Inference for Means

Sections 6.4, 6.5, 6.6, 6.10, 6.11, 6.12, 6.13
• t-distribution
• Formulas for standard errors
• t based inference

Central Limit Theorem
For a sufficiently large sample size, the distribution of sample statistics for a mean or a proportion is normal

• For means, “sufficiently large” is often $n \geq 30$
• If the data are normal, smaller $n$ will be sufficient
• If the data are skewed and/or have outliers, $n$ may have to be much higher than 30
• www.lock5stat.com/statkey

Standard Error Formulas

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Distribution</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion</td>
<td>Normal</td>
<td>$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$</td>
</tr>
<tr>
<td>Difference in Proportions</td>
<td>Normal</td>
<td>$\frac{\hat{p}_1-\hat{p}_2}{\sqrt{n_1 \cdot \hat{p}_1(1-\hat{p}_1) \cdot n_2 \cdot \hat{p}_2(1-\hat{p}_2)}}$</td>
</tr>
<tr>
<td>Mean</td>
<td>$t$, df = $n - 1$</td>
<td>$\frac{s}{\sqrt{n}}$</td>
</tr>
<tr>
<td>Difference in Means</td>
<td>$t$, df = $\min(n_1, n_2) - 1$</td>
<td>$\sqrt{\frac{s^2_1 + s^2_2}{n_1 + n_2 - 2}}$</td>
</tr>
</tbody>
</table>

SE of a Mean
The standard error for a sample mean can be calculated by

$$SE = \frac{\sigma}{\sqrt{n}}$$

Standard Deviation
The standard deviation of the population is
a) $\sigma$
b) $s$
c) $\frac{\sigma}{\sqrt{n}}$

The standard deviation of the sample is
a) $\sigma$
b) $s$
c) $\frac{\sigma}{\sqrt{n}}$
Standard Deviation

The standard deviation of the sample mean is

a) $\sigma$

b) $s$

c) $\frac{\sigma}{\sqrt{n}}$

---

Degrees of Freedom

• The $t$-distribution is characterized by its degrees of freedom ($df$)

• Degrees of freedom are based on sample size
  • Single mean: $df = n - 1$
  • Difference in means: $df = \min(n_1, n_2) - 1$
  • Correlation: $df = n - 2$

• The higher the degrees of freedom, the closer the $t$-distribution is to the standard normal

---

Aside: William Sealy Gosset

---

Question of the Day

How do pheromones in female tears affect men?
Pheromones in Tears

- Tears were collected from human females (they watched a sad movie in isolation)
- Cotton pads were created that had either real female tears or a salt solution that had been dripped down the same female’s face
- 50 men had a pad attached to their upper lip twice, once with tears and once without, order randomized. (matched pairs design!)
- Many variables were measured; we will look first at testosterone level


Statistics: Unlocking the Power of Data

Matched Pairs

- For a matched pairs experiment, we look at the differences for each pair, and do analysis on this one quantitative variable
- Inference for a single mean (mean difference)

Statistics: Unlocking the Power of Data

Pheromones in Tears

- The average difference in testosterone levels between tears and no tears was -21.7 pg/ml
- The average level before sniffing the pads (tears or saline) was 155 pg/ml
- The standard deviation of these differences was 46.5 pg/ml
- The sample size was 50 men

\[ \bar{d} = -21.7 \]
\[ s_d = 46.5 \]
\[ n = 50 \]

\( *pg = \) picogram = 0.001 nanogram = \( 10^{-12} \) gram

Statistics: Unlocking the Power of Data

Pheromones in Tears: Test

1. State hypotheses:
   \[ H_0: \mu_d = 0 \]
   \[ H_1: \mu_d < 0 \]
   \[ \bar{d} = -21.7 \]
   \[ s_d = 46.5 \]
   \[ n = 50 \]
2. Check condition:
   \[ n = 50 \geq 30 \]
3. Calculate standard error:
   \[ SE \approx \frac{s_d}{\sqrt{n}} = \frac{46.5}{\sqrt{50}} = 6.58 \]
4. Calculate t-statistic:
   \[ t = \frac{\bar{d} - \mu_0}{SE} = \frac{-21.7}{6.58} = -3.3 \]
5. Compute p-value:
   \[ df = 49, \text{ p-value} = 0.0009 \]

5. Interpret in context:
   This provides strong evidence that female tears decrease testosterone levels in men, on average.

Statistics: Unlocking the Power of Data
Pheromones in Tears: CI

1. Check conditions: $n = 50 \geq 30 \checkmark$
   $\bar{X}_d = -21.7$
   $SE = 6.58$

2. Find $t^*$: $t^* = 2$

3. Compute the confidence interval: $\bar{X}_d \pm t^* \times SE$
   $-21.7 \pm 2 \times 6.58 = (-34.86, -8.54)$

4. Interpret in context:
   We are 95% confident that female tears on a cotton pad on a man’s upper lip decrease testosterone levels between 8.54 and 34.86 pg/ml, on average.

Pheromones in Tears

Can we conclude that something in the tears (as opposed to just saline trickled down the face) causes this decrease in testosterone levels?

a) Yes  

b) No

Sexual Attraction

- They also had 24 men rate faces on two attributes: sad and sexually arousing
- For sexual arousal, the mean was 439 VAS (visual-analog scale) for tears and 463 VAS for saline
- The standard deviation of the differences was 47 VAS
- Is this evidence that tears change sexual arousal ratings?  
  (a) Yes  (b) No

Give a 90% CI.

Sexual Arousal: Test

- What if we had (accidentally) ignored the paired structure of the data, and analyzed it as two separate groups?
- Two separate groups: one quantitative and one categorical (look at difference in means)
- Paired data: one quantitative (differences), look at mean difference
- How does this affect inference?

Testosterone Revisited
Paired vs Unpaired

If we were to mistakenly analyze this data as two separate groups rather than paired data, we would expect the sample statistic to

a) Increase
b) Decrease
c) Not change

Statistics: Unlocking the Power of Data

If we were to mistakenly analyze this data as two separate groups rather than paired data, we would expect the standard error to

a) Increase
b) Decrease
c) Not change

Statistics: Unlocking the Power of Data

If we were to mistakenly analyze this data as two separate groups rather than paired data, we would expect the p-value to

a) Increase
b) Decrease
c) Not change

Statistics: Unlocking the Power of Data

If we were to mistakenly analyze this data as two separate groups rather than paired data, we would expect the width of the confidence interval to

a) Increase
b) Decrease
c) Not change

Statistics: Unlocking the Power of Data

Paired Test Unpaired

\begin{align*}
H_0 & : \mu_1 = \mu_2 \\
H_A & : \mu_1 < \mu_2
\end{align*}

\begin{align*}
\bar{x}_1 &= -21.7 \\
\bar{x}_2 &= 46.5 \\
n &= 50
\end{align*}

\begin{align*}
SE &= \sqrt{\frac{\bar{x}_1^2}{n_1} + \frac{\bar{x}_2^2}{n_2}} \\
&= \sqrt{\frac{46.5^2}{50} + \frac{101.8^2}{50}} \\
&= \sqrt{132.66 + 154.34} \\
&= 6.58
\end{align*}

\begin{align*}
t &= \frac{\bar{x}_1 - \bar{x}_2}{SE} \\
&= \frac{-21.7 - 46.5}{6.58} \\
&= -3.3
\end{align*}

df = 49
p-value = 0.009

This provides strong evidence that female tears decrease testosterone levels in men.

Paired

\begin{align*}
\text{Confidence Interval} & : \text{statistic} \pm t^* \times SE \\
& = (-21.7 \pm 2 \times 6.58) \\
& = (-34.86, -8.54)
\end{align*}

Unpaired

\begin{align*}
&= \pm 2 \times
\end{align*}

95%: statistic \pm 2 \times SE

\text{Confidence Interval}
Paired vs Unpaired

- What if we had ignored the paired structure of the data, and analyzed it as two separate groups? How does this affect inference?

- If a study is paired, analyze the differences!

To Do

- Read Sections 6.4, 6.5, 6.6, 6.10, 6.11, 6.12, 6.13
- Do HW 6 (due Friday, 11/6)

STAT 250
Dr. Kari Lock Morgan

Intervals and Tests

Section 4.5
- Connecting intervals and tests

Bootstrap and Randomization Distributions

<table>
<thead>
<tr>
<th>Bootstrap Distribution</th>
<th>Randomization Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our best guess at the distribution of sample statistics</td>
<td>Our best guess at the distribution of sample statistics, if H₀ were true</td>
</tr>
<tr>
<td>Centered around the observed sample statistic</td>
<td>Centered around the null hypothesized value</td>
</tr>
<tr>
<td>Simulate sampling from the population by resampling from the original sample</td>
<td>Simulate samples assuming H₀ were true</td>
</tr>
</tbody>
</table>

- Big difference: a randomization distribution assumes H₀ is true, while a bootstrap distribution does not

Which Distribution?

- Let μ be the average amount of sleep college students get per night. Data was collected on a sample of students, and for this sample $x = 6.7$ hours.

- A bootstrap distribution is generated to create a confidence interval for μ, and a randomization distribution is generated to see if the data provide evidence that μ > 7.

- Which distribution below is the bootstrap distribution?

Which Distribution?

- Intro stat students are surveyed, and we find that 152 out of 218 are female. Let $p$ be the proportion of intro stat students at that university who are female.

- A bootstrap distribution is generated for a confidence interval for $p$, and a randomization distribution is generated to see if the data provide evidence that $p > 1/2$.

- Which distribution is the randomization distribution?
Intervals and Tests

- A confidence interval represents the range of plausible values for the population parameter.
- If the null hypothesized value IS NOT within the CI, it is not a plausible value and should be rejected.
- If the null hypothesized value IS within the CI, it is a plausible value and should not be rejected.

Both Father and Mother

“Does a child need both a father and a mother to grow up happily?”

- Let \( p \) be the proportion of adults aged 18-29 in 2010 who say yes. A 95% CI for \( p \) is (0.487, 0.573).
- Testing \( H_0: p = 0.5 \) vs \( H_a: p \neq 0.5 \) with \( \alpha = 0.05 \), we have:
  a) Reject \( H_0 \)
  b) Do not reject \( H_0 \)
  c) Reject \( H_a \)
  d) Do not reject \( H_a \)


Intervals and Tests

- Confidence intervals are most useful when you want to estimate population parameters.
- Hypothesis tests and p-values are most useful when you want to test hypotheses about population parameters.
- Confidence intervals give you a range of plausible values; p-values quantify the strength of evidence against the null hypothesis.

Interval, Test, or Neither?

Is the following question best assessed using a confidence interval, a hypothesis test, or is statistical inference not relevant?

On average, how much more do adults who played sports in high school exercise than adults who did not play sports in high school?

- a) Confidence interval
- b) Hypothesis test
- c) Statistical inference not relevant
**Interval, Test, or Neither?**

Is the following question best assessed using a confidence interval, a hypothesis test, or is statistical inference not relevant?

*Do a majority of adults take a multivitamin each day?*

- a) Confidence interval
- b) Hypothesis test
- c) Statistical inference not relevant

---

**Interval, Test, or Neither?**

Is the following question best assessed using a confidence interval, a hypothesis test, or is statistical inference not relevant?

*Did the Penn State football team score more points in 2014 or 2013?*

- a) Confidence interval
- b) Hypothesis test
- c) Statistical inference not relevant

---

**To Do**

- HW Intervals & Tests (due Friday, 11/6)