ST512-R

Some problems that will be on Homework #4

1. (From exam 2, Fall, 2005) An experiment measures "Ortho-P" reduction after running material through a centrifuge and adding either lime only (L) or an experimental floculant only (F), or both (LF) or neither (C). $N = 12$ total samples are randomized to the four treatment combinations and run through the centrifuge. (Data are available as "centrifuge.dat".)

   (a) Obtain the sum of squares for treatments (based on 3 df).
   Let i, j index treatment combination and sample, respectively.

$$\begin{aligned} SS(Trt) &= \sum_i \sum_j (\bar{y}_{i+} - \bar{y}_{++})^2 \\ &= 3[(59-63)^2 + (61-63)^2 + (50-63)^2 + (82-63)^2] = 1650 \end{aligned}$$

   (b) Obtain sums of squares for three single degree-of-freedom contrasts to quantify

   - interaction between lime and floculant ($SS(l * f)$)

$$\begin{aligned} \hat{\theta}_{l*f} &= (82-61) - (50-59) = 30 \\ SS(\hat{\theta}_{l*f}) &= \frac{30^2}{4/3} = 675 \end{aligned}$$

   - main effect of lime $SS(l)$

$$\begin{aligned} \hat{\theta}_{lime} &= \frac{1}{2}(50+82) - \frac{1}{2}(59+61) = 6 \\ SS(\hat{\theta}_{lime}) &= \frac{6^2}{1/3} = 108 \end{aligned}$$

   - main effect of floculant $SS(f)$.

$$\begin{aligned} \hat{\theta}_{floc} &= \frac{1}{2}(61+82) - \frac{1}{2}(59+50) = 17 \\ SS(\hat{\theta}_{floc}) &= \frac{17^2}{1/3} = 867 \end{aligned}$$

   A check: $SS[Trt] = 1650 = 675 + 108 + 867$

   (c) Does adding lime improve Ortho-P reduction? Obtain 95% confidence intervals for the simple lime effects in the absence or presence of the floculant.

$$\begin{aligned} \hat{\mu}[\text{lime no floculant}] &= 50 - 59 = -9 \\ SE(\hat{\mu}[\text{lime no floculant}]) &= \sqrt{MSE(2/3)} = 5.2 \end{aligned}$$

   95% confidence interval: $-9 \pm 2.31(5.2)$ or -9 ± 12

$$\begin{aligned} \hat{\mu}[\text{lime with floculant}] &= 82 - 61 = 21 \\ SE(\hat{\mu}[\text{lime no floculant}]) &= \sqrt{MSE(2/3)} = 5.2 \end{aligned}$$

   95% confidence interval: $21 \pm 2.31(5.2)$ or 21 ± 12

(d) Obtain an interaction plot to convey the effects of the treatments on Ortho-P reduction.

2. Rao 13.10

(a)

$$F = \frac{MS(Trt)}{MS(E)} = \frac{493}{0.35} = 1407(df = 3, 12, p < .0001)$$

Conclude that the sample treatment means, $\bar{y}_{11} = 10.3, \bar{y}_{12} = 19.6, \bar{y}_{21} = 10.7, \bar{y}_{22} = 34$ differ significantly.

(b) $\hat{mu}[AB] = \bar{y}_{12} - \bar{y}_{11} - (\bar{y}_{22} - \bar{y}_{21} = 14.1$ (also ok to divide this difference of slopes by two, the p-value for a test of additivity is the same)

(c) $t = 27.7$ on $df = 12$, highly significant.

3. Rao 13.14 Well, one example of how it can be misleading to talk about main effects is if you're talking about the differences among the varieties; there are none if when pretreatment 3 is used, but averaging over pretreatments suggests a variety effect.

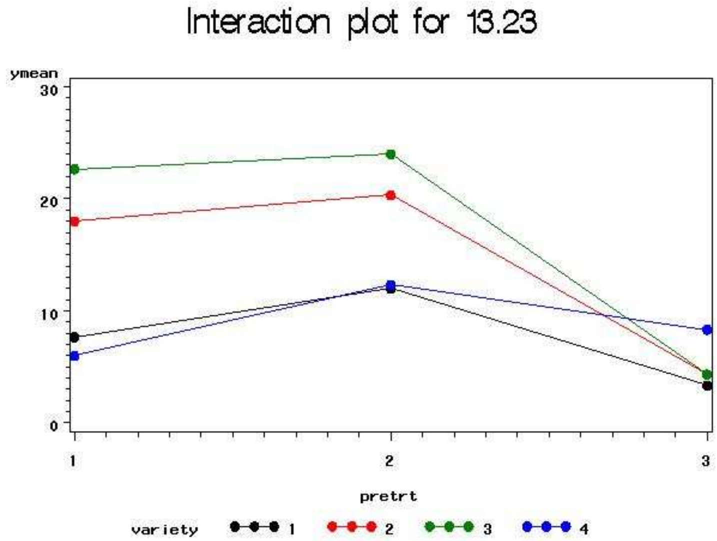
4. Rao 13.23

(a)

```
                    The SAS System                              1
                   The GLM Procedure
                Class Level Information
            Class         Levels    Values
            pretrt             3    1 2 3
            variety            4    1 2 3 4

                                      Sum of
Source                    DF         Squares    Mean Square  F Value  Pr > F
Model                     11     1872.555556     170.232323     8.52  <.0001
Error                     24      479.333333      19.972222
Corrected Total           35     2351.888889
```

(b) Looks like interaction, with pretreatment effects that are most pronounced for varieties 2 and 3. (Pretreatment 3 is not so great, particularly for these two varieties.)



(c) ANOVA table:

```
                    The SAS System                          1
                   The GLM Procedure
                Class Level Information
            Class          Levels    Values
            pretrt              3    1 2 3
            variety             4    1 2 3 4

                                    Sum of
Source                     DF      Squares   Mean Square  F Value  Pr > F
Model                      11  1872.555556    170.232323     8.52  <.0001
Error                      24   479.333333     19.972222
Corrected Total            35  2351.888889

Source                     DF  Type III SS   Mean Square  F Value  Pr > F
pretrt                      2  924.3888889   462.1944444    23.14  <.0001
variety                     3  525.4444444   175.1481481     8.77  0.0004
pretrt*variety              6  422.7222222    70.4537037     3.53  0.0120
```

The interaction is significant. Inspection of the plot indicates that there are significant variety effects, but only when using pretreatments 2 and 3. For pretreatments 1 and 2, it looks like varieties 2 and 3 yield longer root lengths, on average. I tested for pretreatment effects separately for each variety:

```
                      The GLM Procedure
                     Least Squares Means

         pretrt*variety Effect Sliced by variety for y

                             Sum of
    variety       DF        Squares     Mean Square    F Value     Pr > F
    1              2     112.666667       56.333333       2.82     0.0794
    2              2     448.222222      224.111111      11.22     0.0004
    3              2     724.666667      362.333333      18.14     <.0001
    4              2      61.555556       30.777778       1.54     0.2346
```

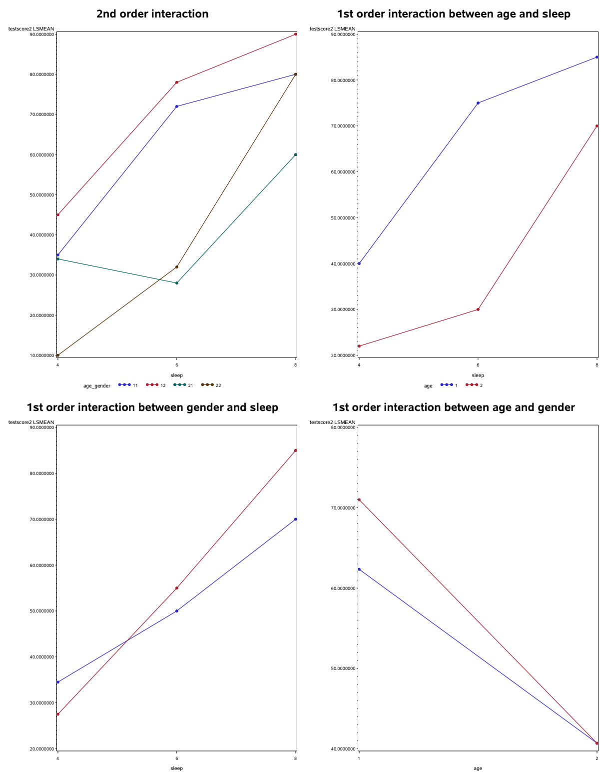
Since the effects were only significant for varieties 2 and 3, I carried out pairwise comparisons for within these two varieties. If a Scheffe adjustment is used, the only differences that are significant are those involving pretreatment 3 and variety 3, which yields significantly lower root lengths than pretreatments 1 or 2:

```
V2: V1 - V3
V3: V2 - V3
```

(Assignment is extended here, updating previous version of HW4, which was shorter.)

5. Consider the data from exercise 13.31 (p. 632).
(See the code entitled "sleep.sas"). Note that while the means for all $t = abc = 12$ treatment combinations are given, the triplicate ($n = 3$) observations are not. The error mean square from the full ANOVA is $MS(E) = 143$ on $df = 24$.

(a)



(b) Without taking into account the sampling variability, it may not be wise to make assessments of interaction (or any) effects from plots alone. (See part (c).)

(c) The GLM Procedure

Class	Levels	Values
age	2	1 2
gender	2	1 2
sleep	3	4 6 8

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	22562.00000	2051.09091	14.34	<.0001
Error	24	3432.00000	143.00000		
Corrected Total	35	25994.00000			

R-Square	Coeff Var	Root MSE	testscore2 Mean
0.867970	22.28247	11.95826	53.66667

Source	DF	Type III SS	Mean Square	F Value	Pr > F
age	1	6084.00000	6084.00000	42.55	<.0001
gender	1	169.00000	169.00000	1.18	0.2878
age*gender	1	169.00000	169.00000	1.18	0.2878
sleep	2	12998.00000	6499.00000	45.45	<.0001
age*sleep	2	1638.00000	819.00000	5.73	0.0093
gender*sleep	2	728.00000	364.00000	2.55	0.0994
age*gender*sleep	2	776.00000	388.00000	2.71	0.0866

At level $\alpha = .05$, the only interaction effect for which there is enough evidence in the experiment to reject the null hypothesis of no population interaction effect in the population is the first-order interaction between age and sleep ($p = .0093, df = 2, 24$). There is no evidence of any effects due to gender. I won't look at the main effect of sleep, as the effect of sleep appears to depend on age.

(d) No conclusions were made in part (b)

(e) For young folks, the mean test score at sleep=4 is significantly lower than at sleep=6 or sleep=8.
For older folks, the mean test score at sleep=8 is significantly higher than at sleep=4 or sleep=6.

Obs	Effect	age	sleep	_age	_sleep	Estimate	StdErr	DF	tValue	Probt	Adjustment	Adjp
1	age*sleep	1	4	1	6	-35.0000	6.9041	24	-5.07	<.0001	Tukey	0.0004
2	age*sleep	1	4	1	8	-45.0000	6.9041	24	-6.52	<.0001	Tukey	<.0001
6	age*sleep	1	6	1	8	-10.0000	6.9041	24	-1.45	0.1604	Tukey	0.6984
13	age*sleep	2	4	2	6	-8.0000	6.9041	24	-1.16	0.2580	Tukey	0.8515
14	age*sleep	2	4	2	8	-48.0000	6.9041	24	-6.95	<.0001	Tukey	<.0001
15	age*sleep	2	6	2	8	-40.0000	6.9041	24	-5.79	<.0001	Tukey	<.0001

(f) It is ok to use Tukey's procedure for post-hoc comparisons.

(g) Estimate the two orthogonal polynomial contrasts for the effect of sleep deprivation on young folks: linear (θ_1) and quadratic (θ_2). (Average over gender.) Report the sums of squares for these two contrasts.

$$\begin{aligned}\hat{\theta}_1 &= -1\bar{y}_{1+1+} + \phantom{(-2)\bar{y}_{1+2+}} + 1\bar{y}_{1+3+} = 45\\ \hat{\theta}_2 &= 1\bar{y}_{1+1+} + (-2)\bar{y}_{1+2+} + 1\bar{y}_{1+3+} = -25\end{aligned}$$

$$\begin{aligned}SS(\theta_1) &= \frac{45^2}{2/6} = 6075\\ SS(\theta_2) &= \frac{(-25)^2}{6/6} = 625\end{aligned}$$

(h) Repeat part g for older folks (to obtain contrasts θ_3 and θ_4 and their sums of squares.)

$$\begin{aligned}\hat{\theta}_3 &= -1\bar{y}_{2+1+} + \phantom{(-2)\bar{y}_{2+2+}} + 1\bar{y}_{2+3+} = 48\\ \hat{\theta}_4 &= 1\bar{y}_{2+1+} + (-2)\bar{y}_{2+2+} + 1\bar{y}_{2+3+} = 32\end{aligned}$$

$$\begin{aligned}SS(\theta_3) &= \frac{48^2}{2/6} = 6912\\ SS(\theta_4) &= \frac{32^2}{6/6} = 1024\end{aligned}$$

(i) Consider two more contrasts, the linear sleep-by-age interaction, $\theta_5 = \theta_1 - \theta_3$ and the quadratic sleep-by-age interaction, $\theta_6 = \theta_2 - \theta_4$. Report sums of squares. Note that θ_5 and θ_6 are orthogonal and $SS(\theta_5) + SS(\theta_6) = SS(\text{age} \times \text{sleep})$.

$$\begin{aligned}\hat{\theta}_5 &= \hat{\theta}_1 - \hat{\theta}_3 = -3 \\ \hat{\theta}_6 &= \hat{\theta}_1 - \hat{\theta}_3 = 57\end{aligned}$$

$$\begin{aligned}SS(\theta_5) &= \tfrac{(-3)^2}{4/6} = 13.5 \\ SS(\theta_6) &= \tfrac{57^2}{12/6} = 1624.5 \\ \text{sum} &= 1638\end{aligned}$$

6. Consider the unbalanced version of the cholesterol data for the four age-gender cohorts discussed in the notes. Report the contrast vectors of the form $(c_{11}, c_{12}, c_{21}, c_{22})$ needed to obtain the least squares estimates of the marginal means for men, for women and for the four combinations of age and gender. Report the estimates based on the data (fill in the table given on 149 of the lecture packet.)

The LSMEAN for men is the linear combination of the sample treatment means ($\hat{\theta} = \sum_i \sum_j \bar{y}_{ij+}$ with expectation $\mu + \frac{1}{2}(\alpha_1 + \alpha_2) + \beta_1$. To have this expectation, the coefficients must satisfy

$$\begin{aligned}c_{11} + c_{12} &= \frac{1}{2} \\ c_{11} + c_{21} &= 1 \\ c_{21} + c_{22} &= \frac{1}{2}\end{aligned}$$

the variance of any linear combination of treatment means is proportional to

$$\frac{c_{11}^2}{n_{11}} + \frac{c_{12}^2}{n_{12}} + \frac{c_{21}^2}{n_{21}} + \frac{c_{22}^2}{n_{22}}$$

and in this case, as a function of c_{11} it looks like

$$\frac{c_{11}^2}{6} + \frac{(\frac{1}{2} - c_{11})^2}{1} + \frac{(1 - c_{11})^2}{21} + \frac{(c_{11} - 1)^2}{7}$$

This may be minimized by setting the derivative wrt c_{11} equal to 0:

$$\frac{c_{11}^2}{6} - 2(\frac{1}{2} - c_{11}) - 2(1 - c_{11}) + \frac{2(c_{11} - \frac{1}{2}}{7} = 0$$

which has solution $c_{11} = 66/97$ and the LSMEAN is

$$(66/97, 1/2 - 66/97, 1 - 66/97, 66/97 - 1/2)(220.7, 162, 289, 212)' = 251.5$$

. Similarly for women, a contrast with expectation $\mu + \frac{1}{2}(\alpha_1 + \alpha_2) + \beta_2$ which has minimum variance at $c_{11} = 18/97$ yielding the linear combination

$$(18/97, 1/2 - 18/97, -18/97, 18/97 + 1/2)(220.7, 162, 289, 212)' = 183.6$$

For the treatment combinations, the coefficient solutions are given in the table below:

Population group	effect	solution for c_{11}	LSMEAN
young men	$\mu + \alpha_1 + \beta_1$	90/97	221.99
young women	$\mu + \alpha_1 + \beta_2$	42/97	154.1
older men	$\mu + \alpha_2 + \beta_1$	42/97	281.1
older women	$\mu + \alpha_2 + \beta_2$	-6/97	213.1