1. Rao 12.3a A sample of \( n = 30 \) subjects were randomly assigned to three therapies/treatments for improving mental capacity. On each subject, pretest (\( Z \)) and posttest (\( Y \)) measurements were made. The data are available on the ST512 website as “raoeg12.1.dat.” The code below should suffice for reading them in:

```plaintext
data scores;
  do therapy=1 to 3; /* loops over three pairs of columns */
    input z y @; /* @ prevents datastep from reading new record */
    output; /* write current values of variables to datastep */
  end;
run;
```

(12.3a) Construct the one-way ANOVA table for comparing the three treatment means when \( z \) is ignored.

```plaintext
proc glm;
  class therapy;
  model y=therapy;
  means therapy; /* will add treatment means to the output */
run;
```

(a) Conduct an \( F \)-test for equality of means. That is, specify a model and a null hypothesis for no therapy effect, then compute the \( F \)-ratio

\[
F = \frac{MS(\text{trt})}{MS(E)}, \quad df = 2, 27
\]

and compare it to the critical value \( F(.05, 2, 27) \).

(b) Plot \( y \) versus \( z \) with a different symbol for each therapy.

(c) By adding \( z \) to the `model` statement in PROC GLM or PROC REG, fit the following analysis of covariance (ANCOVA) model:

\[
Y_i = \beta_0 + \beta_1 x_{i1} + \beta_1 x_{i2} + \beta z_i + E_i
\]

where \( X_{i1} \) and \( x_{i2} \) are indicator variables for therapies 1 and 2 respectively. (Make sure not to declare \( z \) as a `class` variable in PROC GLM.) Report each regression coefficient along with a standard error.

(d) Report the \( F \)-test for a therapy effect, after controlling for the effect of the pretreatment (\( z \)) score.

(e) Report the unadjusted post-test scores for the three therapies.

(f) Report the adjusted post-test score for the three therapies, along with standard errors.
2. Consider designing an experiment to evaluate the potential effectiveness of $t = 5$ weight reducing agents. Suppose that $n$ subjects are to be assigned at random to each of $t = 5$ treatment groups. Suppose that the smallest meaningful effect on weight loss that researchers involved in the study would like to detect is one in which the variance among the weight loss treatment means is at least as great as those given in the alternative below:

$$H_1: \mu_A = \mu_E = 12, \mu_B = 11, \mu_C = 10, \mu_D = 9$$

Suppose further that the standard deviation among weight losses for any treatment group is about $\sigma = 1$. Hold the type I error rate at $\alpha = 0.05$.

(a) Compute the number of subjects necessary to obtain a power of at least $1 - \beta = 0.9$.

(b) Obtain a plot of the power against sample sizes between 2 and 10.

(c) Describe how the power would change if $\sigma$ were actually larger.

(d) Describe how the power would change if the population mean weight gain for agent 1 were $\mu_1 = 15$.

(e) Suppose the $n = 10$ is adopted, and data are observed as given in the table below.

<table>
<thead>
<tr>
<th>Weight losses under five agents</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>12.05</td>
<td>11.02</td>
<td>10.27</td>
<td>9.27</td>
<td>12.17</td>
</tr>
<tr>
<td>std.dev.</td>
<td>.83</td>
<td>1.12</td>
<td>1.03</td>
<td>1.16</td>
<td>0.79</td>
</tr>
</tbody>
</table>

i. Carry out a test for the hypothesis that the treatment has no effect on weight loss. Use $\alpha = 0.05$

ii. After carrying out all pairwise comparisons at familywise error rate $\alpha = .05$, identify which differences are significant. Be clear about which multiple comparison procedure you use.

iii. Explain the difference between strong and weak control of the familywise error rate in this context.

iv. Consider the complex contrast that compares the mean of agents $A$ and $E$ with that of agent $D$.

A. Express this (population) contrast as a vector product involving the vector of (population) treatment means, $\mu' = (\mu_A, \mu_B, \mu_C, \mu_D, \mu_E)$.

B. Report an estimate of the contrast from the data summarized in the table.

C. Report a standard error.

D. Report the sum of squares associated with the contrast.
3. (This problem is just for fun. No need to turn anything in.) Click the link on the ST511 website that takes you to the page of applets by Webster at Univ. of South Carolina. Then click the “Let’s Make a Deal” link.

http://www.stat.tamu.edu/~west/applets/LetsMakeaDeal.html

Read the description of the problem. Now, consider an experiment where you get to observe \( n \) independent 0 - 1 trials and you are interested in competing hypotheses for the success probability \( p \):

\[
H_0 : p = \frac{1}{2} \quad \text{versus} \quad H_1 : p = \frac{2}{3}
\]

Let \( Y \) denote the number of successes out of the \( n \) trials.

(a) Suppose you observe \( n = 30 \) such trials and adopt this critical region for \( Y \):

Reject \( H_0 \) if \( Y \geq 20 \).

Using Table C.5 or an appropriate BINOMIAL applet, obtain the exact significance level of this test.

(b) Using the applet by Lenth, find either the approximate or exact power of the test which uses the critical region we’ve specified.

www.stat.uiowa.edu/~rlenth/Power/.

(c) Collect your data using the switch strategy. Play the game \( n = 30 \) times and report your results. Are your data “statistically significant”, in terms of the test you’ve set up? Briefly state your conclusion regarding \( H_0 \) and \( H_1 \), using \( \alpha \) from part a).
4. Recall the orange yields problem from Quiz 1, where the estimated regression vector for the five varieties that was given by the output was

```
proc reg;
  model y1=xa xb xc xd xe/covb;
run;
```

The SAS System
The REG Procedure

```
Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>4</td>
<td>988.40000</td>
<td>247.10000</td>
<td>3.31</td>
<td>0.0233</td>
</tr>
<tr>
<td>Error</td>
<td>30</td>
<td>2240.00000</td>
<td>74.66667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>34</td>
<td>3228.40000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Model is not full rank. Least-squares solutions for the parameters are not unique. Some statistics will be misleading. A reported DF of 0 or B means that the estimate is biased.

NOTE: The following parameters have been set to 0, since the variables are a linear combination of other variables as shown.

\[
XE = \text{Intercept} - \text{XA} - \text{XB} - \text{XC} - \text{XD}
\]

```
| Variable     | DF | Estimate | Error | t Value | Pr > |t| |
|--------------|----|----------|-------|---------|------|---|
| Intercept    | B  | 31.00000 | 3.26599| 9.49    | <.0001|
| XA           | B  | -7.00000 | 4.61880| -1.52   | 0.1401|
| XB           | B  | 3.00000  | 4.61880| 0.65    | 0.5209|
| XC           | B  | 7.00000  | 4.61880| 1.52    | 0.1401|
| XD           | B  | -6.00000 | 4.61880| -1.30   | 0.2038|
| XE           | 0  | 0        | .      | .       | .    |
```

The estimate is not unique. Obtain another estimate that leads to the same fitted values (\(\hat{Y}\)) and hence, the same ANOVA table.