04/18/2025

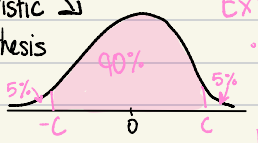
Decision Error

Hypothesis Testing

1) Form a hypothesis: H_0
2) Collect data
3) Create test statistic: $t = \frac{\bar{x} - \mu_0}{\hat{\sigma}/\sqrt{n}}$
4) Make a decision: reject or fail to reject the null hypothesis

- We assume that the null is true

$t = \frac{\bar{x} - \mu_0}{\hat{\sigma}/\sqrt{n}}$ Distribution of t-statistic assuming the null hypothesis is true



Ex: CI for 90%
- The C values bound 90% of the area under the curve, the middle 90% of t-statistics if the null hypothesis is true.

- If your t-statistic is outside the CI you reject the null hypothesis. However there is a chance that your null hypothesis is true and you just have an extreme t-stastic: this is called Type I Error.
- If you increase your confidence, you decrease your type I error rate

	H_0 True	H_0 False
Reject	Type 1 Error α	Correct $1-\beta$ Power
Fail to reject	Correct $1-\alpha$ Confidence	Type II Error β

Type 1 Error
- Reject H_0 when H_0 was true
- "False positive"

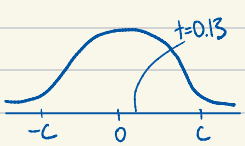
Type II Error
- Fail to reject when H_0 is false
- "Missed opportunity"

- You cannot minimize type I and type II error at the same tim

Example

- Assume the true value of the mean is $\mu = 95$
- Sample: $\bar{x} = 96$, $\hat{\sigma} = 20$, $n = 15$

t-distribution for $\mu_0 = 95$ $\frac{\bar{x} - 95}{\hat{\sigma}/\sqrt{n}}$



what if we were testing $\mu_0 = 100$

We would still fail to reject $\mu_0 = 100$ even though it is untrue. This is a type 2 error

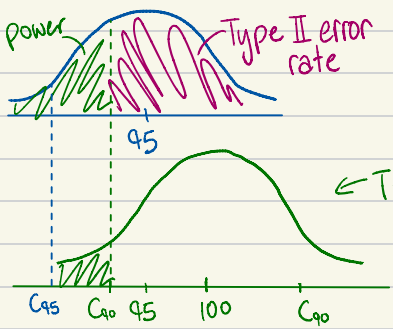
What causes type II error?

- distance between μ and μ_0 - "Effect Size"
- Large σ
- Smaller sample size (n)
  - ↳ With enough evidence you can always reject μ_0 because it will never be exactly identical to μ, but you have to ask yourself, is this difference meaningful? (Ex: a weightloss drug helps people lose ¼ pound)

Example:

Assume the true value of the mean is $\mu = 95$

Here is the distribution of $\bar{x}$:



Note:

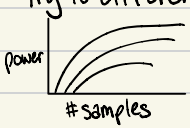
$H_0: \mu = 100$ and H_0 is false

← T-distribution for $\mu_0 = 100$

- Increasing your confidence reduces your Type I Error rate but increases your Type II Error rate.

Power Analysis

- Try 10 different effect sizes and it creates a graph of your power distribution



Read through last 6 slides