Lecture 19 (MWF) Independent sample t-test

Suhasini Subba Rao
Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations

**Review: Example 1**

An airline wants to evaluate the depth perception of pilots over the age of fifty. A random sample of pilots over the age of fifty are used. The sample data of the pilots error is listed below.

<table>
<thead>
<tr>
<th>2.7</th>
<th>2.4</th>
<th>1.9</th>
<th>2.6</th>
<th>2.4</th>
<th>1.9</th>
<th>2.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2</td>
<td>2.5</td>
<td>2.3</td>
<td>1.8</td>
<td>2.5</td>
<td>2.0</td>
<td>2.2</td>
</tr>
</tbody>
</table>

(a) Construct a 95% CI for the mean error.

(b) The mean error for ‘young’ plots is 2, test the hypothesis that the mean error is larger for ‘older’ pilots.
Solution

(a) The average is 2.26 and the sample standard deviation is 0.27 sample standard error is $\frac{0.27}{\sqrt{14}} = 0.0778$.

To evaluate the sample variance we use $\frac{1}{14-1} \sum_{i=1}^{14} (X_i - \bar{X})^2$. There are 14 observations hence the z-transform is a t-distribution with 13 degrees of freedom.

By looking up in the tables we see that $t_{0.025}(13) = 2.16$ (we use this instead of $z_{0.025} = 1.96$). The 95% CI is

$$
\left[ 2.26 - 2.16 \times \frac{0.27}{\sqrt{14}}, 2.26 + 2.16 \times \frac{0.27}{\sqrt{14}} \right].
$$

(b) The hypotheses are $H_0 : \mu \leq 2$ against $H_A : \mu > 2$. We do a one-sided
test so we need \( t_{0.05}(13) = 1.771 \) The non-rejection region is:

\[
\left( -\infty, 2 + 1.771 \times \frac{0.27}{\sqrt{14}} \right] = (-\infty, 2.13].
\]

Therefore the rejection region is

\[
\left( 2 + 1.771 \times \frac{0.27}{\sqrt{14}}, \infty \right) = (2.13, \infty).
\]

Since the sample average \( \bar{X} = 2.26 \) lies in the rejection region there is enough evidence to reject the null and accept the alternative. That is there is evidence to suggest that ‘older’ pilots tend to make a larger error.
• Alternatively we can obtain the p-value

\[
t = \frac{\bar{X} - 2}{\frac{0.27}{\sqrt{14}}} = \frac{2.26 - 2}{\frac{0.27}{\sqrt{14}}} = 3.48.
\]

If we used Statcrunch or JMP we could find that the area to the right of 3.48 (since the alternative is pointing right) is 0.02%. If we only have tables available, since the test is done at the 5% level first look up \( t_{0.05}(13) = 1.771 \) in the tables. Now compare the t-transform with \( t_{0.05}(13) = 1.771 \), since 3.48 > 1.771, the p-value must be less than 5% (easiest seen with a plot). Therefore, there is evidence at the 5% level to reject the null.

• All these methods are equivalent.
Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations

**Solution using JMP output**

![Graph showing distributions and summary statistics]

Everything we need to do a test or construct a CI (of any type) can be found in the above output.
• The data does not deviate much from the \( x=y \) axis on the QQplot. This suggests that the distribution of pilot perception does not deviate much from normality. Therefore, even though the sample size is quite small, both the 95% confidence intervals and the tests will be accurate (i.e. we really will be 95% confident about the location of the mean and the p-values are really those percentages).

• Using this information we can construct, say, a 99% confident interval for the mean (by looking up t-tables with 13df and 0.5%)

\[ 2.26 \pm 3.01 \times 0.075. \]

• We can use this information to do the test \( H_0 : \mu \leq 2 \) against \( H_A : \mu > 2 \).

• For the purposes of an exam, I may only give half the information in the summary statistics and you need to deduce the rest.
Review: Example 2

A new reading program was being evaluated in a fourth grade elementary school. A random sample of 20 students were tested to determine reading speed. Each of their scores was given out of 100. A summary of the data is given below

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Sample average</th>
<th>Sample standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 20$</td>
<td>$X = 82.05$</td>
<td>$s = 10.88$</td>
</tr>
</tbody>
</table>

It was known that on average the reading speed is 84 (using the old reading programs). Test the hypothesis that the new reading program has resulted in a change of reading speed. Do the test at the 1% level.
Notice that in this example, we are only looking for changes in reading speed with the new reading program. I.e. does the program have any effect (regardless of whether the effect was positive or negative). Hence we are testing $H_0 : \mu = 84$ against $H_A : \mu \neq 84$.

The sample size is 20, which means we can tentatively assume that the central limit theorem has kicked in and the sample average $\bar{X}$ is close to normal, but we still need to take our results with a small pinch of salt.

The standard error in this example is $se = 10.88 / \sqrt{20} = 2.43$.

Since this is a two sided test, we need to look up $0.01/2$ in the $t$-tables with 19-degrees of freedom. We get 2.86. Based on this the non-rejection
Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations

region is

\[ [84 - 2.86 \times 2.43, 84 + 2.86 \times 2.43] = [77.1, 90.9]. \]

• Since \( \bar{X} = 82.05 \) lies inside this region, there is not enough evidence to reject the null. In other words, based on the standard error, 82.05 is too close to 84 for us to say that this new program has had any effect on the children’s reading ability.

• The \( t \)-transform is \( t = (82.05 - 84)/2.43 = -0.8 \). Using the \( t \)-distribution with 19df, the area to the LEFT (since this is the smallest area) of \(-0.8\) is 22\%, therefore the p-value is 44\%. This means that that 82.054 is close enough to 84 such that given the sample size there is no evidence that the reading program has changed results.
Comparing populations

Suppose I want to compare the heights of males and females at A&M.

• I can consider all males at Texas A&M as one population and all females at Texas A&M as another population.

• Question 1: Is the mean female height less than that of the male height?

• Question 2: What is the difference in the mean male and mean female heights.

• Suggestion: Compare the sample mean of the male heights with the sample mean of female heights and use to make inference about the heights of male and female A&M students.
Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations

**Plot of data**

In an old STAT651 class I collected data from the students. This included their gender and height. The histograms/boxplot and QQplot of their heights (for males and females) are given below.

![Histograms and QQplots](image)
The QQplot has the horizontal lines because everyone rounded their height to the nearest inch.
Summary of data

• Let $X$ be the height of a randomly selected female and $Y$ the height of randomly selected male. There are $n = 37$ females and $m = 27$ males in the samples. The sample mean for females is $\bar{X} = \frac{1}{37} \sum_{i=1}^{37} X_i = 5.45$ and sample mean for males is $\bar{Y} = \frac{1}{27} \sum_{i=1}^{27} Y_i = 5.92$. The sample standard deviations are $s_X = \sqrt{0.0484}$ and $s_Y = \sqrt{0.0758}$.

• Let $\mu_X$ be the female population mean height and $\mu_Y$ be the male population mean height.

• We are interested in the quantity $\mu_X - \mu_Y$. It will tell us how much larger, how small or whether the male and female heights are equal. Of course, we do not know that difference $\mu_X - \mu_Y$, and need to infer something about $\mu_X - \mu_Y$ from the samples.
• Intuitively it is obvious that to test whether $\mu_X$ and $\mu_Y$ are equal, we need to compare the sample averages $\bar{X}$ and $\bar{Y}$ and look at their difference $\bar{X} - \bar{Y}$. What can the differences in the sample means say about the differences in the true means that is $\mu_X - \mu_Y$ (population mean of females - population mean of males)?

• We would expect that population mean of females is less than population mean of males, in other words population mean of females - population mean of males to be less than zero. Hence we are interested in testing $H_0 : \mu_X - \mu_Y \geq 0$ against $H_A : \mu_X - \mu_Y < 0$.

• We also want know the magnitude of the difference, this means constructing a CI for the mean difference $\mu_X - \mu_Y$.

• Clearly if $\bar{X} - \bar{Y} > 0$ we would be unable to reject the null (why?? - remember $\bar{X} - \bar{Y}$ has to pointing in the same direction as the alternative).
• But if $\bar{X} - \bar{Y} < 0$, then we should use a statistical test.

• The question is how to make the comparison, what is the distribution of $\bar{X} - \bar{Y}$, we look at this now.
Objectives

• To build a confidence interval for the mean difference $\mu_X - \mu_Y$ (this will tell us where the mean difference lies).

• To test the hypothesis that $H_0 : \mu_X - \mu_Y \geq 0$ (mean female height and male height are the same or mean female height is greater than mean male height) against the alternative $H_A : \mu_X - \mu_Y < 0$.

• We can take the test further and test whether $H_0 : \mu_X - \mu_Y \geq -0.3$ against $H_A : \mu_X - \mu_Y < -0.3$.

This is essentially testing whether boys on average are more than 0.3 feet taller than girls.
• We will construct CIs for the difference between the sample means and do hypothesis test. We will rely on JMP output, but understand how all parts of it are calculated (we will be doing the cumbersome calculations).

• However, I will not be making you calculate standard errors and do complex calculations (this is the purpose of statistical software).
Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations

**The necessary ingredients**

- In order to do any test, ie. $H_0 : \mu_X - \mu_Y \geq 0$ against $H_A : \mu_X - \mu_Y < 0$
or to construct CI for $\mu_X - \mu_Y$ we need three magical ingredients:

  - The difference of the sample averages: $\bar{X} - \bar{Y}$.
  - The standard errors of $\bar{X} - \bar{Y}$ (this will turns out to be $\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$, where $\sigma_X$ is the standard deviation/variation in population $X$ and $\sigma_Y$ is the standard deviation/variation in population $Y$).
  - The sample sizes $m$ and $n$ are relatively large (ideally both over 25).
Assumptions and how to check them

- We have two random samples from two different populations $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ (sample size $n$ and $m$ respectively).

- Both samples are independent of each other and independent within the sample. For example the values $X_1, \ldots, X_n$ should have no influence on $Y_1, \ldots, Y_m$ and $X_1$ should not have any influence on $X_2, \ldots, X_n$.

  - Can you think of examples when this may not true?
    * It is likely that data which is observed over time, will be dependent (previous observation has an influence on the current).
  - Checking for independence can be difficult, however there are methods available to see whether there is dependence (i.e. In time series there is the Ljung-Box test).
Distribution of the difference of the sample means $\bar{X} - \bar{Y}$

- Suppose that $\bar{X}$ and $\bar{Y}$ are close to normal, then the difference of the averages has the following distribution:

$$\bar{X} - \bar{Y} \sim \mathcal{N} \left( \mu_X - \mu_Y, \left( \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} \right) \right).$$

where $\sigma_1$ is the standard deviation of $X$ (variability of population of $X$) and $\sigma_2$ is the standard deviation of $Y$ (variability of population $Y$).

- Important points:
  - The distribution is centered about $\mu_X - \mu_Y$, hence I am likely to draw close to $\mu_X - \mu_Y$. 

Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations
Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations

- How ‘close’ the sample means difference is to the true means is determined by the standard error which is $\sqrt{\left(\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}\right)}$.

The larger the sample sizes $n$ and $m$ the smaller the standard error (just like in the one sample case, where we deal with just one sample mean $\bar{X}$, which has standard error $\frac{\sigma^2}{n}$).

- The normality result depends (as usual) on two factors
  * How close the two data sets are to normality.
  * How large the sample size.

If the sample size is small then the data must be close to normal. This can be checked using QQplots.

- Therefore we can make a $Z$-transform of the difference $\bar{X} - \bar{Y}$

$$
\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \sim \mathcal{N}(0, 1).
$$
• In practice $\sigma_1$ and $\sigma_2$ is unknown and has to be estimated from the data.
• Instead estimate the two standard deviations, $s_1$ and $s_2$ from the data and plug them into

$$\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \rightarrow \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}$$

to give the standard error.
Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations

The distribution when the standard deviation is estimated

- Our t-transform is now

\[
\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}}
\]

- The distribution of the above is a classical problem in statistics called the ‘Behren-Fisher problem (you do not have to know about thi)s. Though the distribution of the above is not known exactly, it has an approximate t-distribution with a rather complicated number of degrees of freedom....

- Do not panic! JMP will always give you the degrees of freedom!
Returning to the height example: Assumptions

• Unless many of the students in the 651 class were related it is reasonable to assume that they are independent.

• To check whether the both the sample means $\bar{X}$ (female sample average) and $\bar{Y}$ (male sample average) are normal we look at the data (see next page).

• Description of plots: Since there was rounding in the heights given (which in actually make the estimates of the mean heights slightly biased), the data is not particularly normal. However, the sample sizes are relatively large (37 females and 27 males), the sample means $\bar{X}$ and $\bar{Y}$ will be close to normal. This means that it is fine to use the t-test.
Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations
Making QQplot by pooling the data

If the sample size in both groups are small (say 7 in each group), making a QQplot for each group is not informative. Instead, you can pool the data and make a QQplot of the residual for of the two groups combined. Go to Analyze - Fit Y by X (place data in appropriate places). You should get a dot plot of both groups. Select T-test and save - save residuals (if you believe the variances in each group are different then use standardized residuals). This combines the two groups by subtracting from each group the group means.

Residuals = $X_i - \bar{X}$; Standardized residuals = $(X_i - \bar{X})/s$. 
Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations

**The JMP output**

Below is the JMP output (remember for JMP to work you need to make sure that the X-factor is NOT set as a continuous variable) for the height data.
Interpreting the output

The output gives you all the information that you need to immediately do the test $H_0 : \mu_X - \mu_Y \geq 0$ against $H_A : \mu_X - \mu_Y < 0$ and constructing a 95% confidence interval for $\mu_X - \mu_Y$. We summarise these below:

- The df = 48.29

- The standard error

$$\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}} = \sqrt{\frac{0.0484}{37} + \frac{0.0758}{27}} = 0.064.$$  

Remember, the standard error is measuring the variability of the estimator
of the differences $\bar{X} - \bar{Y}$:

$$\sqrt{\frac{0.0484}{37}} + \sqrt{\frac{0.0758}{27}}$$

measures variability of average female $\bar{X}$ measures variability of average male $\bar{Y}$

• The difference in sample means is $-0.466$.

• The $t$-value for the test $H_0 : \mu_X - \mu_Y \geq 0$ against $H_A : \mu_X - \mu_Y < 0$ is $= (-0.466 - 0)/0.064 = -7.27$.

• Since the alternative is pointing to the left, we need to area to the left of -7.27, this is $\text{Prob} < 0.0001$ (in the output). Therefore, the p-value is less than 0.0001, which means at the 5% level there is sufficient evidence to suggest that males are taller than females.
• The 95% confidence for the difference in the means is $[-0.59, -0.33]$, this means with 95% confidence we believe the average difference lies between $[-0.59, -0.33]$. 
However, we may ask further questions about the difference between male and female height which is not immediately clear from the output, but with a little effort can be deduced.

- What if we have a feeling that the mean difference is over 0.3 feet. Is there evidence in the data that this could be true:

This means we are testing \( H_0 : \mu_X - \mu_Y \geq -0.3 \) against \( H_A : \mu_X - \mu_Y < -0.3 \), which we can test at the 5% level.

In other words we are seeing how plausible is it that we can get a sample mean difference of -0.466 feet when the global difference is greater than or equal to -0.3. If this chance/probability (the p-value) turns out to be too small (smaller than 5% if we use as the level 5%), then we reject the null.
Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations

- Suppose want to know where the actual mean difference lies, this means constructing a confidence interval.

The JMP output gives is a 95% confidence interval for the mean difference [-0.6, 0.34]. Suppose we want greater confidence in the interval, and require a 99% confidence interval.

To answer both questions (in the case of an exam) we can use the output but we also require the critical values for the t-distribution with 48.28df.

<table>
<thead>
<tr>
<th>probability</th>
<th>0.15</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t^*$</td>
<td>1.05</td>
<td>1.3</td>
<td>1.68</td>
<td>2.01</td>
<td>2.4</td>
<td>2.68</td>
</tr>
</tbody>
</table>

This table will always be given in an exam.
Confidence intervals at other levels and tests

• Just as in before to obtain the 99% confidence interval we use the critical value for 0.005 (remember 0.005=0.5%), from the above table we see that it is 2.68. This gives the 99% confidence interval:

\[
[-0.466 - 2.68 \times 0.064, -0.466 + 2.68 \times 0.064] = [-0.63, -0.29]
\]

which (of course) is wider than then 95% CI [-0.6,0.34].

• To do the test \( H_0 : \mu_X - \mu_Y \geq -0.3 \) against \( H_A : \mu_X - \mu_Y < -0.3 \) we make a t-transform

\[
t = \frac{-0.46 - (-0.3)}{0.064} = -2.5
\]

Since -2.5 lies between -2.68 and -2.4, the area to the left of -2.5 is between 0.5% to 1%. Thus the p-value is between 0.5% - 1%, thus
there is evidence to suggest that the mean difference between males and females is greater than 0.3 feet.
Choosing the sample sizes

• The standard error is

\[
\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}
\]

variability of $\bar{X}$  variability of $\bar{Y}$

(if the population standard deviations are known).

• As usual there are two factors which effect the size, the standard deviations $\sigma_1$ and $\sigma_2$ and the sample sizes $n$ and $m$. You **cannot** control the standard deviation (variability) of the population (unless you did something drastic, such as getting rid of the extremes in the population) but, often you **can** control the sample sizes.
Suppose that the standard deviations of both populations are the same, then the standard error is \( \sigma \sqrt{\frac{1}{n} + \frac{1}{m}} \). Suppose \( n + m = 100 \), how to choose the sample sizes to make this as small as possible?

Extreme examples are:

- If \( n = 50 \) and \( m = 50 \), \( \sqrt{\frac{1}{50} + \frac{1}{50}} = 0.2 \)
- If \( n = 99 \) and \( m = 1 \), \( \sqrt{\frac{1}{99} + \frac{1}{1}} = 1.01 \).

In the case that the standard deviations are the same, the standard error is smallest (and this the difference in the sample means is most reliable) when they both have sample sizes.

In general regardless of whether the standard deviations in both populations are the same or not, having one sample size will always
lead to larger standard errors and thus less reliable estimates of the difference in sample means.

• Remember, in the case that the standard deviations are the SAME, having equal sample sizes leads to estimators with the smallest standard errors. However, in the case that the standard deviations are NOT THE SAME, equal sample sizes will not lead to the smallest standard deviation.

• It does not matter if one sample size is much larger than the other, the total variability/reliability on the estimator of the difference is accounted for in the standard error:

\[
\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}
\]

variability of \( \bar{X} \)  variability of \( \bar{Y} \)
Example: Are the diets the same?

Two diets are being compared for effectiveness. 10 volunteers went on diet 1 and 10 different volunteers went on Diet 2. After one month their weight loss (in kilos) was recorded. The data is given below.

<table>
<thead>
<tr>
<th>Diet I</th>
<th>Diet II</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.9</td>
<td>3.5</td>
</tr>
<tr>
<td>2.7</td>
<td>2.5</td>
</tr>
<tr>
<td>3.9</td>
<td>3.8</td>
</tr>
<tr>
<td>2.7</td>
<td>8.1</td>
</tr>
<tr>
<td>2.1</td>
<td>3.6</td>
</tr>
<tr>
<td>2.6</td>
<td>2.5</td>
</tr>
<tr>
<td>2.2</td>
<td>5.0</td>
</tr>
<tr>
<td>4.2</td>
<td>2.9</td>
</tr>
<tr>
<td>5.0</td>
<td>2.3</td>
</tr>
<tr>
<td>0.7</td>
<td>3</td>
</tr>
</tbody>
</table>

Let \( \mu_I \) be the mean weight loss of diet I and \( \mu_{II} \) be the mean weight loss of diet II. Test the hypothesis that the diets are different.
Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations

**Plot of Diet data**

The summary statistics and histogram is given below.

<table>
<thead>
<tr>
<th>Distributions Column 2=0</th>
<th>Summary Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Column 1</strong></td>
<td>Mean 3.72</td>
</tr>
<tr>
<td></td>
<td>Std Dev 1.7344868</td>
</tr>
<tr>
<td></td>
<td>Std Err Mean 0.5484629</td>
</tr>
<tr>
<td>Upper 95% Mean 4.9607771</td>
<td>Lower 95% Mean 2.4792229</td>
</tr>
<tr>
<td>N 10</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distributions Column 2=1</th>
<th>Summary Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Column 1</strong></td>
<td>Mean 2.9</td>
</tr>
<tr>
<td></td>
<td>Std Dev 1.2128637</td>
</tr>
<tr>
<td></td>
<td>Std Err Mean 0.3835507</td>
</tr>
<tr>
<td>Upper 95% Mean 3.7678519</td>
<td>Lower 95% Mean 2.0323481</td>
</tr>
<tr>
<td>N 10</td>
<td></td>
</tr>
</tbody>
</table>
Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations

Checking the assumptions for the Diet Data

QQplot of each group and the residuals
• Since the people are not related and they were randomly allocated to one of the two diets it seems reasonable to assume that both data sets are independent.

• To check the normality assumption (for this small data set) we have made a QQplot. From the limited data that we have, it does not appear to be normal, this means that sample means (average weight loss for both diets will only be too close to normality). This will have an impact on the accuracy of the p-values that we obtain.

• However, there does not appear to be any huge outliers that may have significant impact on the outcome of the test (as we shall see later one outlier can have an dramatic impact on the test).
Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations

The JMP output

**Oneway Analysis of Column 1 By Column 2**

**t Test**

- Assuming unequal variances
  - Difference: -0.8200
  - t Ratio: -1.22517
  - Std Err Dif: 0.6693
  - DF: 16.10337
  - Upper CL Dif: 0.8381
  - Prob > U: 0.2381
  - Lower CL Dif: -2.2381
  - Prob > L: 0.8809
  - Confidence: 0.95
  - Prob < t: 0.1191
The hypothesis test and confidence interval

- Just for revision purposes the critical values for the t with 16.1df is

<table>
<thead>
<tr>
<th>probability</th>
<th>0.15</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t^*$</td>
<td>1.07</td>
<td>1.33</td>
<td>1.74</td>
<td>2.12</td>
<td>2.58</td>
<td>2.91</td>
</tr>
</tbody>
</table>

- We test the hypothesis $H_0 : \mu_I - \mu_{II} = 0$ against $H_A : \mu_I - \mu_{II} \neq 0$, we do the test at the 5% level. We have the t-transform $t = (-0.82 - 0)/0.66 = -1.22$. This lies between -1.33 and -1.07. Therefore the area to the left of -1.22 is between 10 to 15%. Thus the p-value is between 20-30% (as seen by the JMP output which tells us it is 23%). As this is larger than 5% we cannot reject the null at the 5% level. Though we cannot accept the null, the data is consistent with there not being any difference between the two diets.
• The 95% confidence interval for the mean difference between the diets is $[-2.2, 0.59]$ pounds.

• What if we wanted to test if $H_0 : \mu_I - \mu_{II} \leq 0$ against $H_A : \mu_I - \mu_{II} > 0$ at the 5% level?

• What if we wanted to test if $H_0 : \mu_I - \mu_{II} \geq 0$ against $H_A : \mu_I - \mu_{II} < 0$ at the 5% level?
**Warning: Never perform a mean test on the residuals!**

The residuals remove all mean information from the data. The difference in the sample means of the residuals will be zero - therefore the p-value of a 2-sided test will be 100% (and the p-value in a one-sided test will be 50%). To see this in action, here is the output of the independent sample t-test of the residuals of the diet data (considered above):

Naturally the test **CANNOT** detect a difference in the means, because there is no difference.
Formal: comparing populations

We have two samples from two different populations. That is $X_1, \ldots, X_n$ is a size $n$ sample (eg. heights of females in the 651 class) from population 1 (eg. heights of all females) and $Y_1, \ldots, Y_m$ is a size $m$ sample (eg. the heights of males in the 651 class) from population 2 (eg. heights of all males).

The mean of population 1, is $\mu_X$ (eg. mean height of a female) and the mean of population 2 is $\mu_Y$ (eg. mean height of a male). Given the samples we want to make inference about the difference $\mu_X - \mu_Y$.

• It is clear this is an important question. Other examples include:
  – Does a new therapy work better than old therapy?
  – Is there a difference in the performance of one school over another?
  – On average does eating healthy food mean you live longer? On average if one studies more do they get better grades?
Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations

- Is one housing material better than another?

- In the above examples what are the different populations and samples?

- All these questions are important and can lead to quite important decisions, therefore it is important that we do a careful analysis.
Residuals

Consider the output for the following data set

```plaintext
One way Analysis of Normal1 By Cat

<table>
<thead>
<tr>
<th>Normal</th>
<th>Cat</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0.5</td>
</tr>
<tr>
<td>8</td>
<td>0.5</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

T Test

 Assuming unequal variances

| Difference | 9.7665 | 1 Ratio | 47.93479 |
| Std Err Df | 0.22007 | DF | 78.99866 |
| Upper CL Df | 10.1710 | Prob > U | <0.001* |
| Lower CL Df | 9.99600 | Prob > L | <0.001* |
| Confidence | 0.95 | Prob <0.1 | 1.0000 |
```
Lecture 19 (MWF) Independent sample t-test for testing equality of means in two populations