Comparing two population means

- Four models
  - Independent Sampling
    - Large Sample or known variances
      - Z-test
    - The 2 population variances are equal
      - Pooled variance t-test
    - The 2 population variances are unequal
      - t-test for unequal variances
  - Dependent Sampling
    - Matched Pairs t-test

Independent Sampling

Population 1
- \( \mu_1 \) , \( \sigma_1 \)
- \( n_1 \)
- \( \bar{X}_1, s_1 \)

Population 2
- \( \mu_2 \) , \( \sigma_2 \)
- \( n_2 \)
- \( \bar{X}_2, s_2 \)

Dependent Sampling

Population
- \( n \)
- Measurement 1
- Measurement 2
- \( \bar{X}_d, s_d \)

Difference of Two Population means

- \( \bar{X}_1 - \bar{X}_2 \) is Random Variable
- \( \bar{X}_1 - \bar{X}_2 \) is a point estimator for \( \mu_1 - \mu_2 \)
- The standard deviation is given by the formula
  \[ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \]
- If \( n_1 \) and \( n_2 \) are sufficiently large, \( \bar{X}_1 - \bar{X}_2 \) follows a normal distribution.

Difference between two means – large sample Z test

- If both \( n_1 \) and \( n_2 \) are over 30 and the two populations are independently selected, this test can be run.
- Test Statistic:
  \[ Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \]
Example 1

- Are larger houses more likely to have pools?
- The housing data square footage (size) was split into two groups by pool (Y/N).
- Test the hypothesis that the homes with pools have more square feet than the homes without pools. Let \( \alpha = .01 \)

**EXAMPLE 1 - Design**

\[ H_0: \mu_1 \leq \mu_2 \quad H_a: \mu_1 > \mu_2 \]
\[ H_0: \mu_1 - \mu_2 \leq 0 \quad H_a: \mu_1 - \mu_2 > 0 \]
\[ \alpha = .01 \]
\[ Z = (\bar{X}_1 - \bar{X}_2) / (\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}) \]
\( H_0 \) is rejected if \( Z > 2.326 \)

**EXAMPLE 1 - Data**

<table>
<thead>
<tr>
<th>Population 1</th>
<th>Population 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size with pool</td>
<td>Size without pool</td>
</tr>
<tr>
<td>Sample size = 130</td>
<td>Sample size = 95</td>
</tr>
<tr>
<td>Sample mean = 26.25</td>
<td>Sample mean = 23.04</td>
</tr>
<tr>
<td>Standard Dev = 6.93</td>
<td>Standard Dev = 4.55</td>
</tr>
</tbody>
</table>

**EXAMPLE 1 - p-value method**

- Using Technology
- Reject Ho if the p-value < \( \alpha \)

<table>
<thead>
<tr>
<th>Sq ft with pool</th>
<th>Sq ft no pool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>26.25</td>
</tr>
<tr>
<td>Std Dev</td>
<td>6.93</td>
</tr>
<tr>
<td>Observations</td>
<td>130</td>
</tr>
<tr>
<td>Hypothesized Mean Difference</td>
<td>0</td>
</tr>
<tr>
<td>( Z )</td>
<td>4.19</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0000137</td>
</tr>
</tbody>
</table>

**EXAMPLE 1 - Results/Decision**

- Decision: Reject Ho
- Conclusion: Homes with pools have more mean square footage.
Pooled variance t-test

- To conduct this test, three assumptions are required:
  - The populations must be normally or approximately normally distributed (or central limit theorem must apply).
  - The sampling of populations must be independent.
  - The population variances must be equal.

Pooled Sample Variance and Test Statistic

- Pooled Sample Variance:
  \[ s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \]

- Test Statistic:
  \[ t = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \]
  \[ df = n_1 + n_2 - 2 \]

EXAMPLE 2

- A recent EPA study compared the highway fuel economy of domestic and imported passenger cars.
- A sample of 12 imported cars revealed a mean of 35.76 mpg with a standard deviation of 3.86.
- A sample of 15 domestic cars revealed a mean of 33.59 mpg with a standard deviation of 2.16 mpg.
- At the .05 significance level can the EPA conclude that the mpg is higher on the imported cars? (Let subscript 2 be associated with domestic cars.)

- \[ \alpha = .05 \]
- \[ H_0: \mu_1 \leq \mu_2 \quad H_a: \mu_1 > \mu_2 \]
- \[ t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \]
- \[ df = n_1 + n_2 - 2 \]
- \[ t = (1.85) \]
- \[ H_0 \text{ is rejected. Imports have a higher mean mpg than domestic cars.} \]

EXAMPLE 2

- \[ H_0: \mu_1 \leq \mu_2 \quad H_a: \mu_1 > \mu_2 \]
- \[ \alpha = .05 \]
- \[ t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \]
- \[ df = n_1 + n_2 - 2 \]
- \[ t = 1.85 \] \[ H_0 \text{ is rejected. Imports have a higher mean mpg than domestic cars.} \]
Using Technology

- Decision Rule: Reject $H_0$ if p-value $< \alpha$
- Megastat: Compare Two Independent Groups
- Use Equal Variance or Unequal Variance Test
- Use Original Data or Summarized Data

Megastat Result – Equal Variances

<table>
<thead>
<tr>
<th>import</th>
<th>domestic</th>
</tr>
</thead>
<tbody>
<tr>
<td>35.76</td>
<td>33.59</td>
</tr>
<tr>
<td>3.66</td>
<td>2.16</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
</tr>
</tbody>
</table>

Megastat Result – Unequal Variances

<table>
<thead>
<tr>
<th>import</th>
<th>domestic</th>
</tr>
</thead>
<tbody>
<tr>
<td>35.76</td>
<td>33.59</td>
</tr>
<tr>
<td>3.66</td>
<td>2.16</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
</tr>
</tbody>
</table>

Hypothesis Testing - Paired Observations

- Independent samples are samples that are not related in any way.
- Dependent samples are samples that are paired or related in some fashion.
  - For example, if you wished to buy a car you would look at the same car at two (or more) different dealerships and compare the prices.
  - Use the following test when the samples are dependent:

Hypothesis Testing Involving Paired Observations

$$ t = \frac{\bar{X}_d - \mu_d}{s_d / \sqrt{n}} $$

- where $\bar{X}_d$ is the average of the differences
- $s_d$ is the standard deviation of the differences
- $n$ is the number of pairs (differences)

EXAMPLE 3

- An independent testing agency is comparing the daily rental cost for renting a compact car from Hertz and Avis.
- A random sample of 15 cities is obtained and the following rental information obtained.
- At the .05 significance level can the testing agency conclude that there is a difference in the rental charged?
Example 3 - continued

• Data for Hertz
  \[ \bar{X}_1 = 46.67 \]
  \[ s_x = 5.23 \]

• Data for Avis
  \[ \bar{X}_2 = 44.87 \]
  \[ s_x = 5.62 \]

By taking the difference of each pair, variability (measured by standard deviation) is reduced.

\[ \bar{X}_d = 1.80 \]
\[ s_d = 2.513 \]
\[ n = 15 \]

There is a difference in mean price for compact cars between Hertz and Avis. Avis has lower mean prices.

Megastat Output – Example 3

Test for Equal Variances

For the two tail test, the test statistic is given by:

\[ F = \frac{S_1^2}{S_2^2} \]

\[ S_1^2 \] and \[ S_2^2 \] are the sample variances for the two populations.

There are 2 sets of degrees of freedom:
\( n_1 - 1 \) for the numerator, \( n_2 - 1 \) for the denominator.
EXAMPLE 4

A stockbroker at brokerage firm, reported that the mean rate of return on a sample of 10 software stocks was 12.6 percent with a standard deviation of 4.9 percent.

The mean rate of return on a sample of 8 utility stocks was 10.9 percent with a standard deviation of 3.5 percent.

At the .05 significance level, can the broker conclude that there is more variation in the software stocks?

Test Statistic depends on Hypotheses

- Hypotheses
  - $H_0 : \sigma_1 \geq \sigma_2$
  - $H_1 : \sigma_1 < \sigma_2$
  - $H_0 : \sigma_1 \leq \sigma_2$
  - $H_1 : \sigma_1 > \sigma_2$
  - $H_0 : \sigma_1 = \sigma_2$
  - $H_1 : \sigma_1 \neq \sigma_2$

- Test Statistic
  - $F = \frac{s_1^2}{s_2^2}$ use $\alpha$ table
  - $F = \frac{s_1^2}{s_2^2}$ use $\alpha$ table
  - $F = \frac{\max(s_1^2, s_2^2)}{\min(s_1^2, s_2^2)}$ use $\alpha / 2$ table

Excel Example

- Using Megastat – Test for equal variances under two population independent samples test and click the box to test for equality of variances.
- The default p-value is a two-tailed test, so take one-half reported p-value for one-tailed tests.
- Example – Domestic vs Import Data
  - $\alpha = .10$
  - Reject $H_0$ means use unequal variance t-test
  - FTR $H_0$ means use pooled variance t-test

Excel Output

- F-test for equality of variance
  - 14.994 variance: import
  - 4.654 variance: domestic
  - $F = 3.20$, $p = .0438$
  - p-value <.10, Reject Ho

- Use unequal variance t-test to compare means.

Compare Two Means Flowchart