Comparing Two Groups
T-Test

Business Intelligence/Analytics
Mike DeWitt
Institutional Research
August 2018
Am I significant???

https://xkcd.com/539/
Are these two things different?

Differences in means (e.g. GPA 3.2 vs 3.34)
Are these two things different?

Differences in means (e.g. GPA 3.2 vs 3.34)

Difference in proportions (e.g. 50% vs 72%)
Are these two things different?

Differences in means (e.g. GPA 3.2 vs 3.34)
Difference in proportions (e.g. 50% vs 72%)
Differences in distributions (e.g. counts of each race per group)
There’s a test for that

Statistics provides us with several tools
But the first question is what kind of data do we have?
Use your data as your guide

- All Variables
  - Numerical
    - Continuous
    - Discrete
  - Categorical
    - Regular Categorical
    - Ordinal
First Question-What is the “treatment”
First Question - What is the “treatment”

Categorical

Numeric
Second Question-What is the “response”

Continuous
Ratios
Counts
Now we can choose a test...

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Number of Groups</th>
<th>Response</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical</td>
<td>2</td>
<td>Continuous</td>
<td>T-test</td>
</tr>
<tr>
<td>Categorical</td>
<td>&gt;2</td>
<td>Continuous</td>
<td>ANOVA</td>
</tr>
<tr>
<td>Categorical</td>
<td>2</td>
<td>Ratio</td>
<td>Proportion Test</td>
</tr>
<tr>
<td>Categorical</td>
<td>&gt;2</td>
<td>Counts</td>
<td>Chi-Square</td>
</tr>
<tr>
<td>Continuous</td>
<td>&gt;=2</td>
<td>Continuous</td>
<td>Regression</td>
</tr>
</tbody>
</table>
Now we can choose a test...

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Number of Groups</th>
<th>Response</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical</td>
<td>2</td>
<td>Continuous</td>
<td>T-test</td>
</tr>
<tr>
<td>Categorical</td>
<td>&gt;2</td>
<td>Continuous</td>
<td>ANOVA</td>
</tr>
<tr>
<td>Categorical</td>
<td>2</td>
<td>Ratio</td>
<td>Proportion Test</td>
</tr>
<tr>
<td>Categorical</td>
<td>&gt;2</td>
<td>Counts</td>
<td>Chi-Square</td>
</tr>
<tr>
<td>Continuous</td>
<td>&gt;=2</td>
<td>Continuous</td>
<td>Regression</td>
</tr>
</tbody>
</table>
Now we can choose a test...

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Number of</th>
<th>Response</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical</td>
<td>2</td>
<td>Continuous</td>
<td>T-test</td>
</tr>
<tr>
<td>Categorical</td>
<td>&gt;2</td>
<td>Continuous</td>
<td>ANOVA</td>
</tr>
<tr>
<td>Categorical</td>
<td>2</td>
<td>Ratio</td>
<td>Proportion Test</td>
</tr>
<tr>
<td>Categorical</td>
<td>&gt;2</td>
<td>Counts</td>
<td>Chi-Square</td>
</tr>
<tr>
<td>Continuous</td>
<td>&gt;=2</td>
<td>Continuous</td>
<td>Regression</td>
</tr>
</tbody>
</table>

**ALWAYS EXCEPT…**

Most of these tests assume
- Errors are independent and identically distributed
The birth of the t-test

William Sealy Gosset
The birth of the t-test
The birth of the t-test

Needed to compare means of Two Treatments with a Continuous Response

E.g. Alcohol content between two batches
Comparing two (unpaired) means

Comparing Two Independent Groups

Typically what we are interested in...

E.g. Grades between people who took class A vs Class B
Comparing two (paired) means

Comparing The Same Group Twice

Same people see both treatments

E.g. Comparing quotes from a mechanic on the same set of cars
What does the test tell us…

Question: Are the means of these two groups the same?

Confidence Intervals

P-Value
Confidence Intervals

Repeating this experiment many times, 95%* of the confidence intervals calculated will contain the true population mean
Multiplier * Weighted Avg of the Spread

Avg Group 1 - Avg Group 2

Lower

Upper
Confidence Intervals

\[ \text{Avg}_1 - \text{Avg}_2 \pm 1.96 \times s_p \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \]

Avg\(_1\) = Average of Group 1  
Avg\(_2\) = Average of Group 2  
s\(_p\) = Pooled standard deviation  
n\(_1\) = Number of Samples in Group 1  
n\(_2\) = Number of Samples in Group 2

If *Zero is included* in the interval then there is not evidence of a difference
P-Values

A **P-Value** is the probability of observing a result **as or more extreme** than the one I found if the experiment is **repeated many times**.
P-Value Thresholds

P-Value thresholds are arbitrary

By convention $p < 0.05$ or $p < 0.10$ is acceptable

Be sure to state what level you use!
Let’s Practice!

https://www.youtube.com/watch?v=-yZ97arTPGU
P-Value Thresholds

“The value for which P=0.05, or 1 in 20, is 1.96 or nearly 2; it is convenient to take this point as a limit in judging whether a deviation ought to be considered significant or not. “ – RA Fischer
P-Value Thresholds

One Tail P Value (Higher or Lower)

\[ \mu_1 > \mu_2 \text{ or } \mu_1 < \mu_2 \]

Two Tail P Value (Mean Values are Different)

\[ \mu_1 \neq \mu_2 \]