KEY

**Stat 511 Exam II**

**April 7, 2004** **Prof. Vardeman**

1. Consider a $3\times 3$ factorial analysis with factors A and B under an ordinary (fixed effects) linear model. The effects model under the R baseline restriction has parameter vector for mean responses

$$\gamma = \left(\mu^*, \alpha_2^*, \alpha_3^*, \beta_2^*, \beta_3^*, \alpha\beta_{22}^*, \alpha\beta_{23}^*, \alpha\beta_{32}^*, \alpha\beta_{33}^*\right)'$$

8pts a) Write out all 9 cell means in terms of the entries of $\gamma$ in the table below. Let rows correspond to levels of A (1 to 3 top to bottom) and columns correspond to levels of B (1 to 3 left to right).

| | | |
|---|---|---|
| $\mu^*$ | $\mu^*+\beta_2^*$ | $\mu^*+\beta_3^*$ |
| $\mu^*+\alpha_2^*$ | $\mu^*+\alpha_2^*+\beta_2^*+\alpha\beta_{22}^*$ | $\mu^*+\alpha_2^*+\beta_3^*+\alpha\beta_{23}^*$ |
| $\mu^*+\alpha_3^*$ | $\mu^*+\alpha_3^*+\beta_2^*+\alpha\beta_{32}^*$ | $\mu^*+\alpha_3^*+\beta_3^*+\alpha\beta_{33}^*$ |

8pts b) Give below a matrix $\mathbf{C}$ so that the *testable* hypothesis $\mathrm{H}_0: \mathbf{C}\gamma = \mathbf{0}$ is the hypothesis $\mathrm{H}_0: \beta_j = 0\ \forall j$. (As always, $\beta_j = \mu_{.j} - \mu_{..}$) This is $\mu_{.1} = \mu_{.2} = \mu_{.3}$ i.e.

$$\tfrac{1}{3}(3\mu^*+\alpha_2^*+\alpha_3^*) = \tfrac{1}{3}(3\mu^*+\alpha_2^*+\alpha_3^*+3\beta_2^*+\alpha\beta_{22}^*+\alpha\beta_{32}^*)$$

and $\tfrac{1}{3}(3\mu^*+\alpha_2^*+\alpha_3^*) = \tfrac{1}{3}(3\mu^*+\alpha_2^*+\alpha_3^*+3\beta_3^*+\alpha\beta_{23}^*+\alpha\beta_{33}^*)$

i.e. $\beta_2^* + \tfrac{1}{3}\alpha\beta_{22}^* + \tfrac{1}{3}\alpha\beta_{32}^* = 0$ and $\beta_3^* + \tfrac{1}{3}\alpha\beta_{23}^* + \tfrac{1}{3}\alpha\beta_{33}^* = 0$

So use

$$C = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & \tfrac{1}{3} & 0 & \tfrac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \tfrac{1}{3} & 0 & \tfrac{1}{3} \end{pmatrix}$$

7pts c) Give below a matrix $\mathbf{C}$ so that the *testable* hypothesis $\mathrm{H}_0: \mathbf{C}\gamma = \mathbf{0}$ is the hypothesis that considering only levels 2 and 3 of A and levels 2 and 3 of B (the lower right 4 cells in the table) "there are no interactions" (considering only these 4 cells an interaction plot of means would have the "parallelism" property).

This is $\mu_{23} - \mu_{22} = \mu_{33} - \mu_{32}$ i.e.

$$\beta_3^* - \beta_2^* + \alpha\beta_{23}^* - \alpha\beta_{22}^* = \beta_3^* - \beta_2^* + \alpha\beta_{33}^* - \alpha\beta_{32}^*$$

i.e. $-\alpha\beta_{22}^* + \alpha\beta_{23}^* + \alpha\beta_{32}^* - \alpha\beta_{33}^* = 0$

So use

$$C = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & -1 & 1 & 1 & -1 \end{pmatrix}$$

7pts d) Suppose that all cell sample sizes are 1 and write $\mathbf{Y}' = (y_{11}, y_{12}, y_{13}, y_{21}, y_{31}, y_{22}, y_{23}, y_{32}, y_{33})$. The hypothesis in c) is that $\mathrm{E}\,\mathbf{Y}$ is in some subspace of $\Re^9$ of dimension 8. Write below a $9\times 8$ matrix $\mathbf{X}_0$ so that $C(\mathbf{X}_0)$ is this subspace. (Hint: Note that this hypothesis says nothing at all about 5 of the 9 mean responses, and be careful to pay attention to the order of the entries of $\mathbf{Y}$ listed above.)

The hypothesis is that
$\mu_{33} = -\mu_{22} + \mu_{23} + \mu_{32}$
So we may use

$$\underset{9\times 8}{X_0} = \left( \begin{array}{cccccccc} & & & \underset{8\times 8}{I} & & & & \\ \hline 0 & 0 & 0 & 0 & 0 & -1 & 1 & 1 \end{array} \right)$$

2. A data set of Kaplan, et al. (1972) on the metabolism of sulfisoxazole can be found on page 273 of Bates and Watts. This substance was administered intravenously to a subject and blood samples were taken over time. Concentration of sulfisoxazole, $y$ (in $\mu$g/ml), in the plasma was measured. Bates and Watts suggest an analysis of the data based on a model for concentration as a function of time, $t$ (in minutes from injection),

$$y_i = \theta_1 \exp(-\theta_2 t_i) + \theta_3 \exp(-\theta_4 t_i) + \varepsilon_i \qquad (*)$$

Attached to this exam is an R printout that you should use to answer the following questions.

7pts a) There are actually two different non-linear regression models represented on the R printout. The first is model (*) and the second is the less complex model $y_i = \theta_1 \exp(-\theta_2 t_i) + \varepsilon_i$. What **quantitative** support does the printout provide for the use of model (*) rather than the simpler second model?

The error sum of squares for (*) is about 34.4 compared to about 750.8 for the simpler model. Further, both $\hat{\theta}_1$ and $\hat{\theta}_3$ are statistically significant in model (*). (The simpler model is model (*) with $\theta_1$ or $\theta_3$ equal to 0.)

7pts b) In fitting model (*), what one gets as parameter estimates depends upon one's starting values for $(\theta_1, \theta_2, \theta_3, \theta_4)$. In fact, it is possible to produce exactly the same predicted values and $SSE$ as shown on the printout for a different set of parameter estimates $(\hat{\theta}_1^*, \hat{\theta}_2^*, \hat{\theta}_3^*, \hat{\theta}_4^*)$. How is this already obvious from (*)?

Clearly, the roles played by $(\theta_1, \theta_2)$ and $(\theta_3, \theta_4)$ in model (*) are completely symmetric. If $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4)$ minimizes $(Y-\hat{Y})'(Y-\hat{Y})$, so also will

$$(\hat{\theta}_1^*, \hat{\theta}_2^*, \hat{\theta}_3^*, \hat{\theta}_4^*) = (\hat{\theta}_3, \hat{\theta}_4, \hat{\theta}_1, \hat{\theta}_2)$$

Henceforth base your responses on model (*).

9pts c) Based on model (*), an initial concentration is $\theta_1+\theta_3$. Give approximate 95% confidence limits for this value. (You don't need to do arithmetic, but YOU MUST PLUG IN NUMBERS.)

Use $(\hat{\theta}_1+\hat{\theta}_3) \pm t_8$ (std err for $\hat{\theta}_1+\hat{\theta}_3$)

$$(81.2+162.6) \pm 2.306\sqrt{(1,1)\begin{pmatrix}37.3 & -32.4\\ -32.4 & 48.2\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix}}$$

9pts d) It is potentially of interest to know the value of

$$t_5 = \text{the time at which concentration is } .5(\theta_1+\theta_3)$$

$t_5$ is some function of $(\theta_1,\theta_2,\theta_3,\theta_4)$, say $\tau(\theta_1,\theta_2,\theta_3,\theta_4)$. It is possible to show (don't try to do so) that here

$$\tau(\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3,\hat{\theta}_4)=2.06 \quad\text{and}\quad \nabla\tau=\left(\left.\frac{\partial\tau}{\partial\theta_i}\right|_{(\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3,\hat{\theta}_4)}\right)=\cancel{(14.8,-.0884,1.40,-.00416)}\ (-.01664,-.43523,.08311,240.36)$$

Give approximate 95% confidence limits for $t_5$. (Again, don't do arithmetic, BUT DO PLUG IN.)

Use $\tau(\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3,\hat{\theta}_4) \pm t_8$ (std error for $\tau(\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3,\hat{\theta}_4)$)

$$2.06 \pm 2.306\sqrt{\begin{matrix}(-.017,-.453,.0831,240.4)\\ \cancel{(14.8,-.09,1.4,-.004)}\end{matrix}\begin{pmatrix}37.3 & -.5 & -32.4 & -.04\\ -.5 & .04 & 1.2 & .001\\ 32.4 & 1.2 & 48.2 & .06\\ -.04 & .001 & .06 & .00008\end{pmatrix}\cancel{\begin{pmatrix}14.8\\ -.09\\ 1.4\\ -.004\end{pmatrix}}\begin{pmatrix}-.017\\ -.435\\ .0831\\ 240.4\end{pmatrix}}$$

3. Consider a scenario in which 12 samples of a large lot of material are sent 3 apiece to 4 different labs for hardness testing. At the labs, each specimen is tested twice. For

$$y_{ijk} = \text{the hardness measured on the } k\text{th test of the } j\text{th specimen at the } i\text{th lab}$$

where $i=1,2,3,4$ and $j=1,2,3$ and $k=1,2$, suppose that

$$y_{ijk}=\mu_i+\phi_{ij}+\varepsilon_{ijk} \qquad (**)$$

for constants $\mu_1,\mu_2,\mu_3,\mu_4$, the $\phi_{ij}$ iid $\mathrm{N}(0,\sigma_\phi^2)$ independent of the iid $\mathrm{N}(0,\sigma^2)$ random variables $\varepsilon_{ijk}$.

8pts a) For the $y_{ijk}$ written in dictionary order in $\mathbf{Y}$, what are $\mathbf{X},\boldsymbol{\beta},\mathbf{Z}$, and $\mathbf{u}$ so the model can be written in standard mixed linear model form?

$$\underset{24\times4}{X}=\begin{pmatrix}\underset{6\times1}{1} & 0 & 0 & 0\\ 0 & \underset{6\times1}{1} & 0 & 0\\ 0 & 0 & \underset{6\times1}{1} & 0\\ 0 & 0 & 0 & \underset{6\times1}{1}\end{pmatrix}\quad \beta=\begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\\ \mu_4\end{pmatrix}\quad \underset{24\times12}{Z}=\begin{pmatrix}\underset{2\times1}{1} & & & 0\\ & \underset{2\times1}{1} & & \\ & & \ddots & \\ 0 & & & \underset{2\times1}{1}\end{pmatrix}\quad \underset{12\times1}{\underset{\sim}{u}}=\begin{pmatrix}\phi_{11}\\ \phi_{12}\\ \phi_{13}\\ \phi_{21}\\ \vdots\\ \phi_{23}\\ \vdots\\ \phi_{41}\\ \vdots\\ \phi_{43}\end{pmatrix}$$

The second R printout attached to this exam gives an analysis for 24 hypothetical hardness values under model (**). Use it to answer the following questions.

8pts b) Under the mixed linear model used in this problem, measurements made on the same specimen are correlated. What is an estimate of that correlation?

The covariance between observations on the same specimen is $\sigma_\phi^2$. The variance for any observation is $\sigma_\phi^2+\sigma^2$ So

$$\widehat{\text{correlation}} = \frac{\hat{\sigma}_\phi^2}{\hat{\sigma}_\phi^2+\hat{\sigma}^2} = \frac{(1.17)^2}{(1.17)^2+(.17)^2} = .979$$

8pts c) Give approximate 95% confidence limits for both $\sigma_\phi$ and $\sigma$.

For $\sigma_\phi$:

$(.72, 1.92)$

For $\sigma$:

$(.12, .26)$

8pts d) Give a sensible point (single number) prediction and standard error for the sample mean of two hardness tests made on a 13th specimen sent to Lab 1 for hardness testing. (PLUG IN.)

The obvious prediction for $\mu_1+\phi_{new}+\bar{\epsilon}_{new}$ is $\underbrace{\hat{\mu}_1+0}_{=101.435}$ with standard error

$$\sqrt{(\text{s.e. for } \hat{\mu}_1)^2+\hat{\sigma}_\phi^2+\tfrac{1}{2}\hat{\sigma}^2} = \sqrt{(.68)^2+(1.17)^2+\tfrac{1}{2}(.17)^2}$$

6pts e) R reports "8 degree of freedom" standard errors for its estimates of each of $\mu_1,\mu_2,\mu_3$, and $\mu_4$. Those estimates are $\bar{y}_{1..},\bar{y}_{2..},\bar{y}_{3..}$, and $\bar{y}_{4..}$. Consider the values $\bar{y}_{ij.}$ and say why "8" makes sense and how you think the standard error(s) could be computed from these sample means using simple computations.

$\bar{y}_{i..} = \mu_i + \bar{\phi}_{i.} + \bar{\epsilon}_{i..}$ so $\text{Var}\,\bar{y}_{i..} = \frac{1}{3}\sigma_\phi^2+\frac{1}{6}\sigma^2 = \frac{1}{3}(\sigma_\phi^2+\frac{1}{2}\sigma^2)$

Now $\bar{y}_{ij.} = \mu_i+\phi_{ij}+\bar{\epsilon}_{ij.}$ so that $\bar{y}_{i1.}, \bar{y}_{i2.}$ and $\bar{y}_{i3.}$ are iid $N(\mu_i, \sigma_\phi^2+\frac{1}{2}\sigma^2)$. So the sample variance of $\bar{y}_{i1.}, \bar{y}_{i2.}$ and $\bar{y}_{i3.}$ provides a 2-degree of freedom estimate of $\sigma_\phi^2+\frac{1}{2}\sigma^2$. Pooling 4 of these provides an 8-degree of freedom estimate of $\sigma_\phi^2+\frac{1}{2}\sigma^2$, which when divided by 3 provides an estimate of $\text{Var}\,\bar{y}_{i..}$.