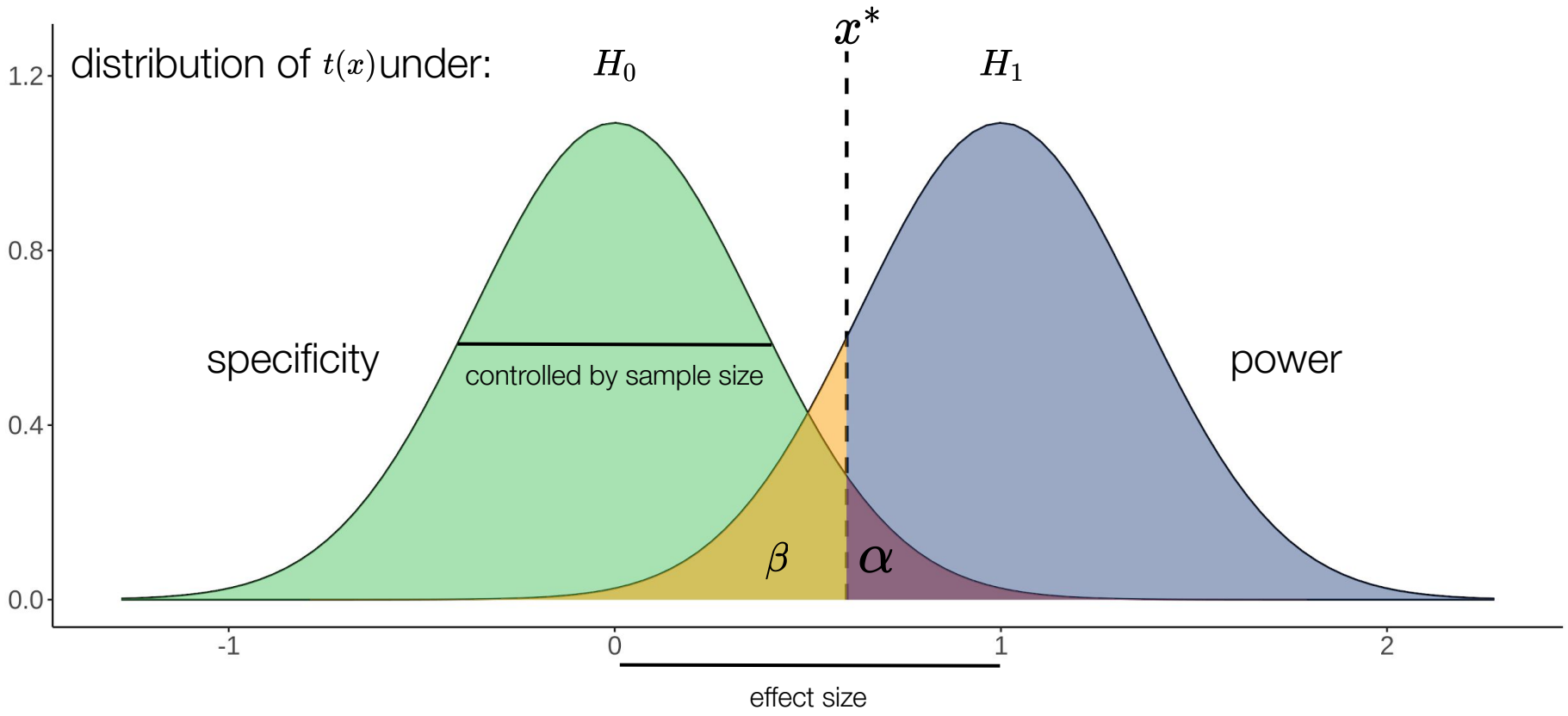
what about the **power**?

The alternative as a means of determining power

if H_0 is not true, H_1 characterizes the true distribution of the data



A complete picture of hypothesis testing

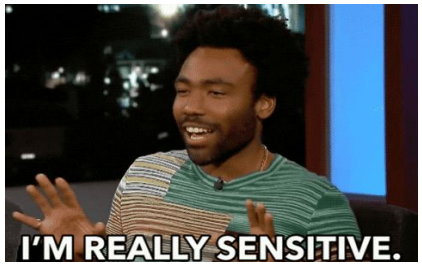


There is no free lunch

low type I error

https://gph.is/2fgMwNJ

small effect detection



https://gph.is/2rXvPyi

pick 3

high power

https://gph.is/2qjWXXP

small sample size

https://gph.is/g/ZWdK7LL

Determining sample size

recall procedure for hypothesis testing:

1. fix significance level α
2. compute critical value $x^*(\alpha, n)$
3. calculate the test statistic $t(x)$
4. if $t(x) > x^*$, reject H_0

use it to find optimal sample size:

1. fix desired α, **power**, **effect size**
2. solve for n:

$$\textbf{power} = \mathrm{P}\left(t(x) > x^*(\alpha, n) \mid H_1\right)$$

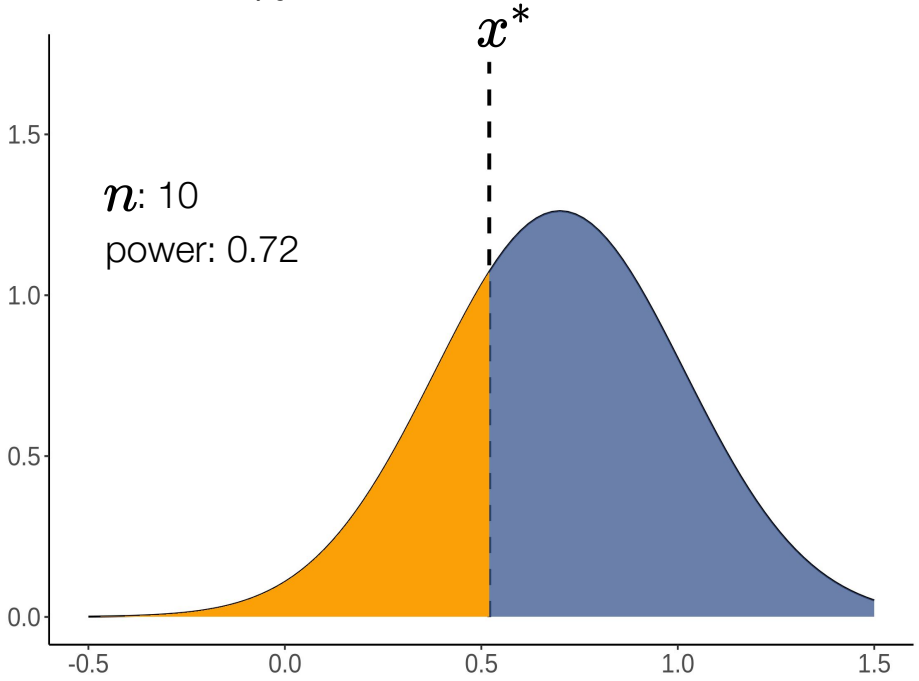
"the n that makes the probability of rejecting H_0 when it is false equal to **power**"

A sample size calculation algorithm

problem: difficult to solve for n

fix desired α, **power**, and **effect size**

1. fix a sample size n
2. calculate critical value $x^*(\alpha, n)$
3. compute power, $\mathrm{P}\left(t(x) > x^*(\alpha, n) \mid H_1\right)$
4. repeat 1-3 until computed power is at least as big as desired **power**

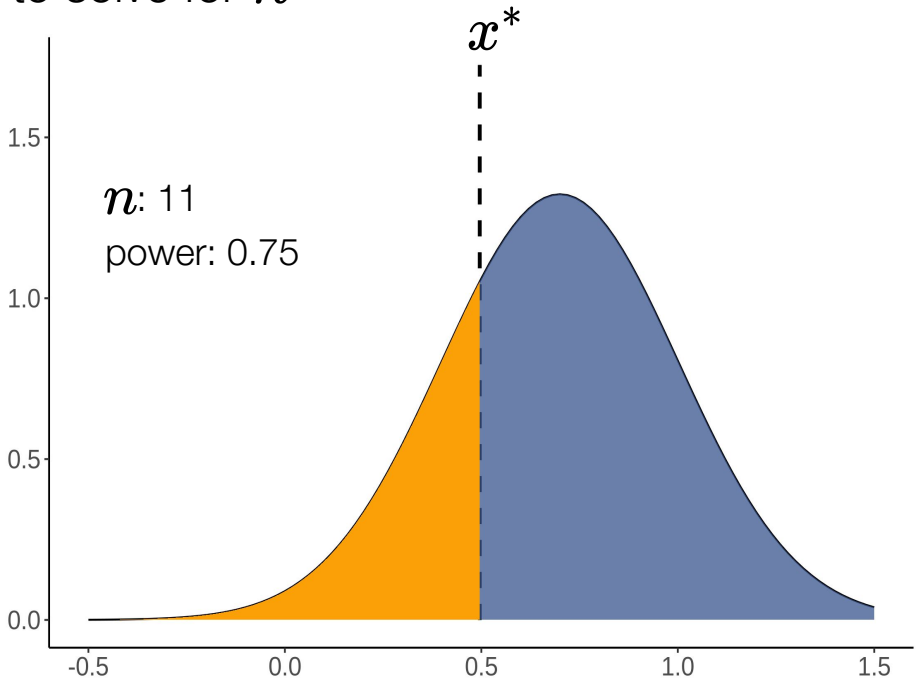


A sample size calculation algorithm

problem: difficult to solve for n

fix desired α, **power**, and **effect size**

1. fix a sample size n
2. calculate critical value $x^*(\alpha, n)$
3. compute power, $\mathrm{P}\left(t(x) > x^*(\alpha, n) \mid H_1\right)$
4. repeat 1-3 until computed power is at least as big as desired **power**

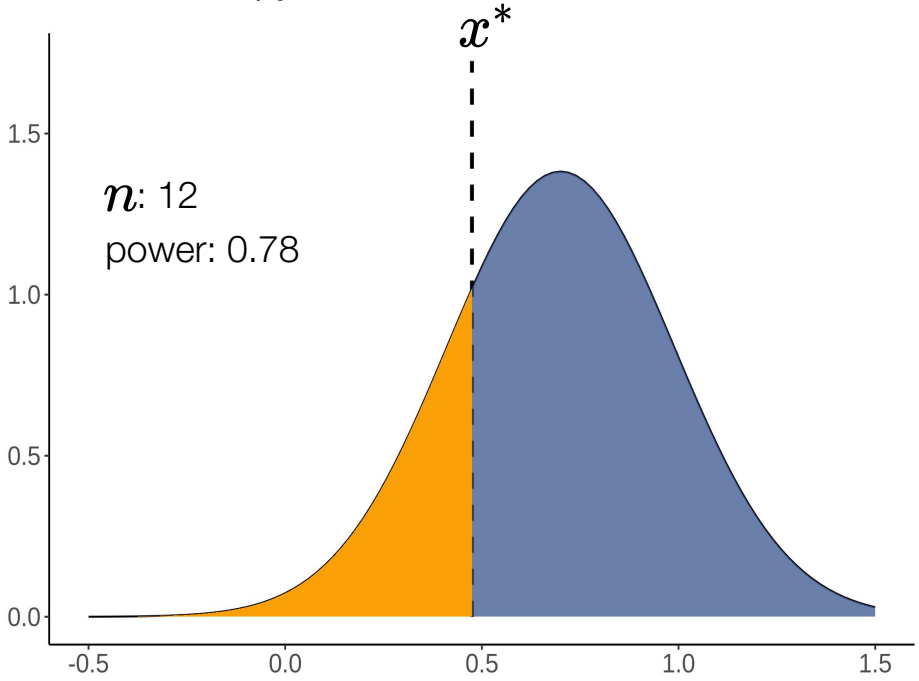


A sample size calculation algorithm

problem: difficult to solve for n

fix desired α, **power**, and **effect size**

1. fix a sample size n
2. calculate critical value $x^*(\alpha, n)$
3. compute power, $\mathrm{P}\left(t(x) > x^*(\alpha, n) \mid H_1\right)$
4. repeat 1-3 until computed power is at least as big as desired **power**

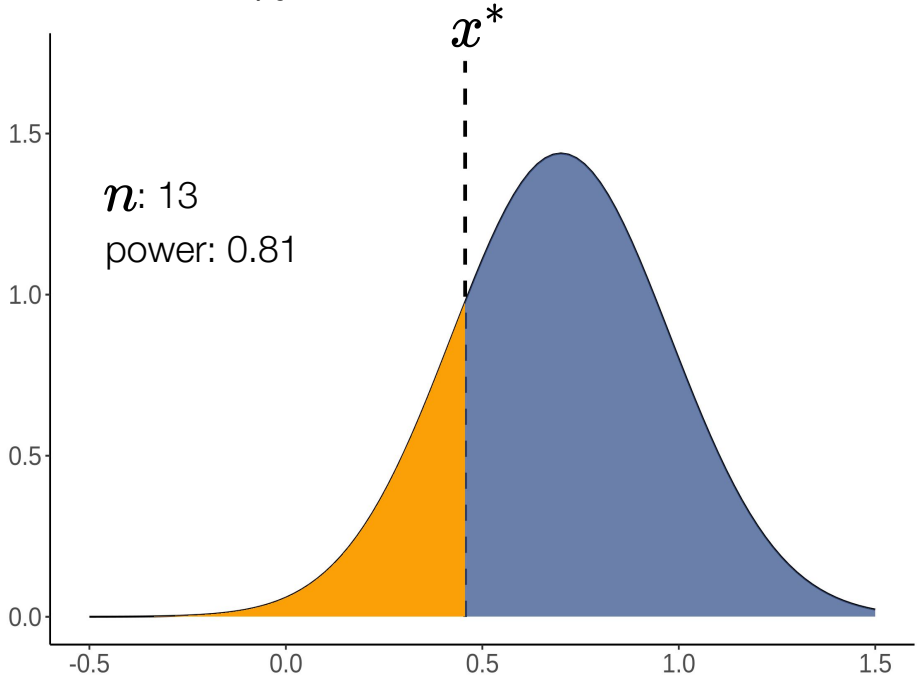


A sample size calculation algorithm

problem: difficult to solve for n

fix desired α, **power**, and **effect size**

1. fix a sample size n
2. calculate critical value $x^*(\alpha, n)$
3. compute power, $\mathrm{P}\left(t(x) > x^*(\alpha, n) \mid H_1\right)$
4. repeat 1-3 until computed power is at least as big as desired **power**

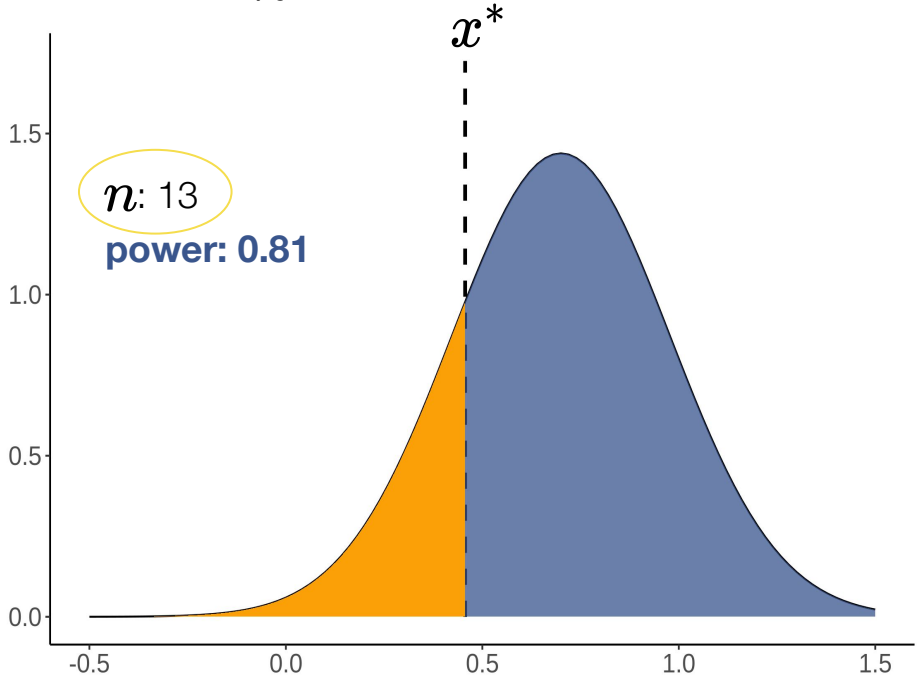


A sample size calculation algorithm

problem: difficult to solve for n

fix desired α, **power**, and **effect size**

1. fix a sample size n
2. calculate critical value $x^*(\alpha, n)$
3. compute power, $\mathrm{P}\left(t(x) > x^*(\alpha, n) \mid H_1\right)$
4. repeat 1-3 until computed power is at least as big as desired **power**



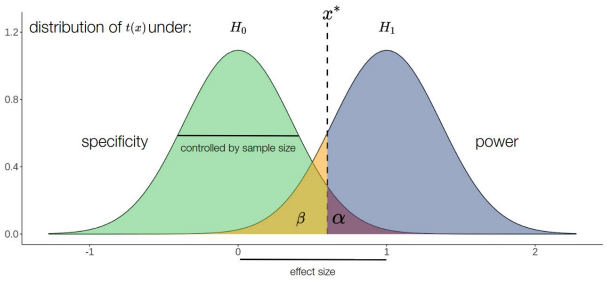
Conclusion

low power studies are a waste of resources

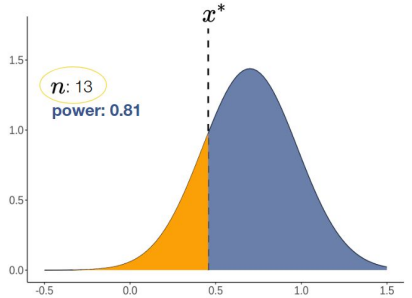
power: 0.2

PPV= $\frac{\blacksquare}{\blacksquare + \blacksquare}$: 0.80

trade-off between α, power, n, and effect size



sample size can be used to increase power

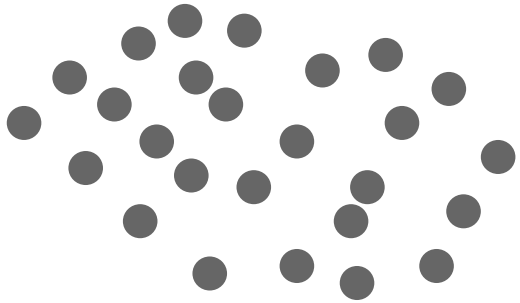


/giankdiluvi/power

dashboard, code, slides:

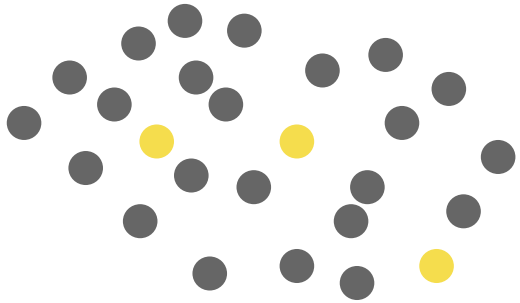
Sample size for survey methodology

we want to **sample** from a **population**



Sample size for survey methodology

we want to **sample** from a **population**

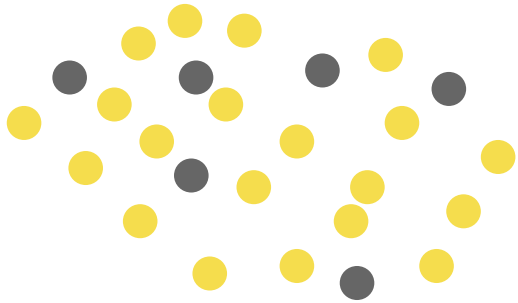


too small...

Sample size for survey methodology

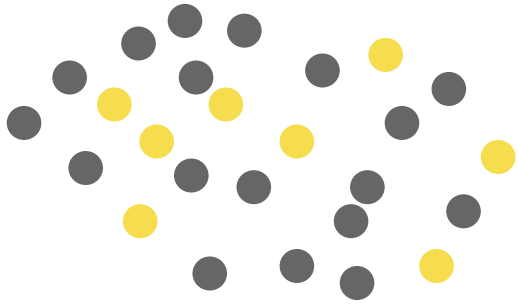
we want to **sample** from a **population**

too large...



Sample size for survey methodology

we want to **sample** from a **population**



about right!

trade-off between **sample size**, how large an **error** you are willing to tolerate, and **how often**

$$P(\text{error} > \blacksquare) < \blacksquare$$

"willing to make an error larger than ■ only ■% of the time"

(classical textbook by [Cochran 77])