1. In an experiment on Plant Growth, 10 plants received treatment 1, another 10 plants received treatment 2, and another 10 plants were used as controls (giving 30 plants in all). The growth was measured and the sums of these measurements for each of the 3 samples were 46.61 for treatment 1, 55.26 for treatment 2 and 50.32 for the controls. The sum of squares of all 30 measurements was 786.3183. Obtain the analysis of variance table.

Solution: The sum of all observations is 152.19, so the total sum of squares is

\[ 786.3183 - \frac{152.19^2}{30} = 14.25843 \, , \]

the treatment sum of squares is

\[ \frac{50.32^2 + 46.61^2 + 55.26^2}{10} - \frac{152.19^2}{30} = 3.7663 \, , \]

and the residual sum of squares is thus \(14.25843 - 3.7663 = 10.4921\). The anova table is thus

<table>
<thead>
<tr>
<th>Source of var’n</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>3.7663</td>
<td>2</td>
<td>1.8832</td>
<td>4.8461</td>
</tr>
<tr>
<td>Residual</td>
<td>10.4921</td>
<td>27</td>
<td>0.3886</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14.2584</td>
<td>29</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. In an effort to improve the quality of tapes, the effects of four kinds of coatings, A, B, C and D on the reproducing quality of sound were compared. The measurements of sound distortion are given below:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10</td>
<td>15</td>
<td>8</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>B</td>
<td>14</td>
<td>18</td>
<td>21</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>17</td>
<td>16</td>
<td>14</td>
<td>15</td>
<td>18</td>
</tr>
<tr>
<td>D</td>
<td>12</td>
<td>15</td>
<td>17</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

```
The sum of squares of all 22 observations is $10^2 + 15^2 + \cdots + 15^2 = 5112$.

(a) Obtain the anova table for these four samples.

**Solution:** The sum of all observations is 330, so

$$\text{Total SS} = \frac{5112^2 - 330^2}{22} = 162,$$

$$\text{Between SS} = \frac{60^2}{5} + \frac{68^2}{4} + \frac{112^2}{7} + \frac{90^2}{6} - \frac{330^2}{22} = 68$$

Thus by taking differences we get

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>68</td>
<td>3</td>
<td>22.6</td>
<td>4.434</td>
</tr>
<tr>
<td>Within</td>
<td>94</td>
<td>18</td>
<td>5.2</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>162</td>
<td>21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) Test the hypothesis

$$H: \text{“the four samples above were randomly drawn from identical normal populations”}$$

using large values of the $F$-ratio as evidence against $H$.

**Solution:** Under $H$, the $F$-ratio has the $F_{3,18}$ distribution. So using large values of it as evidence, the p-value is $P(F_{3,18} > 4.434)$, which from the tables is between 1% and 2.5%, providing some evidence against $H$.

3. Suppose we have two samples $y_{11}, \ldots, y_{1n_1}$ and $y_{21}, \ldots, y_{2n_2}$. Student’s two-sample $t$-statistic is

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},}$$

where $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$, for $i = 1, 2$, are the sample means,

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

is the pooled sample variance and $s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$, for $i = 1, 2$, are the sample variances. Show that $t^2$ is exactly the one-way anova
$F$-ratio when there are only 2 treatments. (Hint: write $\bar{y}.$ in terms of $\bar{y}_1$ and $\bar{y}_2$).

**Solution:** The overall mean $\bar{y}.$ can be written as

$$\bar{y}.$ = \frac{n_1\bar{y}_1 + n_2\bar{y}_2}{n_1 + n_2}.$$  

So

$$\bar{y}_1 - \bar{y}.$ = \frac{(n_1 + n_2)\bar{y}_1 - [n_1\bar{y}_1 + n_2\bar{y}_2]}{n_1 + n_2} = \frac{n_2(\bar{y}_1 - \bar{y}_2)}{n_1 + n_2}.$$  

Similarly

$$\bar{y}_2 - \bar{y}.$ = \frac{n_1(\bar{y}_2 - \bar{y}_1)}{n_1 + n_2},$$  

So the Between/Treatment Sum of Squares becomes

$$n_1(\bar{y}_1 - \bar{y}.)^2 + n_2(\bar{y}_2 - \bar{y}.)^2 = \frac{(n_1n_2^2 + n_1^2n_2)(\bar{y}_1 - \bar{y}_2)^2}{(n_1 + n_2)^2} = \frac{(n_1 + n_2)n_1n_2(\bar{y}_1 - \bar{y}_2)^2}{(n_1 + n_2)^2} = \frac{(\bar{y}_1 - \bar{y}_2)^2}{\frac{1}{n_1} + \frac{1}{n_2}},$$

since

$$\frac{1}{n_1} + \frac{1}{n_2} = \frac{n_1 + n_2}{n_1n_2}.$$  

Also, it is straightforward to show that $s_p^2$ is exactly the Within/Residual MS. So then when we have two treatments, $t^2 = F$ (in that case the Treatment SS and Treatment MS are the same).

4. (a) The Cauchy inequality says the following: for any two sets of numbers $a_1, \ldots, a_t$ and $b_1, \ldots, b_t,$

$$\left( \sum_{i=1}^{t} a_i b_i \right)^2 \leq \left( \sum_{i=1}^{t} a_i^2 \right) \left( \sum_{i=1}^{t} b_i^2 \right),$$

3
with equality only if \( a_i = kb_i \) for some constant \( k \), for all \( i \). Use this to show that

\[
\frac{\sum_{i=1}^{t} a_i b_i}{\sqrt{\sum_{i=1}^{t} a_i^2}} \leq \sqrt{\sum_{i=1}^{t} b_i^2}.
\]

**Solution:** follows easily using, for \( b \geq 0 \),

\[
a^2 \leq b \iff -\sqrt{b} \leq a \leq \sqrt{b} \iff |a| \leq \sqrt{b}.
\]

(b) Let \( \bar{y}_1, \ldots, \bar{y}_t \) be a set of sample means. Suppose \( c_1, \ldots, c_t \) satisfy \( \sum_{i=1}^{t} c_i = 0 \). Then show that

\[
\sum_{i=1}^{t} c_i \bar{y}_i = \sum_{i=1}^{t} c_i (\bar{y}_i - \bar{y}) ,
\]

where \( \bar{y} = \sum_{i=1}^{t} n_i \bar{y}_i / \sum_{i=1}^{t} n_i \) is the overall mean.

**Solution:**

\[
\sum_{i=1}^{t} c_i (\bar{y}_i - \bar{y}) = \sum_{i=1}^{t} c_i \bar{y}_i - \bar{y} \sum_{i=1}^{t} c_i = \sum_{i=1}^{t} c_i \bar{y}_i - 0.
\]

(c) Use the results of (a) and (b) to show that for any \( \sum_{i=1}^{t} c_i = 0 \),

\[
\frac{|\sum_{i=1}^{t} c_i \bar{y}_i|}{\sqrt{\sum_{i=1}^{t} c_i^2 / n_i}} \leq \sqrt{\sum_{i=1}^{t} n_i (\bar{y}_i - \bar{y})^2} ,
\]

with equality only if \( c_i = kn_i (\bar{y}_i - \bar{y}) \) for some \( k \), and all \( i \). Hence show that the most significant sample contrast is \( \sqrt{(t-1)F} \), where \( F \) is the usual \( F \)-ratio.

**Solution:**

\[
\frac{|\sum_{i=1}^{t} c_i \bar{y}_i|}{\sqrt{\sum_{i=1}^{t} c_i^2 / n_i}} = \frac{|\sum_{i=1}^{t} c_i (\bar{y}_i - \bar{y})|}{\sqrt{\sum_{i=1}^{t} c_i^2 / n_i}} \quad \text{using (b)}
\]

\[
= \sum_{i=1}^{t} \frac{c_i}{\sqrt{n_i}} \sqrt{n_i (\bar{y}_i - \bar{y})} \quad \text{= } \sum_{i=1}^{t} \frac{c_i \sqrt{n_i}}{\sqrt{n_i}} (\bar{y}_i - \bar{y}_i) \quad \text{= } \sqrt{\sum_{i=1}^{t} c_i^2 / n_i}
\]

\[
\leq \sqrt{\sum_{i=1}^{t} n_i (\bar{y}_i - \bar{y})^2}.
\]
using the result of part (a), with \( a_i = \frac{c_i}{\sqrt{n_i}} \) and \( b_i = \sqrt{n_i}(\bar{y}_i - \bar{y}_.) \), which therefore indicates that equality only holds if \( a_i = kb_i \), i.e. that \( c_i = kn_i(\bar{y}_i - \bar{y}_.) \), for all \( i \) and some \( k \). But the RHS above is \( \sqrt{\text{TrSS}} \), so

\[
\max_{\sum_i c_i = 0} \frac{|\sum_{i=1}^t c_i \bar{y}_i|}{\sigma \sqrt{\sum_{i=1}^t c_i^2 / n_i}} = \frac{\sqrt{\text{TrSS}}}{\sigma} = \sqrt{\frac{(t-1)\text{TrMS}}{\text{ResMS}}} = \sqrt{(t-1)\bar{F}}.
\]