1. In an experiment, four treatments were randomly assigned to blood samples to investigate the effect of the treatments on the clotting time of plasma. The data of the experiment is given in the following table. The measurements are the clotting times of plasma, in minutes.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.4</td>
<td>9.4</td>
<td>9.8</td>
<td>12.2</td>
<td></td>
</tr>
<tr>
<td>12.8</td>
<td>15.2</td>
<td>12.9</td>
<td>14.4</td>
<td></td>
</tr>
<tr>
<td>9.6</td>
<td>9.1</td>
<td>11.2</td>
<td>9.8</td>
<td></td>
</tr>
<tr>
<td>9.8</td>
<td>8.8</td>
<td>9.9</td>
<td>12.0</td>
<td></td>
</tr>
<tr>
<td>8.4</td>
<td>8.2</td>
<td>8.5</td>
<td>8.5</td>
<td></td>
</tr>
<tr>
<td>8.6</td>
<td>9.9</td>
<td>9.8</td>
<td>10.9</td>
<td></td>
</tr>
<tr>
<td>8.9</td>
<td>9.0</td>
<td>9.2</td>
<td>10.4</td>
<td></td>
</tr>
<tr>
<td>7.9</td>
<td>8.1</td>
<td>8.2</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>9.300</td>
<td>9.713</td>
<td>9.938</td>
<td>11.025</td>
</tr>
<tr>
<td>sd</td>
<td>1.550</td>
<td>2.294</td>
<td>1.514</td>
<td>1.815</td>
</tr>
</tbody>
</table>

(i) Conduct the residual analysis to check model adequacy of the one-way ANOVA model.

*Go to R session*

(ii) Suppose we are only interested in investigating the following three differences: (a) $\mu_1 - \mu_4$, (b) $\mu_2 - \mu_4$ and (c) $\mu_3 - \mu_4$, where $\mu_j$ is the mean effect of treatment $j$. Test the significance of these differences at level 0.05 by using an appropriate multiple comparison procedure.

*The ANOVA table:*

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tmt</td>
<td>3</td>
<td>13.016</td>
<td>4.339</td>
<td>1.3096</td>
</tr>
<tr>
<td>Residuals</td>
<td>28</td>
<td>92.763</td>
<td>3.313</td>
<td></td>
</tr>
</tbody>
</table>

1
The three test statistics are computed below:

\[
L_1 = \frac{\hat{\mu}_1 - \hat{\mu}_4}{\sqrt{\text{MSE}}} \sqrt{\frac{n}{2}} = \frac{9.3 - 11.025}{\sqrt{3.313}} \sqrt{4} = -1.895;
\]

\[
L_2 = \frac{\hat{\mu}_2 - \hat{\mu}_4}{\sqrt{\text{MSE}}} \sqrt{\frac{n}{2}} = \frac{9.713 - 11.025}{\sqrt{3.313}} \sqrt{4} = -1.4422;
\]

\[
L_3 = \frac{\hat{\mu}_3 - \hat{\mu}_4}{\sqrt{\text{MSE}}} \sqrt{\frac{n}{2}} = \frac{9.938 - 11.025}{\sqrt{3.313}} \sqrt{4} = -1.1950.
\]

The Dunnett’s criterion is appropriate for these comparisons. The critical value of the Dunnett’s criterion at 0.05 level for two sided test is \(d_{3,28,0.05} \approx 2.47\). Comparing the absolute values of the \(L_j\)’s with this critical value shows that none of the differences is significant.

2. A manufacturer of television sets is interested in the effect on tube conductivity of four different types of coating for color picture tubes. The following conductivity data are obtained:

<table>
<thead>
<tr>
<th>Coating Type</th>
<th>Conductivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>143 141 150 146</td>
</tr>
<tr>
<td>2</td>
<td>152 149 137 143</td>
</tr>
<tr>
<td>3</td>
<td>134 136 132 127</td>
</tr>
<tr>
<td>4</td>
<td>129 127 132 129</td>
</tr>
</tbody>
</table>

(i) Is there a difference in conductivity due to coating type? Use level \(\alpha = 0.05\).

The ANOVA table:

\[
\begin{array}{llllll}
\text{Df} & \text{Sum Sq} & \text{Mean Sq} & \text{F value} & \text{Pr(>F)} \\
\text{tmt} & 3 & 844.69 & 281.56 & 14.302 & 0.0002881 *** \\
\text{Residuals} & 12 & 236.25 & 19.69 & \\
\end{array}
\]

From the ANOVA table, the F statistic for effect of coating type is 14.302 with a p-value 0.00029. Therefore, at level 0.05, the difference is significant.

(ii) Compute the 95% simultaneous confidence intervals for differences of all pairs of mean effects of the coating types.

The intervals are computed either by the formula

\[
\sum_{i=2}^{4} c_i \hat{\beta}_i \pm \frac{q_{0.05}(4, 12)}{\sqrt{2}} \sqrt{\frac{\text{MSE}}{4} \sum_{i=2}^{4} c_i^2},
\]
where the cᵢ's are the components of the contrast vector corresponding to the pairwise differences, q₀.₀₅(4, 12) = 4.2, or by the formula

\[ \hat{\mu}_i - \hat{\mu}_j \pm q_{0.05}(4,12) \sqrt{\frac{MSE}{4}}, \]

The intervals are computed as follows:

d₁₂  -9.567826  9.067826  
d₁₃  3.432174  22.067826  
d₁₄  6.432174  25.067826  
d₂₃  3.682174  22.317826  
d₂₄  6.682174  25.317826  
d₃₄  -6.317826  12.317826

(iii) Analyze the residuals and draw conclusions about the model adequacy.

Go to R session.

3. In an experiment, the amount of radon released in showers was investigated. Rodon enriched water was used in the experiment, and six different orifice diameters were tested in shower heads. The data from the experiment are shown below:

<table>
<thead>
<tr>
<th>Orifice Diameter</th>
<th>Radon released %</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.31</td>
<td>80  83  83  85</td>
</tr>
<tr>
<td>0.51</td>
<td>75  75  79  79</td>
</tr>
<tr>
<td>0.71</td>
<td>74  73  76  77</td>
</tr>
<tr>
<td>0.91</td>
<td>67  72  74  74</td>
</tr>
<tr>
<td>1.11</td>
<td>62  62  67  69</td>
</tr>
<tr>
<td>1.31</td>
<td>60  61  64  66</td>
</tr>
</tbody>
</table>

(i) Does the size of the orifice affect the mean percentage of radon released? Use level \( \alpha = 0.05 \).

(ii) Compute the \( p \)-value of \( F \)-statistic of the test in (i).

(i) and (ii) can be answered from the following ANOVA table:

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tmt</td>
<td>5</td>
<td>1133.38</td>
<td>226.68</td>
<td>0.852</td>
</tr>
<tr>
<td>Residuals</td>
<td>18</td>
<td>132.25</td>
<td>7.35</td>
<td></td>
</tr>
</tbody>
</table>
Since the p-value is less than 0.05, the size of the orifice affect the mean percentage of radon significantly at level 0.05.

The p-value can be computed using \(1 - \text{pf}(30.852, 5, 18)\).

(iii) If the test in (i) is significant, what contrasts can you detect significant? Use level \(\alpha = 0.05\).

The sample means are computed as follows:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>82.75</td>
<td>77.00</td>
<td>75.00</td>
<td>71.75</td>
<td>65.00</td>
<td>62.75</td>
</tr>
</tbody>
</table>

The constraints with the following constraint vectors seem to be significant:

\[c(1, -1, 0, 0, 0, 0)\]
\[c(1, 0, -1, 0, 0, 0)\]
\[c(1, 0, 0, -1, 0, 0)\]
\[c(1, 0, 0, 0, -1, 0)\]
\[c(0, 1, 0, -1, 0, 0)\]
\[c(0, 1, 0, 0, -1, 0)\]
\[c(0, 1, 0, 0, 0, -1)\]
\[c(0, 1/3, 1/3, 1/3, -1/2, -1/2)\]
\[c(-1, 1/3, 1/3, 1/3, 0, 0)\]

etc.

The Sheffe’s criterion should be used for the comparison. The Sheffe’s critical value is given by \(c_{0.05} = \sqrt{5F_{0.05}(5, 18)} = 3.7235\). The test statistics corresponding to the above contrasts are computed using the formula

\[
\sum_{i=1}^{6} c_i \bar{y}_i \pm c_{0.05} \sqrt{\frac{MSE}{n} \sum_{i=1}^{6} c_i^2}
\]

where \(MSE = 7.35\), or

\[
\sum_{i=2}^{6} c_i \hat{\beta}_i \pm c_{0.05} \sqrt{\text{Var}(\sum_{i=2}^{6} c_i \hat{\beta}_i)}
\]

as
\[
\begin{array}{c}
d1 & 3.000000 \\
d2 & 4.043478 \\
d3 & 5.739130 \\
d4 & 9.260870 \\
d5 & 10.434783 \\
d6 & 2.739130 \\
d7 & 6.260870 \\
d8 & 7.434783 \\
d9 & 8.655276 \\
d10 & -5.218478 \\
\end{array}
\]

Compared with the critical value, except \( d1 \) and \( d6 \), all the other contrasts are significant at level 0.05.

4. Lysozyme levels in the gastric juice of 29 patients with peptic ulcer and of 30 normal controls are given below.

Patient group:

0.2 0.3 0.4 1.1 2.0 2.1 3.3 3.8 4.5 4.8 4.9 5.0 5.3 7.5 9.8 10.4 10.9 11.3 12.4 16.2 17.6 18.9 20.7 24.0 25.4 40.0 42.2 50.0 60.0

Normal group:

0.2 0.3 0.4 0.7 1.2 1.5 1.5 1.9 2.0 2.4 2.5 2.836 4.8 4.8 5.4 5.7 5.8 7.5 8.7 8.8 9.1 10.3 15.6 16.1 16.5 16.7 20.0 20.7 33.0

(i) Carry out a residual analysis to see whether the assumptions of constant variance and normality are valid for the data.

*Go to R session.*

(ii) Using an appropriate method to suggest a data transformation such that the transformed data have approximately a constant variance.

*Method 1: Try the transformation \( x^\lambda \) for several values of \( \lambda \). For each transformed data, compute ratio of the sample variances of the two groups, the \( \lambda \) value which yields the ratio closest to 1 is suggested.

For \( \lambda = (-0.5, -1, 0.001, 0.5, 2) \), for following are the ratios

1.071910, 1.050815, 1.267593, 2.072238, 14.939097.
The reciprocal transformation $x^{-1}$ is suggested.

Method 2: For $\alpha = (3, 2, 3/2, 2, 1, 1/2, -1/2, -1)$ compute

$$\frac{\max(s_i/\bar{y}_i^\alpha)}{\min(s_i/\bar{y}_i^\alpha)},$$

find $\alpha$ such that the ratio is closest to 1. Then take transformation $x^\lambda$ with $\lambda = 1 - \alpha$.

The computed ratios are

\begin{align*}
\text{alpha} & \quad 3.000 & 2.00 & 1.500 & 2.00 & 1.000 & -0.500 & -1.000 \\
\text{sn} & \quad 3.222 & 1.73 & 1.268 & 1.73 & 1.077 & 1.469 & 2.737 & 3.735
\end{align*}

The log transform is suggested.

(iii) Conduct the Mann-Whitney-Wilcoxon test based on the data: (a) without adjustment on ties, and (b) with adjustment on ties. Compare the results.

Ranks of Group 1:

\begin{align*}
&[1] \quad 1.5 \quad 3.5 \quad 5.5 \quad 8.0 \quad 13.5 \quad 15.0 \quad 19.0 \quad 21.0 \quad 22.0 \quad 24.0 \quad 26.0 \quad 27.0 \quad 28.0 \quad 32.5 \quad 37.0 \\
&[16] \quad 39.0 \quad 40.0 \quad 41.0 \quad 42.0 \quad 45.0 \quad 48.0 \quad 49.0 \quad 51.5 \quad 53.0 \quad 54.0 \quad 56.0 \quad 57.0 \quad 58.0 \quad 59.0
\end{align*}

Ranks of Group 2:

\begin{align*}
&[1] \quad 1.5 \quad 3.5 \quad 5.5 \quad 7.0 \quad 9.0 \quad 10.5 \quad 10.5 \quad 12.0 \quad 13.5 \quad 16.0 \quad 17.0 \quad 18.0 \quad 20.0 \quad 24.0 \quad 24.0 \\
&[16] \quad 29.0 \quad 30.0 \quad 31.0 \quad 32.5 \quad 34.0 \quad 35.0 \quad 36.0 \quad 38.0 \quad 43.0 \quad 44.0 \quad 46.0 \quad 47.0 \quad 50.0 \quad 51.5 \quad 55.0
\end{align*}

From the table, we obtain

$$\bar{R}_1 = 33.66, \bar{R}_2 = 26.47,$$

$$\chi^2 = \frac{12n_1n_2(\bar{R}_1 - \bar{R}_2)^2}{n^2(n + 1)} = \frac{12 \times 29 \times 30 \times (33.66 - 26.47)^2}{59^2 \times 60} = 2.5830.$$  

(a) The p-value of the test without adjustment for ties is $Pr(\chi^2_1 > 2.5830) = 0.1080$.

(b) To adjust for ties, compute $f$ as follows:

\begin{align*}
&[1] \quad 0.2 \quad 0.2 \quad 0.3 \quad 0.3 \quad 0.4 \quad 0.4 \quad 0.7 \quad 1.1 \quad 1.2 \quad 1.5 \quad 1.5 \quad 1.9 \quad 2.0 \quad 2.0 \quad 2.1 \\
&[16] \quad 2.4 \quad 2.5 \quad 2.8 \quad 3.3 \quad 3.6 \quad 3.8 \quad 4.5 \quad 4.8 \quad 4.8 \quad 4.8 \quad 4.9 \quad 5.0 \quad 5.3 \quad 5.4 \quad 5.7 \\
&[31] \quad 5.8 \quad 7.5 \quad 7.5 \quad 8.7 \quad 8.8 \quad 9.1 \quad 9.8 \quad 10.3 \quad 10.4 \quad 10.9 \quad 11.3 \quad 12.4 \quad 15.6 \quad 16.1 \quad 16.2 \\
&[46] \quad 16.5 \quad 16.7 \quad 17.6 \quad 18.9 \quad 20.0 \quad 20.7 \quad 20.7 \quad 24.0 \quad 25.4 \quad 33.0 \quad 40.0 \quad 42.2 \quad 50.0 \quad 60.0
\end{align*}
\[ T = 8, \ t_1 = t_2 = t_3 = t_4 = t_5 = 2, t_6 = 3, t_7 = t_8 = 2 \]
\[ \sum_{i=1}^{T} t_i(t_i - 1)(t_i + 1) = 2 \times 3 \times 7 + 3 \times 2 \times 4 = 66. \]
\[ n(n - 1)(n + 1) = 205320. \]
\[ f = 1 - 66/205320 = 0.9968. \]

The adjusted statistic is \( \chi^2/0.9968 = 2.5913 \). The corresponding p-value is 0.1074.