Lecture 13 - Tests of Two Means (cont.)

Example - Reading and Writing

200 randomly selected high school students took a reading and writing test and their scores are shown below. At a first glance, does there appear to be a difference between the average reading and writing test score?

<table>
<thead>
<tr>
<th>id</th>
<th>read</th>
<th>write</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70</td>
<td>57</td>
</tr>
<tr>
<td>2</td>
<td>86</td>
<td>44</td>
</tr>
<tr>
<td>3</td>
<td>141</td>
<td>63</td>
</tr>
<tr>
<td>4</td>
<td>172</td>
<td>47</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>200</td>
<td>137</td>
<td>63</td>
</tr>
</tbody>
</table>

Do you think reading and writing scores are independent?
**Parameter and point estimate**

- **Parameter of interest**: Average difference between the reading and writing scores of *all* high school students.  
  \[ \mu_{\text{diff}} \]

- **Point estimate**: Average difference between the reading and writing scores of *sampled* high school students.  
  \[ \bar{x}_{\text{diff}} \]

**Setting the hypotheses**

If in fact there was no difference between the scores on the reading and writing exams, what would you expect the average difference to be?

What are the hypotheses for testing if there is a difference between the average reading and writing scores?

- **\( H_0 \)**: There is no difference between the average reading and writing score.
  \[ \mu_{\text{diff}} = 0 \]

- **\( H_A \)**: There is a difference between the average reading and writing score.
  \[ \mu_{\text{diff}} \neq 0 \]

**Example - Zinc**

Trace metals in drinking water affect the flavor and unusually high concentrations can pose a health hazard. Data were collected by measuring zinc concentration at the bottom and at the surface of 10 randomly sampled wells in Wake country.

We would like to evaluate whether the true average concentration of zinc at the bottom of the well water exceeds that of the surface water. Data are given below.

<table>
<thead>
<tr>
<th>well</th>
<th>zinc</th>
<th>location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.43</td>
<td>bottom</td>
</tr>
<tr>
<td>2</td>
<td>0.266</td>
<td>bottom</td>
</tr>
<tr>
<td>3</td>
<td>0.567</td>
<td>bottom</td>
</tr>
<tr>
<td>4</td>
<td>0.531</td>
<td>bottom</td>
</tr>
<tr>
<td>5</td>
<td>0.707</td>
<td>bottom</td>
</tr>
<tr>
<td>6</td>
<td>0.716</td>
<td>bottom</td>
</tr>
<tr>
<td>7</td>
<td>0.651</td>
<td>bottom</td>
</tr>
<tr>
<td>8</td>
<td>0.589</td>
<td>bottom</td>
</tr>
<tr>
<td>9</td>
<td>0.469</td>
<td>bottom</td>
</tr>
<tr>
<td>10</td>
<td>0.723</td>
<td>bottom</td>
</tr>
</tbody>
</table>

We are testing to see if the average difference is different than 0.

\[
\bar{x}_{\text{diff}} = -0.545, \quad s = 8.89, \quad n = 200
\]

\[
T = \frac{\bar{X} - \mu}{SE} = \frac{-0.545 - 0}{8.89/\sqrt{200}} = -0.877
\]

\[
p\text{-value} = P(T < -0.877 \text{ or } T > 0.877) = 2 \times P(T < -0.877) = 2 \times 0.19 = 0.38
\]
Calculating Power - Two Means

- **Step 0:** Pick a meaningful effect size \( \delta \) and a significance level \( \alpha \)

- **Step 1:** Calculate the range of values for the point estimate beyond which you would reject \( H_0 \) at the chosen \( \alpha \) level.

- **Step 2:** Calculate the probability of observing a value from preceding step if the sample was derived from a population where \( \mu = \mu_{H_0} + \delta \)

Example - Salmon contamination

Mirex is a carcinogenic insecticide that is known to enter certain watersheds in runoff from agricultural fields. The EPA recommends that any fish caught in these areas be screened for mirex, with any values above 0.08 ppm being considering unsafe. Researchers want to test 150 randomly caught salmon from a river in the pacific northwest for potential organic contaminants. Based on previous research they expect the sample standard deviation of mirex to be around 0.05 ppm.

What would the power of a test be to detect an average concentration of mirex that is 0.01 ppm above the EPA’s limit?

Example - Reading

An educator believes that new reading activities for elementary school children will improve reading comprehension scores. She plans to randomly assign third graders to an 8-week program in which some will use these activities (18 students) and others will experience traditional teaching methods (20 students). At the end of the experiment, both groups will take a reading comprehension exam.

Previous studies using this exam have resulted in sample standard deviations of 8 points. What would the power of a test be to detect a 5 points improvement in the treatment group?
cases: 38 students
variable(s): (Dep) score - numerical, (Indep) treatment - categorical
test: Test of Two Means (unpaired)
  independence w/in groups: random assignment, 18, 20 both < 10% of all students
  independence btw groups: random assignment, different students in each condition
  sample size/skew: Assume neither distribution is extremely skewed
parameter of interest: $\mu_{new} - \mu_{trad}$
point estimate: $\bar{x}_{new} - \bar{x}_{trad}$
hypotheses: (two-tailed)
  $H_0 : \mu_{new} = \mu_{trad}$
  $H_A : \mu_{new} \neq \mu_{trad}$
effect size: $\delta = 5$