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

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{n(1-p)}{n}} )</td>
</tr>
<tr>
<td>Difference in Proportions</td>
<td>Normal</td>
<td>( \sqrt{\frac{n_1(1-p_1) + n_2(1-p_2)}{n_1 + n_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>( \frac{s_1}{\sqrt{n_1}} + \frac{s_2}{\sqrt{n_2}} )</td>
</tr>
<tr>
<td>Correlation</td>
<td>t, df = n - 2</td>
<td>( \frac{1-r^2}{\sqrt{n-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 sample is

a) \( \sigma \)
b) \( s \)
c) \( \frac{\sigma}{\sqrt{n}} \)

The standard deviation of the sample mean is

a) \( \sigma \)
b) \( s \)
c) \( \frac{\sigma}{\sqrt{n}} \)
Statistics: Unlocking the Power of Data

**t-distribution**

- For quantitative data, we use a t-distribution instead of the normal distribution.
- This arises because we have to estimate the standard deviations.
- The t distribution is very similar to the standard normal, but with slightly fatter tails (to reflect the uncertainty in the sample standard deviations).

**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**

Aside:

Normality Assumption

- The t-distribution assumes the data come from a normal distribution.
- Note: this assumption is about the original data, not the distribution of the statistic.
- For large sample sizes (\( n \gtrsim 30 \)) the t-distribution is still a good approximation.
- For small sample sizes (\( n < 30 \)), we can only use the t-distribution if the distribution of the data is approximately symmetric and bell-shaped.

Normality Assumption

- One small problem: for small sample sizes, it is very hard to tell if the data actually comes from a normal distribution!
Small Samples

• If sample sizes are small, only use the \( t \)-distribution if the data looks reasonably symmetric and does not have any extreme outliers.
• Even then, remember that it is just an approximation!
• In practice/life, if sample sizes are small, you should just use simulation methods (bootstrapping and randomization)

Mental Muscle

• Participants were asked to either perform actual arm pointing motions or to mentally imagine equivalent arm point motions
• They had to complete the motion multiple times, and the time in seconds to complete the motions was recorded
• Is there a difference in average time for actual movement and mental movements?

Descriptive Statistics

Test for Difference in Means

\[
t = \frac{\text{statistic} - \text{null}}{\text{SE}}
\]

\[
\text{null} = 0
\]

\[
\bar{x}_m - \bar{x}_a = 7.338 - 7.162 = 0.176
\]

\[
SE = \sqrt{\frac{s_m^2}{n_m} + \frac{s_a^2}{n_a}} = \sqrt{\frac{1.218^2}{8} + \frac{0.703^2}{8}} = 0.5
\]

\[
t = \frac{0.176 - 0}{0.5} = 0.352
\]

Conclusion

The p-value is 0.8236. Our conclusion is

a) Reject \( H_0 \); we have evidence that actual and mental movement take different amounts of time
b) Do not reject \( H_0 \); we do not have evidence that actual mental movement take different amounts of time
c) Reject \( H_0 \); we do not have evidence that actual and mental movement take different amounts of time
d) Do not reject \( H_0 \); we have evidence that actual mental movement take different amounts of time
### Fatigue

- Participants then developed muscle fatigue by holding a heavy weight out horizontally as long as they could.
- After becoming fatigued, the same experiment was repeated.
- Is there a difference between actual and mental movements after muscle fatigue?

### Interval for Difference in Means

- Give a 95% confidence interval for the difference in average time \( \text{prefatigue} \).

\[
\text{sample statistic} \pm t^* \times SE
\]

### Interval with t-distribution

- \( t^* \) comes from t-distribution. 95% interval:

\[
\pm \frac{0.176}{2.364} \times 0.5 = (-1.01, 1.36)
\]

### Your Turn!

Is the difference significant?

- a) Yes
- b) No

Give a 99% interval.
Matched Pairs
- Does the actual movement take longer after the muscle has been fatigued?
- This is a matched pairs design! (Everyone has a value pre fatigue and post fatigue)
- Compute the differences for each person; post fatigue time – pre fatigue time
- New variable measures the additional amount of time needed after fatigue.

Inference
We now do inference for
a) single mean
b) difference in means
c) single proportion
d) difference in proportions
e) correlation

Test for a Mean

\[ H_0 : \mu = 0 \]
\[ H_a : \mu > 0 \]
\[ \bar{x} = 0.875 \]
\[ SE = \frac{s}{\sqrt{n}} = \frac{1.146}{\sqrt{8}} = 0.405 \]
\[ t = \frac{\text{statistic} - \text{null}}{SE} = \frac{0.875 - 0}{0.405} = 2.16 \]

Conclusion
The p-value is 0.034. Our conclusion is...

a) Reject \( H_0 \); we have evidence that actual movement takes longer after fatigue
b) Do not reject \( H_0 \); we do not have evidence that actual movement takes longer after fatigue
c) Reject \( H_0 \); we do not have evidence that actual movement takes longer after fatigue
d) Do not reject \( H_0 \); we have evidence that actual movement takes longer after fatigue
**Conclusion**

Can we conclude that fatigue causes the increase in actual movement time?

a) Yes
b) No

**Interval for a Mean**

- Give a 95% confidence interval for amount of additional time the actual movement takes after fatigue.

\[
\text{sample statistic} \pm t^* \times SE
\]

\[
0.875 \quad 2.364 \quad 0.405
\]

\[
0.875 \pm 2.364 \times 0.405 = (-0.08, 1.83)
\]

**Your Turn!**

- Perform the same test, but for the mental movement. Does the time needed to complete mental movement change after fatigue?

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>8</td>
</tr>
<tr>
<td>Mean</td>
<td>-1.237</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.539</td>
</tr>
</tbody>
</table>

a) Yes
b) No

- Give a confidence interval for the difference.

**To Do**

- Read Sections 6.4, 6.5, 6.6, 6.10, 6.11, 6.12, 6.13
- Do HW 6b (due Friday, 4/10)