4 Announcements

- Friday 9/15/00 Lab in LSRC 247. Look over lab and bring questions.
- Reading in M&M to supplement the homework: “The power of the $t$-test,” p. 517-518. Good example of power calculation for a one-sample $t$-test.
- For Tuesday class:
  - Ch. 3, R&S
  - p. 515-517, M&M, Robustness of the $t$ procedures

4.1 Power of a Statistical Test

- What does “not significant” really mean?
  - positively demonstrating that a treatment had no effect
  - failing to demonstrate that a treatment does have an effect
- How many samples are needed to demonstrate a statistically significant difference in means?

4.2 Errors in a Statistical Test

The random sampling process can lead to two kinds of errors:

<table>
<thead>
<tr>
<th>Decision</th>
<th>Truth</th>
<th>$H_0$ true</th>
<th>$H_A$ true</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept $H_0$</td>
<td>correct decision, $1-\alpha$</td>
<td>Type II error = $\beta$</td>
<td></td>
</tr>
<tr>
<td>Reject $H_0$</td>
<td>Type I error = $\alpha$</td>
<td>correct decision, $1-\beta$</td>
<td></td>
</tr>
</tbody>
</table>

4.3 Implications of these errors: an illustration

- Suppose the EPA performs an hypothesis test to determine whether a chemical company is violating the law, such as:
  $H_0 : \mu \leq 3$ ppm; $H_A : \mu > 3$ ppm
- Suppose that data has been collected and that the null hypothesis is not rejected.
- Results:
  - No action is taken to reduce discharge of the pollutant.
  - A government report is printed that says, “Our statistical analysis has shown that the company is in compliance.”

- Perspectives:
  1. EPA
  2. Manager of the company producing the pollutant
  3. Nearby residents
  4. Conservation biologist

4.4 Power

- The chance of reporting a statistically significant difference when the treatment really produces an effect
- Power=1-Type II error=1 $- \beta$
- Defined with respect to a specific alternative hypothesis being considered.
- If an hypothesis test has power equal to 0.8, there is an 80% chance of actually reporting a statistically significant effect when one is present.
4.5 EPA PCB example

- Recall the EPA example from the Review handout. This time, consider a two-sided test. (Previously, we considered a one-sided test.)
- Recall: \( n = 30, \bar{Y} = 3.1 \text{ ppm}, s = 0.5 \text{ ppm}, \alpha = 0.01 \)
- Test \( H_0 : \mu = 3 \text{ vs. } H_A : \mu \neq 3. \)

\[
T = \frac{3.1 - 3}{\frac{s}{\sqrt{n}}} = 1.099. \tag{1}
\]

- Reject \( H_0 \) if \( |T| > t_{29,1-\alpha\over 2} = 2.76 \)
- Find the \((1 - \frac{\alpha}{2})^{th}\) quantile of a \( t \)-distribution from the command line in Splus:

\[
> qt(.995,29) \tag{1} \]

[1] 2.756386

- Thus, I can't reject the null hypothesis at \( \alpha = 0.01 \) and we don't have enough evidence to shut the plant down.

- Confidence interval for the population mean under \( H_0 \): 

\[
- |T| < 2.76
\]

- We can rewrite the above statement in terms of the population mean and get a 99% confidence interval,

\[
-2.76 < T < 2.76 \iff -2.76 < \frac{\bar{Y} - \mu}{s/\sqrt{n}} < 2.76 \iff \bar{Y} - 2.76 \frac{s}{\sqrt{n}} < \mu < \bar{Y} + 2.76 \frac{s}{\sqrt{n}} \iff 2.85 < \mu < 3.35
\]

4.5.1 Power calculation for EPA example

1. What are the implications of concluding that the company is in compliance (and can continue PCB discharge) when in fact it is not?
2. Specify an alternative hypothesis: Test \( H_0 : \mu = 3 \text{ vs. } H_A : \mu = 3.2 \).
3. Define acceptance region for \( H_0 \):

- Rejection Region is \( |T| > 2.76 \) for \( \alpha = 0.01 \)
- Acceptance Region is \( |T| < 2.76 \) under \( H_0 \)
4. Restate acceptance region in terms of \( \bar{Y} \) under the condition that \( H_0 : \mu = 3 \)

\[
-2.76 < T < 2.76 \iff -2.76 < \frac{\bar{Y} - \mu}{s/\sqrt{n}} < 2.76 \iff \mu - 2.76 \frac{s}{\sqrt{n}} < \bar{Y} < \mu + 2.76 \frac{s}{\sqrt{n}} \iff 2.75 < \bar{Y} < 3.25
\]

5. Calculate type II error, or \( \beta \), under \( H_A \):

\[
\beta = P(2.75 < \bar{Y} < 3.25 | H_A : \mu = 3.2) = P(\frac{2.75 - 3.2}{0.5/\sqrt{30}} < \frac{\bar{Y} - 3.2}{0.5/\sqrt{30}} < \frac{3.25 - 3.2}{0.5/\sqrt{30}}) = P(-4.93 < T < 0.55) = P(T < 0.55) - P(T < -4.93) \approx P(T < 0.55) = 0.71
\]

6. Power=1-\beta=29%

7. Conclusion: This test has only 29% power to detect an increase in PCB levels of 0.2 ppm. Thus, if this is the level of interest to regulators, it would be inappropriate to use the results of this test to conclude that the company is in compliance and can continue its operations.
4.6 Factors affecting the power (type II error) of a test

The ability to detect an effect (such as a non-zero difference in means) with a given level of confidence depends on:

1. size of the difference or treatment effect
   - How far is the truth from the null hypothesis?
   - It is harder to miss a big difference.
   - Power increases as the effect (the true differences) is larger.

2. sample size
   - The power of averaging extracts the signal from the noise.
   - Power increases as the sample size is increased.
   - An example: Studies with only a few samples that fail to reject null hypothesis may arrive at this result because the statistical procedure lacked the power to detect a treatment effect even though one existed.

3. inherent variability in the data
   - Noisy data hides the truth.
   - Power increases as the standard deviation of the observations is decreased.

4. the level (α) of the Type I error you are willing to tolerate
   - It is easier to find a difference if you take a bigger chance on a false positive.
   - Power increases as α increases.

4.7 Required sample size and power

- Test $H_0: \mu_D = 0$ versus $H_1: \mu_D \neq 0$
- Equal sample sizes.
- Useful approximate formula to calculate required sample size to achieve power of $1 - \beta$.

$$n = 2 \left( z_{1-\beta} + z_{1-\sigma/2} \right)^2 \left( \frac{\sigma}{\theta} \right)^2$$

- Use this formula only when $n \geq 10$ for each group.
- For $\alpha = .05$, power=80% in diazinon example ($\theta = 36.58$), we require 16 days of day and night measurements to detect a 1 standard deviation difference in diazinon readings.
5. Power = 1 - β = 10.3%
6. Conclusion: This test has insufficient power to detect an increase of 10 mg/m³ for the conditions stated above.

4.8 Another power example

Is there an increase in mean diazimon levels at night?

\[ H_0: \mu_D = 0 \text{ versus } H_A: \mu_D > 0 \text{ for } \alpha = 0.05 \]

\[ \mu_D = \text{mean increase in diazimon concentration at night} \]

\[ \mu_D = \text{mean increase in diazimon concentration during the day} \]

- Test: \( H_0: \mu_D = 0 \) versus \( H_A: \mu_D > 0 \) for \( \alpha = 0.05 \)
- \( s = 5.58, n = 11 \) as before.
- \( \bar{D} = 11.812 \)

- Reject \( H_0 \) if \( T > 1.1068 \)

- Conclusion: We fail to reject \( H_0 \) at \( \alpha = 0.05 \) and conclude that based on this sample, there is not strong evidence of an increase in diazimon levels at night.

4.8.1 What is the power of the test to detect an increase of 10 mg/m³?

1. Specify an alternative hypothesis: \( H_A: \mu_D = 10 \)
2. Define acceptance region for \( H_A \):
   - Rejection Region is \( T > 1.1812 \) for \( \alpha = 0.05 \)
   - Acceptance Region is \( T < 1.1812 \)
3. Restate acceptance region in terms of \( \bar{D} \):
   - \( \bar{D} - \frac{10}{\sqrt{11}} \)
   - \( \bar{D} - \frac{10}{\sqrt{11}} < 1.1812 \)
   - \( \bar{D} < \frac{10}{\sqrt{11}} \)
4. Calculate type II error, or \( \beta \), under \( H_A \):
   - \( \beta = P(\bar{D} < 10) = 0.05 \)

5. Power = 1 - \( \beta \) = 95%