ST 512 Homework assignment #3 (more problems may be added)

Dr. Jason A. Osborne

1. Rao 9.17

   (a) Rao 9.17a - there are $k = 4$ contrasts of interest,

$$\begin{aligned}\theta_1 &= \mu_1 - \frac{\mu_2 + \mu_3}{2}\\ \theta_2 &= \mu_1 - \mu_2\\ \theta_3 &= \mu_2 - \mu_3\\ \theta_4 &= \frac{\mu_1 + \mu_2}{2} - \mu_3\end{aligned}$$

   To obtain simultaneous 95% (or 90%)confidence intervals for these $k = 4$ contrasts using the Bonferroni adjustment, $t(\frac{0.05}{8}, 9) = 3.111$ (or $t(\frac{0.05}{4}, 9) = 2.685$) is needed. This could be stolen from table C.10, p827. Alternatively, the TINV function in SAS or the web APPLETS could be used to get this. Anyhoo, the confidence intervals all take the form

$$\hat{\theta}_i \pm 3.111\widehat{SE}(\hat{\theta}_i) \text{ for simultaneous95\% confidence}$$

$$\hat{\theta}_i \pm 2.685\widehat{SE}(\hat{\theta}_i) \text{ for simultaneous90\% confidence}$$

   yielding with simultaneous 95% confidence,

$$\begin{array}{lcc} 0.23 \pm 3.111(0.046) & \text{or} & (0.09, 0.37)\\ 0.15 \pm 2.111(0.053) & \text{or} & (-0.02, 0.31)\\ 0.17 \pm 2.111(0.053) & \text{or} & (0.00, 0.34)\\ 0.24 \pm 2.111(0.046) & \text{or} & (0.10, 0.39)\end{array}$$

   or with 90% confidence,

$$\begin{array}{lcc} 0.23 \pm 2.685(0.046) & \text{or} & (0.11, 0.35)\\ 0.15 \pm 2.685(0.053) & \text{or} & (0.00, 0.29)\\ 0.17 \pm 2.685(0.053) & \text{or} & (0.03, 0.31)\\ 0.24 \pm 2.685(0.046) & \text{or} & (0.12, 0.37)\end{array}$$

(b) Rao 9.17c

```
                       The GLM Procedure
          Tukey's Studentized Range (HSD) Test for y
 Comparisons significant at the 0.05 level are indicated by ***.

                    Difference
       trt             Between     Simultaneous 95%
   Comparison            Means    Confidence Limits

   1    - 2            0.14500    -0.00394  0.29394
   1    - 3            0.31500     0.16606  0.46394  ***
   2    - 1           -0.14500    -0.29394  0.00394
   2    - 3            0.17000     0.02106  0.31894  ***
   3    - 1           -0.31500    -0.46394 -0.16606  ***
   3    - 2           -0.17000    -0.31894 -0.02106  ***

               Bonferroni (Dunn) t Tests for y

   1    - 2            0.14500    -0.01148  0.30148
   1    - 3            0.31500     0.15852  0.47148  ***
   2    - 1           -0.14500    -0.30148  0.01148
   2    - 3            0.17000     0.01352  0.32648  ***
   3    - 1           -0.31500    -0.47148 -0.15852  ***
   3    - 2           -0.17000    -0.32648 -0.01352  ***

                    Scheffe's Test for y

   1    - 2            0.14500    -0.01065  0.30065
   1    - 3            0.31500     0.15935  0.47065  ***
   2    - 1           -0.14500    -0.30065  0.01065
   2    - 3            0.17000     0.01435  0.32565  ***
   3    - 1           -0.31500    -0.47065 -0.15935  ***
   3    - 2           -0.17000    -0.32565 -0.01435  ***
```

2. Using the data from Rao 10.2, test the adequacy of the simple linear regression model in which the mean soil water content is linear in depth. Which model would you select, the linear regression model or the one-factor ANOVA model with 4 treatment means?

$$\begin{aligned}\text{SLR Model} \quad \mu(x) &= \beta_0 + \beta_1 x \\ \text{One-factor ANOVA Model} \quad \mu_{ij} &= \mu + \tau_i\end{aligned}$$

$$\begin{aligned} F &= \frac{SS[\text{lack-of-fit}]/(3-1)}{MS[E]_{full}} \\ &= \frac{(SS[R]_{full} - SS[R]_{red})/(3-1)}{MS[E]_{full}} \\ &= \frac{0.00024011/2}{0.000173888} \\ &= 0.69. \end{aligned}$$

This F-ratio is not significant on $2, 12$ degrees of freedom. There is no evidence of lack-of-fit of the simple linear regression model.

SAS was used to fit 1) the SLR 2) the one-factor ANOVA model and 3) the model in which a class variable `xlof` is added to the (reduced) SLR to get the (full) one-factor ANOVA model. Output follows:

The REG Procedure

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.00752	0.00752	45.28	<.0001
Error	14	0.00233	0.00016619		
Corrected Total	15	0.00985			

The GLM Procedure

Class	Levels	Values
x	4	10 40 60 120

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.00776450	0.00258817	14.89	0.0002
Error	12	0.00208650	0.00017388		
Corrected Total	15	0.00985100			

```
                     The GLM Procedure
           Class         Levels     Values
           xlof               4     10 40 60 120

                                  Sum of
Source                DF         Squares    Mean Square  F Value  Pr > F
Model                  3      0.00776450     0.00258817    14.89  0.0002
Error                 12      0.00208650     0.00017388
Corrected Total       15      0.00985100

Source                DF     Type I SS    Mean Square  F Value  Pr > F
x                      1      0.00752439     0.00752439    43.27  <.0001
xlof                   2      0.00024011     0.00012006     0.69  0.5202
```

3. Consider the experiment described as Rao Example 8.2 (p. 280-281). Let the five treatment means be denoted $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$. Consider these four contrasts:

$$\begin{aligned}
\theta_1 &= \mu_2 + \mu_3 - \mu_4 - \mu_5 \\
\theta_2 &= \mu_2 - \mu_3 + \mu_4 - \mu_5 \\
\theta_3 &= \mu_2 - \mu_3 - \mu_4 + \mu_5 \\
\theta_4 &= \mu_1 - \frac{1}{4}(\mu_2 + \mu_3 + \mu_4 + \mu_5)
\end{aligned}$$

(a) Is this set of contrasts mutually orthogonal? YES.

(b) Compute the sum of squares associated with each contrast.

$$\begin{aligned}
SS(\theta_1) &= \frac{(-5.86)^2}{4/4} &= 34.3 \\
SS(\theta_2) &= \frac{(-1.075)^2}{4/4} &= 1.16 \\
SS(\theta_3) &= \frac{(1.045)^2}{4/4} &= 1.09 \\
SS(\theta_4) &= \frac{(1.185)^2}{\frac{1}{4}(1^2+4(1/4)^2} &= 4.49
\end{aligned}$$

(c) Compute the sum of the four sums of squares computed in part (a) $\sum_i SS(\theta_i) = 41.1$

(d) Compute the treatment sum of squares in the ANOVA. $SS(Trt) = 41.1$

(e) Briefly describe the "effect" being estimated by each contrast, using language of the experiment.
Respectively, the main effect of light type, the main effect of intensity, the light type-by-intensity interaction, the difference between dark and the average of the four lighted treatments

4. Consider designing an experiment to evaluate the potential ...

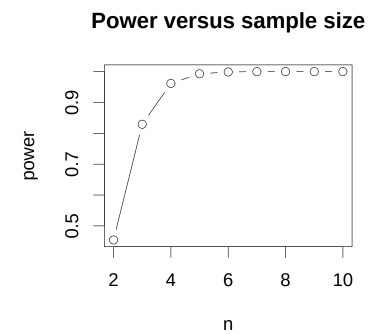
   (a) With $n = 3$ or $n = 4$, the critical values are $F(.05, 4, 10) = 3.48$ and $F(.05, 4, 15) = 3.06$, respectively. The noncentrality parameters are

   $$\gamma = n\frac{\sum \tau_i^2}{\sigma^2} = 6.8n = 13.6(n = 3) \text{ or } 20.4(n = 4)$$

   which is large enough to get the type II error rate β down to something less than 0.1

   $$\begin{aligned} \Pr(F < 3.48; \nu_1 = 4, \nu_2 = 10, \gamma = 20.4) &= .17 > 0.1 \\ \Pr(F < 3.05; \nu_1 = 4, \nu_2 = 15, \gamma = 27.2) &= .038 < 0.1 \end{aligned}$$

   So, $n = 4$ is big enough.

   (b)

   (c) Power goes down as σ increases.

   (d) If $\mu_A = 15$ instead of $\mu_A = 12$ while everything else remained the same, the noncentrality parameter would increase and so would the power.

   (e)

   $$F = \frac{SS[Trt]/(5-1)}{MS[E]} = \frac{58.6/4}{0.986} = 14.9(\text{Significant with } p < 0.0001)$$

   Using Tukey's $HSD = 1.27$ with $fwe = .05$, we see that means with the same superscripts in the table below do not differ significantly:

   i. Weight losses under five agents

	D	C	B	A	E
mean	9.27	10.27	11.02	12.05	12.17
grouping	z	zx	xy	y	y

   ii. Strong control means protection under any alternative configuration of the treatment means. Weak control means protection only under the null hypothesis where all treatment means are equal.