1. An experiment randomizes $N = 30$ peanut plants to three treatment groups: a slow-release fertilizer (S), a fast-release fertilizer (F) and a standard fertilizer for control (C). Two measurements were taken on each plant, yield $y_i$ and height $z_i$. for $i = 1, 2, \ldots, 30$:

<table>
<thead>
<tr>
<th>Slow</th>
<th>Fast</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>$z$</td>
<td>$y$</td>
</tr>
<tr>
<td>18.4</td>
<td>43.8</td>
<td>18.9</td>
</tr>
<tr>
<td>17.3</td>
<td>41.8</td>
<td>18.0</td>
</tr>
<tr>
<td>20.0</td>
<td>45.5</td>
<td>17.3</td>
</tr>
<tr>
<td>21.0</td>
<td>49.1</td>
<td>21.7</td>
</tr>
<tr>
<td>20.9</td>
<td>49.8</td>
<td>19.9</td>
</tr>
<tr>
<td>21.7</td>
<td>56.5</td>
<td>20.7</td>
</tr>
<tr>
<td>20.4</td>
<td>51.0</td>
<td>19.0</td>
</tr>
<tr>
<td>20.9</td>
<td>53.0</td>
<td>18.6</td>
</tr>
<tr>
<td>18.7</td>
<td>42.0</td>
<td>19.7</td>
</tr>
<tr>
<td>21.3</td>
<td>46.2</td>
<td>18.1</td>
</tr>
</tbody>
</table>

Relevant SAS code and output appear at the end of the following questions.

(a) Let $x_iS$ denote an indicator variable for the slow-release treatment:

$$x_iS = \begin{cases} 1 & \text{plant } i \text{ receives treatment “S”} \\ 0 & \text{else} \end{cases}$$

Similarly define another indicator variable for the fast-release treatment, $x_iF$.

$$x_iF = \begin{cases} 1 & \text{plant } i \text{ receives treatment “F”} \\ 0 & \text{else} \end{cases}$$

(b) Let $E_i \overset{iid}{\sim} N(0, \sigma^2)$. Consider the model

$$Y_i = \beta_0 + \beta_S x_iS + \beta_F x_iF + \beta_h z_i + E_i \text{ for } i = 1, 2, \ldots, 30$$

$$\dim(X) = (30 \times 4)$$

(c) Using the output, estimate the mean yield as a linear function of height $z$ separately for each treatment group.

$$\mu_S(z) = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_h z = 10.86 + 0.19z$$

$$\mu_F(z) = \hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_h z = 8.96 + 0.19z$$

$$\mu_C(z) = \hat{\beta}_0 + \hat{\beta}_h z = 9.79 + 0.19z$$
(d) In light of the analysis of covariance, is it plausible that the fertilizer treatment has no effect on yield? Report a p-value. Use $\alpha = 0.05$ to conduct a test.

$$H_0 : \beta_1 = \beta_2 = 0 \text{ (after controlling for } z)$$

$$F = \frac{R(\beta_1, \beta_2 | \beta_0, \beta_h)/2}{MS[E]}$$

$$= \frac{(\text{Type III MS for trt})/MS[E]}{10.3(p = 0.0005 \text{ on } df = 2, 26)}$$

"So, after controlling for height, there is significant variability in the mean yields among the three fertilizer groups."

(e) For a given height $z$, estimate the difference between mean yield among a population of peanut plants receiving the control (C) fertilizer and among a population receiving the fast-release (F) fertilizer. Report a standard error and a p-value for a test that this difference is 0.

$$\hat{\mu}_F(z) - \hat{\mu}_C(z) = \beta_F$$

$$\hat{\mu}_F(z) - \hat{\mu}_C(z) = \hat{\beta}_F$$

$$= -0.84(\text{SE} = 0.46, p = 0.0820)$$

(f) Complete the table of unadjusted means and means adjusted to the average plant height of $\bar{z} = 48.9$.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>$n$</th>
<th>unadj. mean</th>
<th>adj. mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>10</td>
<td>20.06</td>
<td>20.2</td>
</tr>
<tr>
<td>F</td>
<td>10</td>
<td>19.19</td>
<td>18.25</td>
</tr>
<tr>
<td>C</td>
<td>10</td>
<td>18.52</td>
<td>19.08</td>
</tr>
</tbody>
</table>

$$\hat{\mu}_S(\bar{z}) = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_h \bar{z} = 10.86 + 0.19\bar{z} = 20.15$$

$$\hat{\mu}_F(\bar{z}) = \hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_h \bar{z} = 8.96 + 0.19\bar{z} = 18.25$$

$$\hat{\mu}_C(\bar{z}) = \hat{\beta}_0 + \hat{\beta}_h \bar{z} = 9.79 + 0.19\bar{z} = 19.08$$

(g) This question omitted due to time constraints: Consider an analysis where height $z$ is ignored: use one-factor analysis of variance to test the hypothesis that the mean yields are equal for the three fertilizer treatments. Note that $F(0.05, 2, 27) = 3.35$

$$SS[Trt] = 11.9$$


$$F = \frac{11.9/2}{54.5/27} = 2.95$$

"The observed $F$-ratio does not exceed the critical value and the fertilizer means do not differ significantly in an analysis that does not control for height, $z"
proc glm order=data;
class trt;
model y=trt z/solution;
means trt;
lsmeans trt/stderr;

The GLM Procedure

Class	Levels	Values
trt	3	S F C

Sum of
Source	DF	Squares	Mean Square	F Value	Pr > F
Model	3	45.53934660	15.17978220	18.94	<.0001
Error	26	20.83432007	0.80132000
Corrected Total	29	66.37366667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
trt	2	11.92466667	5.96233333	7.44	0.0028
z	1	33.61467993	33.61467993	41.95	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	2	16.50499575	8.25249787	10.30	0.0005
z	1	33.61467993	33.61467993	41.95	<.0001

Standard
Parameter	Estimate	Error	t Value	Pr > |t|
Intercept	9.792157111 B	1.37696313	7.11	<.0001
trt	S	1.073260162 B	0.40676418	2.64	0.0139
trt	F	-0.837780957 B	0.46309641	-1.81	0.0820
trt	C	0.000000000 B	. . ..
z	0.192074007	0.02965563	6.48	<.0001

NOTE: The X'X matrix has been found to be singular, and a generalized
inverse was used to solve the normal equations. Terms whose
estimates are followed by the letter 'B' are not uniquely estimable.

Level of ---------y---------
trt	N	Mean	Std Dev	N	Mean	Std Dev
S	10	20.0600000	1.44775689	10	47.8700000	4.85456257
F	10	19.1900000	1.33287492	10	53.2900000	6.69949418
C	10	18.5200000	1.47557898	10	45.4400000	5.72619711

Standard
trt	y LSMEAN	Error	Pr > |t|
S	20.2514338	0.2846148	<.0001
F	xx.xxxxxxxx	0.3119925	<.0001
C	19.1781736	0.3007634	<.0001

3
2. A researcher in food science conducts a randomized experiment to study the effect of blanching time on vitamin C content of shelled soybeans. She randomizes \( N = 12 \) serving-size allotments of soybeans to \( t = 4 \) blanching time treatments and measures the amount of vitamin C in \( mg \), in each allotment. The data are summarized below:

<table>
<thead>
<tr>
<th>Blanching Time</th>
<th>Sample Size, ( n )</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 minute</td>
<td>3</td>
<td>29.8</td>
<td>0.46</td>
<td>0.21</td>
</tr>
<tr>
<td>2 minutes</td>
<td>3</td>
<td>26.0</td>
<td>0.44</td>
<td>0.19</td>
</tr>
<tr>
<td>3 minutes</td>
<td>3</td>
<td>24.0</td>
<td>0.40</td>
<td>0.16</td>
</tr>
<tr>
<td>4 minutes</td>
<td>3</td>
<td>23.0</td>
<td>0.53</td>
<td>0.28</td>
</tr>
</tbody>
</table>

\[ \text{avg} = 25.7 \]

Let the vitamin C content of the \( j^{th} \) allotment in the \( i^{th} \) time group be denoted \( y_{ij} \). An incomplete ANOVA table appears on the next page.

(a) Using the fact that \( \sum_{i=1}^{4} (\bar{y}_{i+} - 25.7)^2 = 27.08 \), report the sum of squares for treatment in a one-factor analysis of variance of these data.

\[
SS[\text{Trt}] = \sum_{i=1}^{4} \sum_{j=1}^{5} (\bar{y}_{i+} - \bar{y}_{++})^2
\]
\[
= 3(27.08) = 81.24
\]

(b) Using either the table above, or the fact that \( \sum_{i=1}^{4} \sum_{j=1}^{3} (\bar{y}_{ij} - \bar{y}_{i+})^2 = 1.68 \), report the mean square for error.

\[
MS[E] = \frac{SS[E]}{t(n-1)}
\]
\[
= 1.68/8 = 0.21
\]

(c) Sketch a plot of the treatment means versus blanching time.
(d) Formulate and test the hypothesis that blanching time has no effect on vitamin C, using level of significance $\alpha = 0.05$. (Specify a model for $Y_{ij}$ and corresponding hypotheses, $H_0$ and $H_1$.) You may use one of the following critical values: $F(0.05, 3, 3) = 9.28$, $F(0.05, 3, 8) = 4.07$. A partial ANOVA table is provided for convenience.

Model: $Y_{ij} = \mu_i + E_{ij}$

where $E_{ij} \overset{iid}{\sim} N(0, \sigma^2)$. The hypotheses are

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_5$ vs $H_1 : \text{not all equal}$

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>81.24</td>
<td>27.08</td>
<td>3</td>
<td>128.95</td>
</tr>
<tr>
<td>Error</td>
<td>1.68</td>
<td>0.21</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>82.92</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Since $F_{obs} = 128.95 > 4.07 = F^*$, $H_0$ is rejected and the blanching time has a significant (deleterious) effect on vitamin C content.

(e) Consider a simple linear regression model which takes mean vitamin C content to be linear in blanching time. In testing for lack-of-fit of this model, what are the degrees of freedom associated with lack-of-fit and pure error?

$df_{LOF} = 3 - 1 = 2$, $df_{PE} = N - 4 = 8$

(f) The regression sum of squares from such a model is $SS[R] = 75.3$. Obtain the $F$-ratio for a test of lack of fit.

$$F = \frac{SS[LOF]/2}{MS[PE]} = \frac{(SS[trt] - SS[R])/2}{MS[PE]} = 14.23$$

(g) Draw a conclusion based on the $F$-ratio for lack-of-fit, using significance level $\alpha = 0.05$. Interpret this result as it pertains to the researcher’s analysis of the effect of blanching time on vitamin C content in shelled soybeans. Critical values of potential interest include $F(0.05, 1, 12) = 4.75$, $F(0.05, 2, 8) = 4.46$, $F(0.05, 3, 3) = 9.28$. Since $F_{obs} = 14.23 > 4.46 = F(0.05, 2, 8)$, there is evidence of lack-of-fit and the linear regression model is inadequate.

(h) Obtain sums of squares for linear and quadratic orthogonal polynomial contrasts of the equally space blanching time means.

$$SS(\hat{\theta}_1) = \frac{(-3\bar{y}_1 - 1\bar{y}_2 + \bar{y}_3 + 3\bar{y}_4)^2}{(20/3)}$$

$= 75.3 = SS[R]$  

$$SS(\hat{\theta}_2) = \frac{(\bar{y}_1 - 1\bar{y}_2 - \bar{y}_3 + \bar{y}_4)^2}{(20/3)}$$

$= 5.9 = R(\beta_2|\beta_1, \beta_0)$ (in a quadratic regression)
3. A nutrition experiment is undertaken to investigate the effect of $t = 4$ different rations of food on growth in young rats. In a balanced design, $n = 10$ rats were randomly assigned to each of the $t = 4$ groups, for a total sample size of $N = 40$. Some summary statistics are

$$SS[E] = 325.6, \quad MS[E] = 9.04$$

$$SS[Trt] = 372.3$$

(a) If interest lies in all pairwise comparisons among these means, which multiple comparisons procedures is more powerful? (Circle one: [Tukey] / Scheffé).

(b) Using any multiple comparison procedure that controls the familywise error rate $\alpha = 0.05$, identify which of the $\binom{4}{2} = 6$ pairwise differences are significant in the table below. You may use the fact $t(0.025/4, 36) = 2.79$ and $t(0.025/6, 36) = 2.81$ and $q(0.05, 4, 36) = 3.81$ and $F(0.05, 3, 36) = 2.87$.

<table>
<thead>
<tr>
<th>Ration group</th>
<th>$n$</th>
<th>Mean</th>
<th>MCP grouping</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>$\bar{y}_{1+} = 7.9$</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>$\bar{y}_{2+} = 11.7$</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>$\bar{y}_{3+} = 11.7$</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>$\bar{y}_{4+} = 16.5$</td>
<td>C</td>
</tr>
</tbody>
</table>

Two means with the same letter do not differ significantly.

Using Tukey,

$$HSD = q(0.05, 4, 36)\sqrt{MS[E]/n} = 3.81\sqrt{9.04/10} = 3.62$$

Using Scheffé,

$$MSD = \sqrt{(4 - 1)F(0.05, 4, 36)MS[E](2/n)} = \sqrt{3 \times 2.879.04(2/10)} = 3.95$$

Using Bonferroni,

$$MSD = t(0.025/4, 4, 36)\sqrt{MS[E](2/n)} = 2.81\sqrt{9.04(2/10)} = 3.75$$

(c) Define the familywise error rate for this analysis.

$$fwe = \alpha = 0.05 = Pr( \text{at least one type I error})$$
4. A researcher will conduct an experiment, using level of significance $\alpha = 0.05$, to investigate possible differences in coverages ($y$) of $t = 3$ species of plants grown on greenroofs. She will partition a greenroof into $3 \times n$ sections and randomize these sections to the three species. If she were to repeat the experiment many times, she thinks she would see the means ($\mu_i$) and standard deviations ($\sigma_i$) given in the (RESEARCH hypothesis) table below.

<table>
<thead>
<tr>
<th>Species</th>
<th>$i$</th>
<th>$\mu_i$</th>
<th>$\sigma_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delosperma</td>
<td>1</td>
<td>41</td>
<td>10</td>
</tr>
<tr>
<td>Reflexum</td>
<td>2</td>
<td>53</td>
<td>10</td>
</tr>
<tr>
<td>Album</td>
<td>3</td>
<td>41</td>
<td>10</td>
</tr>
</tbody>
</table>

(a) Under the competing NULL hypothesis that all treatment means are equal, what is the probability that she will declare that the sample means differ significantly?

$\alpha = 0.05$

(b) Suppose she uses $n = 5$ sections per species. Specify the sampling distribution of the $F$-ratio, $MS[Trt]/MS[E]$, under the RESEARCH hypothesis specified in the table above.

*Non-central $F$ distribution w/ $df = 2, 12$ and $\gamma = 0.96 \times 5 = 4.8$*

(c) Use the plot below to find the power of the test with $n = 5$ sections per species.

$1 - \beta \approx 0.5$

(d) What would the type II error rate be if she increases the sample size to $n = 7$?

$\beta \approx 0.35$