Hypothesis Testing

Steps for hypothesis testing

1) state hypothesis
   - null (H_0) - hypothesis of no difference
   - research hypothesis (H_A)
2) calculate probability of getting sample mean
3) decide whether to
   - reject H_0 ($p < .05$)
   - retain H_0 ($p > .05$)
4) make general conclusion

Null hypothesis

- states that there is no difference btwn our statistic and what we would find by chance
  - can compare sample to population (e.g. $H_0 - \mu_1 = \mu$)
  - can compare two samples (e.g. $H_0 - \mu_1 = \mu_2$)
- why use null
  - cannot prove something to be true, but can prove something to be false (theoretical)
  - by starting w/ null, we can put together sampling distribution

Statistical conclusion

- if data are very different from expected → reject H_0
- if data isn't different → fail to reject null hypothesis (retain)
- why do we say "fail to reject?"
  - hypothesis doesn't prove anything
  - failure to reject could be due to a variety of factors
- when retaining null, you aren't actually saying two groups are equal

Significance level

- we set a limit on what values or sufficiently unlikely
  - this value is usually .05 ($p < .05$)
  - ↳ when $p < .05$, we have a less than 5% chance that the difference we see is sampling error
- when p value is less than .05 - sufficiently unlikely and we reject the null
  - we want a p value less than .05
  - indicates that difference is large and meaningful
- when p value is greater than .05 - likely retain the null hypothesis
  - indicates difference is small/due to sampling error

Type One/Two error

- type one error: rejecting null when it is true
  - finding an effect when there isn't one
- type two error: failing to reject null when null is false
  - not finding an effect that is there

decision	reality: H_0 true (no diff)	reality: H_0 false (diff)
reject H_0 (say diff)	type one error $p = \alpha$	correct (power) $p = 1 - \beta$
retain H_0 (say no diff)	correct $p = 1 - \alpha$	type two error $p = \beta$

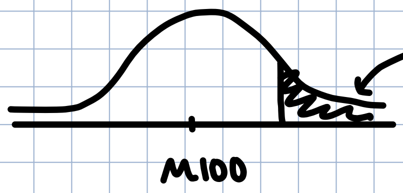
One Tailed & Two Tailed tests

- one tailed (directional) test: decide that scores are in the top 5% (or bottom 5%) are very unlikely → greater than/less than symbol)
- two-tailed (undirectional test): decide that scores at either extreme are unlikely
- decide which test before testing

EXAMPLE:

IQ = μ100

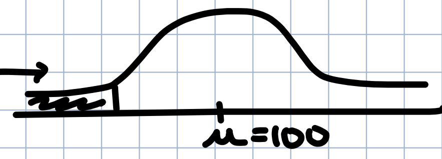
HA: μ > 100 (I think sample is better than pop.)



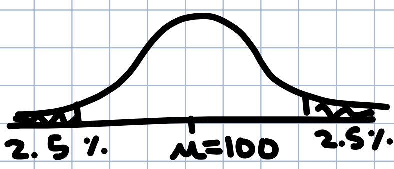
critical region (5%)
α = .05

Ha: μ < 100 (I think sample is less)

critical region 5% (α = .05)



HA: $\mu \neq 100$



*a priori

I believe that our stats class is smarter than average. class has mean IQ of 106. Is our class rly intelligent?

1) HO: μ = 100
   *our class is no different than the rest of pop in IQ scores
   Ha: μ > 100
   *our class will score higher

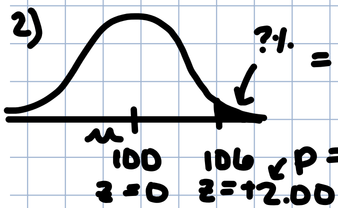
$\mu = 100$
$\sigma = 15$
$N = 25$
$M = 106$

$z_{obt} = \frac{M - \mu}{\sigma_M}$

$z_{obt} = \frac{106 - 100}{3}$

$\sigma_M = \sigma / \sqrt{N}$
$= 15 / \sqrt{25}$
$= 3$

2)



= +2.00

p = .0228

p = 2.3 %. due to chance! → reject Ho → class is intelligent

What is power?

- power: prob. of rejecting false null
    - ability to detect difference if one exists
- factors of effect:
    - true diff. btwn pop means: increases in diff = increases in power
    - N: increases in sample size = increases in power
    - σ_M (standard error): increase in variance → decreases in power
    - α: increases in α = increases in power
    - one tail test increases power
- determine power a priori