Homework Solutions

Do all tests at the 5% level and quote p-values when possible. When answering each question uses sentences and include the relevant JMP output and plots (do not include the data in your solutions and make sure your plots and output are embedded in the solution).

All students must produce their own solution.

1. There has been speculation that breakfast cereal manufacturers add more sugar to cereals targeted at children than adults.

The sugar content of 24 varieties of cereals are given in http://www.stat.tamu.edu/~suhasini/teaching651/HW/Cereal.csv (note that the target audience is also given in this data set).

(a) Based on the discussion at the start of this question, what is the hypothesis of interest.

Let $\mu_{\text{child}}$ denote the mean sugar content in child’s cereals, $\mu_{\text{adult}}$ the mean sugar content in adult’s cereals.

Then the null hypothesis and alternative hypothesis are:

$$H_0: \mu_{\text{child}} - \mu_{\text{adult}} \leq 0 \quad \text{versus} \quad H_A: \mu_{\text{child}} - \mu_{\text{adult}} > 0.$$

(b) Based on this data set, which cereals contain more sugar, what is the difference between them?

The difference in sample means is $\bar{X}_{\text{child}} - \bar{X}_{\text{adult}} = 4.75$, which is larger than 0. For this sample of cereals there is more sugar for children cereals than adult cereals. We need to see whether this is statistical significant (i.e. if a similar result holds over the population of cereals).

(c) Do an independent sample t-test to test this assertion.

The JMP output of the t-test is given below.

![Oneway Analysis of Sugars By Target](image)

We see the difference is statistically significant with a p-value of 1.46%. As 1.46% < 5%, there is some evidence to suggest that the cereals targeted at children contain more sugar.
(d) Make a QQplot of the residuals to check for normality.

The total sample size is 24 (I think 10 in one group 14 in another - please check), this means that the sample size is not large enough for the central limit theorem to hold reliably. Therefore we need to check if the residuals are close to normal. If they are, we can say the p-value obtained in (c) is close to the truth. Below we give a QQplot of the data subtracted from their group mean.

We see that the points do not appear to deviate hugely from the line. Therefore, the difference of the sample means $\bar{X}_{\text{adult}} - \bar{X}_{\text{child}}$ will be close to normal and the p-value of 1.46% is reliable.

(e) Give a 95% CI for the mean difference in sugar levels.

From the output in part (c), we observe that 95% confidence interval of the mean difference is $[0.57308, 8.92692]$. The margin of error is large because the sample size is small.

2. Who are faster at paper mazes, Adults or Children?

A random sample of 10 adults and 10 children (between the ages 6-10) was taken. Each was asked to complete a maze and their times recorded. The data is given in http://www.stat.tamu.edu/~suhasini/teaching651/HW/maze.csv.

(a) For this data set, who are faster at completing a maze, children or adults, give the average difference in times.

From the graph listed below, we know the difference between children and adults is $\bar{X}_{\text{child}} - \bar{X}_{\text{adult}} = 4.167$, this means in this sample adults are on average faster than children.
(b) Give a 95% confidence interval for the mean difference in maze times.

The 95\% confidence interval for the mean difference in maze times is $[-3.327, 11.661]$.

(c) Use an independent sample t-test to test the research hypothesis that there is a difference between the mean adult and child times.

A scatter plot, t-test and Wilcoxon test are given below.

The hypothesis of interest is $H_0 : \mu_{\text{adult}} - \mu_{\text{child}} = 0$ vs $H_A : \mu_{\text{adult}} - \mu_{\text{child}} \neq 0$.

The $p$ value is 0.2505. We cannot reject the null hypothesis. This means despite there being a difference in average maze times between adults and children, the difference is not statistically significant. Such a difference can be obtained under the null hypothesis.

(d) The sample sizes are small. Therefore do a Wilcoxon sum rank test too. The results in (c) and (d) concur?

Refer to the table above, we know:

$$T_{\text{adult}} = 96 \quad T_{\text{child}} = 114.$$ 

Since the sample sizes are the same we can use either value. The non-rejection region for the two-sided test (at the 5\% level) is $[79, 131]$ (this means under the null of no difference we would expect the ranks to lie in this interval). Since 96 lies in this interval we cannot reject the null (note that also 114 lies in this interval - so it does not matter which value we choose). Therefore just like the independent sample t-test we cannot reject the null for the Wilcoxon Sum Rank test.

3. A group of dieticians are investigating the efficacy of a certain diet in losing weight.
A small group of 6 individuals are placed on a diet. The weights before going on the diet and after going on the diet for two months are recorded. For these 6 individuals we see the average weight loss is 11 pounds.

(a) First apply an independent sample t-test to see whether the difference is statistically significant (remember to state the hypothesis of interest).

The data is given in http://www.stat.tamu.edu/~suhasini/teaching651/HW/Diet_independent.csv. Let \( \mu_{\text{before}} \) be the mean weight before going on the diet, \( \mu_{\text{after}} \) be the mean weight after going on the diet. Since we want to see whether diet results in weight loss the hypothesis of interest is

\[
H_0: \mu_{\text{before}} - \mu_{\text{after}} \leq 0 \quad \text{versus} \quad H_A: \mu_{\text{before}} - \mu_{\text{after}} > 0.
\]

The average weight loss for this group of individuals is 11 pounds. Using the independent sample t-test, the p-value is 0.3578(35.87%) > 0.05. The results of this test suggest that the difference of 11 pounds is not statistically significant.

However, the independent sample t-test is conducted under the assumption that both groups are independent of each. Clearly since the same individual is used in the before and after group this seems like a very dubious assumption. Remember if the assumption is wrong, it can lead to misleading results. In part (b) we make a scatter plot of the data.

(b) There is a clear matching in the data (since the same person is weighed before and after diet). The same data, but with the appropriate matching is given in http://www.stat.tamu.edu/~suhasini/teaching651/HW/Diet_Matched.csv. Use the Fit Y by X option to make a scatter plot of the weights before and after the diet. Is there a clear dependence between these variables (give the plot and discuss what you see)?

A scatter plot is of the before against the after weights is given below.
As one would expect there is a clear linear dependence between the before and after weights. In other words, people who tend to be heavier at the start also tend to be heavier after going on the diet. The clear linear dependence in the data, strongly suggests that a matched t-test is more appropriate than an independent sample t-test.

(c) Apply the Matched paired t-test to this data (remember do the appropriate one-sided test).

The JMP output of the Matched paired t-test is given below.

We see that 5 of the 6 people lost weight and one gained a little weight. Testing \( H_0 : \mu_d \leq 0 \) vs. \( H_A : \mu_d > 0 \) using the matched paired t-test, we see the difference is statistically significant with a p-value of 1.63%.

(d) By comparing the outputs of the independent sample t-test and matched paired t-test explain what the main differences are and why the conclusions are different.

We should use the conclusion of matched paired t-test.

The independent sample t-test includes additional variation occurring from the independence of the observations, while a paired t-test is not subject to this variation. This is because the paired observations are dependent.

(e) Apply the Sign test to this data (remember do a one-sided test), you can use JMP
to do this.

Now we test $H_0: \text{Median} \leq 0$ vs $H_A: \text{Median} > 0$. We see that the p-value for this test is 10.94%. This is relatively large; it is relatively ‘simple’ to get 5 positives out of 6 under the null that the chance of the positive and negative is equal. We recall that the sign test is very conservative and it is difficult to reject the null with the sign test.

(f) Apply the Wilcoxon sign test (remember do a one-sided test). Please do this by hand.

Recall we test the hypothesis $H_0: \mu_{\text{before}} - \mu_{\text{after}} \leq 0$ vs $H_A: \mu_{\text{before}} - \mu_{\text{after}} > 0$. The sums of positives and negatives are

$$T_+ = (2 + 3 + 4 + 5 + 6) = 20$$
$$T_- = 1.$$ 

Since the test is a one-sided test we focus on $T_- = 1$ (it would be ‘small’ under the alternative). To decide on the threshold use the Wilcoxon Sign-rank tables. If the rank is less than 2 (5% level, one-sided test) we reject the null. Since $1 = T_- < 2$, we reject the null using the Wilcoxon Sign-rank test.

(g) Briefly compare your conclusions from the three matched procedures in (c), (e) and (f).

The matched paired t-test assumes that the sample mean of the difference is normal; as the sample size is very small (n=6), we require that the observations cannot deviate much from normality. For the matched paired t-test we were able to reject the null.

Given that the sample size is so small, it makes sense to also try the nonparametric tests, where no distributional assumptions are placed on the data. We saw we could reject the null using the Wilcoxon sign-rank test (this test takes into account both sign and rank), however, the sign test is very conservative and we are unable to reject the null using the sign test.

4 Do have fertilizers have an impact on the height of marigolds?

150 marigolds (each in boxes of size 50) were given three different fertilizer treatments (Control, Fertilizer A and Fertilizer B) from seed until day 90. At Day 90 they were measured.

The data is given in

http://www.stat.tamu.edu/~suhasini/teaching651/HW/MarigoldFertilizerComparison.csv.

(a) What is the average height of marigolds in each of the three groups?

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Fertilizer A</th>
<th>Fertilizer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>average</td>
<td>9.64</td>
<td>12.1</td>
<td>6.03</td>
</tr>
<tr>
<td>sd</td>
<td>2.16</td>
<td>2.91</td>
<td>3.3</td>
</tr>
<tr>
<td>s.e</td>
<td>0.305</td>
<td>0.41</td>
<td>0.47</td>
</tr>
</tbody>
</table>
We observe for this data set that there are differences in the sample means for the different treatment groups. Below we to an ANOVA to see whether these differences are statistically significant.

(b) Use ANOVA to test the hypothesis that the treatment has no influence on the mean height $H_0 : \mu_C = \mu_A = \mu_B$ vs $H_A : \text{At least one mean is different.}$

![Oneway Analysis of Height By Treatment](image)

**Oneway Anova**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Ratio</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>2</td>
<td>932.5530</td>
<td>466.275</td>
<td>57.9156</td>
<td>&lt;0.0001 *</td>
</tr>
<tr>
<td>Error</td>
<td>147</td>
<td>1193.4894</td>
<td>8.391</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Total</td>
<td>149</td>
<td>2116.0424</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Means for Oneway Anova**

<table>
<thead>
<tr>
<th>Level</th>
<th>Number</th>
<th>Mean</th>
<th>Std Error</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>50</td>
<td>9.6464</td>
<td>0.40127</td>
<td>8.853</td>
<td>10.439</td>
</tr>
<tr>
<td>Fertilizer A</td>
<td>50</td>
<td>12.1059</td>
<td>0.40127</td>
<td>11.313</td>
<td>12.899</td>
</tr>
<tr>
<td>Fertilizer B</td>
<td>50</td>
<td>6.0347</td>
<td>0.40127</td>
<td>5.242</td>
<td>6.828</td>
</tr>
</tbody>
</table>

The $F$-value is extremely large (57.9 - telling is that it very difficult to obtain these differences under the null that all the populations means are the same). The corresponding p-value for the test is less than 0.01%, thus there is strong evidence to suggest that at least one mean is different.

(c) Make a QQplot of the residuals.

A QQplot of the residuals is given below.
The residuals do not appear to be close to normal; it looks mildly left skewed. However the sample sizes are large (50 in each group) that we assume that the sample means are close to normal and the p-value reliable. Moreover, the p-value is so small, that even if the distribution of the sample means deviated from normal with such a small p-value we could reject the null.

(d) Use ANOVA to test for equality of means on the residuals. Is the result what you expect?

The residuals for each group have sample same sample mean, zero, therefore the ANOVA would not be able to reject the null. Since the residuals are completely consistent with the null of the means being the same. The p-value will be one. This is exactly what we see when we do the test (observe the p-value is one in the output).
Oneway Analysis of Height centered by Treatment By Treatment

Oneway Anova

<table>
<thead>
<tr>
<th>Summary of Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residuals</td>
</tr>
<tr>
<td>Adj. Residuals</td>
</tr>
<tr>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>Mean of Response</td>
</tr>
<tr>
<td>Observations (or Sum Wgts)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analysis of Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
</tr>
<tr>
<td>Treatment</td>
</tr>
<tr>
<td>Error</td>
</tr>
<tr>
<td>C. Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Means for Oneway Anova</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>Fertilizer A</td>
</tr>
<tr>
<td>Fertilizer B</td>
</tr>
</tbody>
</table>

Std Error uses a pooled estimate of error variance.