Comparing means or proportions between two groups

Confidence intervals and p-values

Key Concepts

- Many times in public health and medicine, research is centered about comparing some key summary quantity (mean, proportion) between two (or more) populations based on samples from the populations to be compared
  - Mean weight loss on diet type A versus mean weight loss on diet type B
  - Proportion of responders to treatment versus placebo
  - Etc...
- Study Design Types
  - Paired
  - Unpaired (we will focus on unpaired, but same ideas apply to paired analyses)

Key Question

- After accounting for sampling variability, is there a difference in the quantities being compared? In other words, how do our sample results corroborate (or not) with the idea of a difference at the population level?
  - Just because our estimated means (proportions) are different in value, it does not mean the true means (proportions) are different

- One Approach: Confidence Interval
  - Estimate the difference in means (proportions) between the groups being compared, a standard error for the difference and create a 95% CI: see if 95% CI includes

General Approach:
(sample estimated) ± 2× (estimated standard error of sample estimate)
Confidence Interval Approach

- General Approach:
  
  \( (\bar{x}_1 - \bar{x}_2) \pm 2 \times \hat{SE}(\bar{x}_1 - \bar{x}_2) \)

- Ex:
  
  \[ (\hat{p}_1 - \hat{p}_2) \pm 2 \times SE(\hat{p}_1 - \hat{p}_2) \]

Comparing Two Independent Groups: Example 1

1 Samaha, F., et al. A Low-Carbohydrate as Compared with a Low-Fat Diet in Severe Obesity, New England Journal of Medicine, 348: 21

Abstract from article

ABSTRACT: Patients with bulimia (binge-purge syndrome) frequently complain that they consume a very restrictive diet to avoid gaining weight. To investigate this claim, 23 hospitalized bulimic patients were assessed daily for body weight, caloric intake, macronutrient diet content, activity measures, and body composition estimates during weight-stable periods. Bulimic patients ate fewer kilocalories per kilogram body weight (22.1 ± 4.6 kcal/kg) than did age-matched normal women (29.7 ± 6.5 kcal/kg) but had similar activity levels and body composition. Clinical variables, such as history of laxative abuse, anorexia, or obesity, and physiological characteristics, such as body weight, activity level, or dietary content, could not account for this difference in caloric consumption. Bulimic patients tended to eat a diet lower in fat and higher in protein than did control subjects. These results agree with observations of increased efficiency of caloric utilization in obese patients and support patient complaints of a tendency to gain weight easily.

95% CI for mean difference in calories per kilogram body weight: (patients without bulimia compared to bulimic patients): 3.6 to 11.6 calories
Example 3: Comparing Two Proportions

- We will motivate by using data from the Pediatric AIDS Clinical Trial Group (ACTG) Protocol 076 Study Group.
- Study design:
  - "We conducted a randomized, double-blinded, placebo-controlled trial of the efficacy and safety of zidovudine (AZT) in reducing the risk of maternal-infant HIV transmission"
  - 363 HIV infected pregnant women were randomized to AZT or placebo
- Results:
  - The absolute difference in proportions of infected children (by 18 months) between the AZT group and the placebo group was -15% with 95% CI -22% to -8%


Example 4: Comparing Two Proportions

- Outbreak of Gastroenteritis in Maryland high school
- Sample of 263 students who bought lunch on single day: some ate sandwiches, others ate other meals
- 48% of the 225 students who ate sandwiches fell ill, as compared with 10% of the 38 students who ate other foods:
  - This is an observed difference in proportions of 38%, sandwich eaters compared to non-sandwich eaters
  - A 95% CI for the difference in proportions is 26% to 50%

Same Key Question, Another Approach

- After accounting for sampling variability, is there a difference in the quantities being compared? In other words, how do our sample results corroborate (or not) with the idea of a difference at the population level?
  - Just because our estimated means (proportions) are different in value, it does not mean the true means (proportions) are different
- Second Approach: Hypothesis Testing
  - Specify two competing, all inclusive possibilities for the true differences in means (proportions)
  - Use your data to make a decision: ie use your estimated difference in sample means (proportions) and its estimated standard error to figure out which possibility is more consistent with your study results
Hypothesis Testing Approach

- General Approach: Specifying two exhaustive, mutually exclusive possibilities for true (null and alternative)
  - (null) $H_0$: difference in means (proportions) = 0
  - (alternative) $H_A$: difference in means (proportions) ≠ 0

- Start by assuming null is truth, i.e.: no difference
- Figure out how far your sample estimate is from null in terms of standard errors (distance)
- Translate “distance” into probability: probability of seeing a result as far (extreme) or farther than yours just by chance (sampling error) when null is truth; this probability is called the p-value
- Decide whether p-value is “low enough” to call your results inconsistent with the null as truth (reject the null or “fail to reject” if p-value not “low enough”)

Confidence Interval Approach

- General Approach: Assume null as truth (true difference = 0)
- Compute standardized “distance” between your estimate of truth and null value of 0

$$\frac{(\bar{x}_1 - \bar{x}_2)}{SE(\bar{x}_1 - \bar{x}_2)}$$

$$\frac{(\hat{p}_1 - \hat{p}_2)}{SE(\hat{p}_1 - \hat{p}_2)}$$
Confidence Interval Approach

- Using sampling distribution centered at assumed truth (null, true difference = 0), figure out where your result falls under curve
- Use “standardized distance” to place on curve
- Translate into probability of being “as far” or “farther” away from center (0) - this is the p-value

Comparing Two Independent Groups: Example 1

- "A Low Carbohydrate as Compared with a Low Fat Diet in Severe Obesity"¹
  - 132 severely obese subjects randomized to one of two diet groups
  - Subjects followed for six month period

At the End of Study Period

- "Subjects on the low-carbohydrate diet lost more weight than those on a low fat diet (95% confidence interval for the difference in weight loss between groups, -1.6 to -6.2 kg; p < .01)"

¹ Samaha, F., et al. A Low-Carbohydrate as Compared with a Low-Fat Diet in Severe Obesity, New England Journal of Medicine, 348: 21

Comparing Two Independent Groups: Example 1

- Abstract from article

ABSTRACT: Patients with bulimia (binge-purge syndrome) frequently complain that they consume a very restrictive diet to avoid gaining weight. To investigate this claim, 23 hospitalized bulimic patients were assessed daily for body weight, caloric intake, macronutrient diet content, activity measures, and body composition estimates during weight-stable periods. Bulimic patients ate fewer kilocalories per kilogram body weight (22.1 ± 4.6 kcal/kg) than did age-matched normal women (29.7 ± 6.5 kcal/kg) but had similar activity levels and body composition. Clinical variables such as history of laxative abuse, anorexia, or obesity, and physiological characteristics, such as body weight, activity level, or dietary content, could not account for this difference in caloric consumption. Bulimic patients tended to eat a diet lower in fat and higher in protein than did control subjects. These results agree with observations of increased efficiency of caloric utilization in obese patients and support patient complaints of a tendency to gain weight easily.

95% CI for mean difference in calories per kilogram body weight: (patients without bulimia compared to bulimic patients): 3.6 to 11.6 calories (p < .01)
Example 3: Comparing Two Proportions

- We will motivate by using data from the Pediatric AIDS Clinical Trial Group (ACTG) Protocol 076 Study Group.
- Study design
  - "We conducted a randomized, double-blinded, placebo-controlled trial of the efficacy and safety of zidovudine (AZT) in reducing the risk of maternal-infant HIV transmission.
  - 363 HIV infected pregnant women were randomized to AZT or placebo.
- Results
  - The absolute difference in proportions of infected children (by 18 months) between the AZT group and the placebo group was -15% with 95% CI -22% to -8%.
  - The result was statistically significant (p<0.001).


Example 4: Comparing Two Proportions

- Outbreak of Gastroenteritis in Maryland high school.
- Sample of 263 students who bought lunch on single day: some ate sandwiches, others ate other meals.
- 48% of the 225 students who ate sandwiches fell ill, as compared with 10% of the 38 students who ate other foods:
  - This is an observed difference in proportions of 38%, sandwich eaters compared to non-sandwich eaters.
  - A 95% CI for the difference in proportions is 26% to 50%.
  - The result is statistically significant (p<0.001).

p-values

- p-values are probabilities (numbers between 0 and 1).
- Small p-values mean that the sample results are unlikely when the null is true.
- The p-value is the probability of obtaining a result as or more extreme than you did by chance alone assuming the null hypothesis H0 is true.
  - How likely your sample result (and other result less likely) are if null is true.
**p-values**

- The p-value is NOT the probability that the null hypothesis is true!
- The p-value alone imparts no information about scientific/substantive content in result of a study
- For example: in the AZT/infant transmission study had you only been told
  - The result was "statistically significant" (p < .001)
  - What is the direction of the comparison?
  - What is the magnitude of the association?

**p-values**

- If the p-value is small either a very rare event occurred and $H_0$ is true
  - OR
  - $H_0$ is false
- Type I error
  - Claim $H_0$ is true when in fact $H_0$ is true
  - The probability of making a Type I error is called the alpha-level ($\alpha$-level) or significance level

Note on the p-value and the alpha-Level

- If the p-value is less then some pre-determined cutoff (e.g. .05), the result is called "statistically significant"
- This cutoff is the $\alpha$-level
- The $\alpha$-level is the probability of a type I error
- It is the probability of falsely rejecting $H_0$ when $H_0$ true
- Idea: keep chance of " making a mistake" when $H_0$ true low and only reject if sample result "unlikely"
  - Unlikeliness threshold determined by $\alpha$-level
Note on the p-value and the alpha-Level

- Truth versus decision made by hypothesis testing

<table>
<thead>
<tr>
<th></th>
<th>H₀</th>
<th>Hₐ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TRUTH</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DECISION</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reject H₀</td>
<td>Type I Error</td>
<td>Power</td>
</tr>
<tr>
<td></td>
<td>alpha-level</td>
<td>1-beta</td>
</tr>
<tr>
<td>Not</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reject H₀</td>
<td></td>
<td>Type II Error</td>
</tr>
<tr>
<td></td>
<td></td>
<td>beta</td>
</tr>
</tbody>
</table>

Connection: Hypothesis Testing and CIs

- The confidence interval gives plausible values for the population parameter
  - "data take me to the truth"
- Hypothesis testing postulates two choice for the population parameter
  - "here are two possibilities for the truth, data help me choose one"

95% Confidence Interval

- If 0 is not in the 95% CI, then we would reject H₀ that \( \mu = 0 \) at level \( \alpha = .05 \) (the p-value < .05)
- Why?
- With confidence interval we start at sample mean difference and go 2 standard errors in either direction (or slightly more in small samples)
If 0 is not in the 95% CI, then this must mean the parameter is > 2 standard errors away from 0 (either above or below). Hence, the distance (t) will be > 2 or < -2: and the resulting p-value < .05

95% Confidence Interval and p-value

- So, in the weight change/diet type example, the 95% confidence interval tells us that the p-value is less than .05, but it doesn’t tell us that it is case p = .009.
- The confidence interval and the p-value are complementary
- However, you can’t get the exact p-value from just looking at a confidence interval, and you can’t get a sense of the scientific/substantive significance of your study results by looking at a p-value

More on the p-value

- Lack of statistical significance is not the same as lack of scientific significance: must evaluate in context of study, sample size
- Small n can sometimes produce a non-significant even though the magnitude of the association at the population level is real and important (our study just can’t detect it)
- “Low power” in small sample studies makes not rejecting hard to interpret
- Sometimes small studies are designed without power in mind just to generate preliminary data