Psych 2111
Prof. Matt Jones

Lecture 15: Three Views of Hypothesis Testing

What are we doing?
- In science, we run lots of experiments
- In some cases, there's an "effect"; in others there's not
  - Some manipulations have an impact; some don't
  - Some drugs work; some don't
- Goal: Tell these situations apart, using the sample data
  - If there is an effect, reject the null hypothesis
  - If there's no effect, retain the null hypothesis
- Challenge
  - Sample is imperfect reflection of population, so we will make mistakes
  - How to minimize those mistakes?

Analogy: Prosecution
- Think of cases where there is an effect as "guilty"
  - No effect: "innocent" experiments
- Scientists are prosecutors
  - We want to prove guilt
  - Hope we can reject null hypothesis
- Can't be overzealous
  - Minimize convicting innocent people
  - Can only reject null if we have enough evidence
- How much evidence to require to convict someone?
- Tradeoff
  - Low standard: Too many innocent people get convicted (Type I Error)
  - High standard: Too many guilty people get off (Type II Error)

Binomial Test
- Example: Halloween candy
  - Sack holds boxes of raisins and boxes of jellybeans
  - Each kid blindly grabs 10 boxes
  - Some kids cheat by peeking
- Want to make cheaters forfeit candy
  - How many jellybeans before we call kid a cheater?
- Work out probabilities for honest kids
  - Binomial distribution
  - Don't know distribution of cheaters

- Make a rule so only 5% of honest kids lose their candy
- For each *individual* above cutoff
  - Don't know for sure they cheated
  - Only know that if they were honest, chance would have been less than 5% that they'd have so many jellybeans
- Therefore, our rule catches as many cheaters as possible while limiting Type I errors (false convictions) to 5%

Testing the Mean

- Example: IQs at various universities
  - Some schools have average IQs (mean = 100)
  - Some schools are above average
  - Want to decide based on *n* students at each school
- Need a cutoff for the sample mean
  - Find distribution of sample means for Average schools
  - Set cutoff so that only 5% of Average schools will be mistaken as Smart

Unknown Variance: t-test

- Want to know whether population mean equals μ_0
  - H_0: $\mu = \mu_0$
  - H_1: $\mu \neq \mu_0$
- Don't know population variance
  - Don't know distribution of M under H_0
  - Don't know Type I error rate for any cutoff
- Use sample variance to estimate standard error of M
  - Divide $M - \mu_0$ by standard error to get t
  - We know distribution of t exactly

Two-tailed Tests

- Often we want to catch effects on either side
- Split Type I Errors into two critical regions
  - Each must have probability $\alpha/2$

An Alternative View: p-values

- <u>p-value</u>
  - Probability of a value equal to **or more extreme** than what you actually got
  - Measure of how consistent data are with H_0
- $p > \alpha$
  - t is within t_{crit}
  - Retain null hypothesis
- $p < \alpha$
  - t is beyond t_{crit}
  - Reject null hypothesis; accept alternative hypothesis

Three Views of Inferential Statistics

- Effect size & confidence interval
  - Values of μ_0 that don't lead to rejecting H_0
- Test statistic & critical value
  - Measure of consistency with H_0
- p-value & α
  - Type I error rate
- Can answer hypothesis test at any level
  - Result predicted by H_0 vs. confidence interval
  - Test statistic vs. critical value
  - p vs. α