STAT 8260 Exam 2 – Thursday, April 10, 2008
SHOW ALL WORK

Name: Answer Key

1. Consider the cell means version of the one-way anova model with three groups, two replicates/group:

$$y_{ij} = \mu_i + e_{ij} \quad i = 1, 2, 3; j = 1, 2;$$

where $\{e_{ij}\} \overset{iid}{\sim} N(0, \sigma^2)$ and $\mathbf{y}^T = (y_{11}, y_{12}, y_{21}, y_{22}, y_{31}, y_{32}) = (1, 3, 2, 6, 1, -1)$. Consider the hypothesis $H_0 : \mu_2 = \frac{\mu_1+\mu_3}{2}$.

a. **(12 pts)** Express H_0 the form of the general linear hypothesis. In addition, give two different ways to express the hypothesis $\widetilde{H}_0 : \mu_1 = \mu_2 = \mu_3$ in the form of the general linear hypothesis.

$H_0: C\beta = \underline{t}$ where $C = \left(-\frac{1}{2}, 1, -\frac{1}{2}\right)$, $\beta = \begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\end{pmatrix}$, $\underline{t} = \underline{0}$

$\widetilde{H}_0$ can be written as $C\beta = \underline{t}$ where β, $\underline{t}$ as above and either

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

or ~~$C = (1 \ 0 \ -1$~~

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

(lots of other possibilities)

b. **(9 pts)** Show that the model space under H_0 is $\mathcal{L}(\mathbf{j}_6, \mathbf{x})$ where $\mathbf{x} = (-1, -1, 0, 0, 1, 1)^T$.

Under H_0: $\mu_2 = \frac{\mu_1+\mu_3}{2}$ the mean of $\underline{y}$ is

$$\underline{X}\beta = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} \quad \text{where } \mu_2 = \frac{\mu_1+\mu_3}{2}$$

or

$$\begin{array}{c} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \\ \underline{x}_1 \ \ \underline{x}_2 \ \ \underline{x}_3 \end{array} \begin{pmatrix} \mu_1 \\ (\mu_1+\mu_3)/2 \\ \mu_3 \end{pmatrix} = \mu_1\left(\underline{x}_1 + \tfrac{1}{2}\underline{x}_2\right) + \mu_3\left(\underline{x}_3 + \tfrac{1}{2}\underline{x}_2\right)$$

$$= \mu_1 \underbrace{\begin{pmatrix} 1 \\ 1 \\ 1/2 \\ 1/2 \\ 0 \\ 0 \end{pmatrix}}_{\equiv \underline{v}_1} + \mu_2 \underbrace{\begin{pmatrix} 0 \\ 0 \\ 1/2 \\ 1/2 \\ 1 \\ 1 \end{pmatrix}}_{\equiv \underline{v}_2}$$

$$= \mathcal{L}(\underline{v}_1, \underline{v}_2), \text{ but notice } \underline{v}_1 = \underbrace{\underline{j}_6 - \frac{\underline{x}+\underline{j}_6}{2}}, \ \underline{v}_2 = \underbrace{\frac{\underline{x}+\underline{j}_6}{2}}$$

linear comb. of $\underline{x}, \underline{j}_6$

$$= \mathcal{L}\left(\underline{j}_6 - \frac{\underline{x}+\underline{j}_6}{2}, \ \frac{\underline{x}+\underline{j}_6}{2}\right)$$

$$= \mathcal{L}(\underline{j}_6, \underline{x})$$

c. **(15 pts)** Conduct the test of H_0 as a full and reduced model F test. You need only compute the F test statistic and give its distribution under H_0. You don't need to compute the p-value, critical value or give the conclusion (reject/fail to reject).

Reduced model has model space $\mathcal{L}(\underline{j}_6, \underline{x}) = C(\underline{X})$

where $\underline{X} = \begin{pmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$ so reduced model can be written as

$\underline{y} = \underline{X}\underline{\beta} + \underline{e}$ where $\underline{\beta} = \begin{pmatrix} \mu_1 \\ \mu_3 \end{pmatrix}$

this is just a simple linear regression model, with intercept μ_1 & slope μ_3. In reduced model $\hat{y} = p(y \mid \mathcal{L}(j_6, \underline{x}))$

$$= \left(\frac{1}{\sqrt{6}}\underline{j}_6, \frac{1}{2}\underline{x}\right)\begin{pmatrix} \frac{1}{\sqrt{6}}\underline{j}_6^T \\ \frac{1}{2}\underline{x}^T \end{pmatrix}\underline{y} = \left(\frac{1}{\sqrt{6}}\underline{j}_6, \frac{1}{2}\underline{x}\right)\begin{pmatrix} \frac{1}{\sqrt{6}}\underline{j}_6^T y \\ \frac{1}{2}\underline{x}^T y \end{pmatrix}$$

$$= \left(\frac{1}{\sqrt{6}}\underline{j}_6, \frac{1}{2}\underline{x}\right)\begin{pmatrix} \frac{1}{\sqrt{6}}(12) \\ -2 \end{pmatrix} = \begin{pmatrix} \frac{1}{6}(12) + (-\frac{1}{2})(-2) \\ \frac{1}{6}(12) + (-\frac{1}{2})(-2) \\ \frac{1}{6}(12) + 0(-2) \\ \frac{1}{6}(12) + 0(-2) \\ \frac{1}{6}(12) + 0(-2) \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 2 \\ 2 \\ 1 \\ 1 \end{pmatrix} \Rightarrow \underline{y} - \hat{\underline{y}} = \begin{pmatrix} -2 \\ 0 \\ 0 \\ 4 \\ 0 \\ -2 \end{pmatrix}$$

$\Rightarrow SSE(RM) = \|y - \hat{y}\|^2 = 4 + 0 + 0 + 16 + 0 + 4 = 24$

$\hat{y}$ in FM is $\left(\frac{y_1+y_2}{2}, \frac{y_1+y_2}{2}, \frac{y_3+y_4}{2}, \frac{y_3+y_4}{2}, \frac{y_5+y_6}{2}, \frac{y_5+y_6}{2}\right)^T$

$= (2, 2, 4, 4, 0, 0)^T \Rightarrow y - \hat{y} = (-1, 1, -2, 2, 1, -1)^T$

$\Rightarrow SSE(FM) = (-1)^2 + 1^2 + (-2)^2 + 2^2 + 1^2 + (-1)^2 = 12$

$$\Rightarrow F = \frac{(SSE(RM) - SSE(FM))/(dfE(RM) - dfE(FM))}{SSE(FM)/dfE(FM)}$$

$$= \frac{(24-12)/1}{12/3} = 3 \qquad F \sim F(1,3) \text{ under } H_0$$

d. **(9 pts)** Suppose now that we obtain an additional observation not in the original data set from group 1 but its observed value ($y_{13} = 7$) seems unusually large and thus inconsistent with the original data. To investigate this, compute a 95% prediction interval for a new observation from group 1 based only on the original 6 observations. For this prediction interval, suppose that σ^2 is known to be 1.

Model is $\underline{y} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} + \underline{e}$

interested in predicting y when $\underline{x}_0^T = (1, 0, 0)$,

$$\hat{y}_0 = \underline{x}_0^T \hat{\beta} = (1,0,0)\begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \end{pmatrix} = \hat{\mu}_1 = \frac{y_1 + y_2}{2} = \frac{1+3}{2} = 2$$

If σ^2 were unknown, we'd use $\hat{y}_0 \pm t_{1-\alpha/2}(n - ncols(X))\sqrt{s^2(1 + x_0^T(X^TX)^{-1}x_0)}$

but since σ^2 is known, we can use a z-interval:

$$\hat{y}_0 \pm \underbrace{z_{1-\alpha/2}}_{=1.96 \text{ for } \alpha=.05} \sqrt{\sigma^2(1 + x_0^T(X^TX)^{-1}x_0)}, \quad (X^TX)^{-1} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}^{-1} = \frac{1}{2} I$$

$$\Rightarrow \hat{y}_0 \pm z_{1-\alpha/2}\sqrt{\sigma^2(1 + x_0^T(X^TX)^{-1}x_0)} = 2 \pm 1.96\sqrt{1(1 + \tfrac{1}{2}\underbrace{x_0^Tx_0}_{=1})}$$

$$= 2 \pm 1.96\sqrt{1.5} = (-.40, 4.40)$$

e. **(8 pts)** Compute the sample multiple correlation coefficient between $\mathbf{y}$ and $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$, where $\mathbf{x}_i$ is the indicator for the ith group.

computed in part c

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{12}{28} = .5714$$

$$\left(SST = \sum_i (y_i - \bar{y})^2 = \sum (y_i - 2)^2 = 1^2 + 1^2 + 0^2 + 4^2 + 1^2 + 3^2 = 28 \right)$$

$\Rightarrow$ sample mult. corr. coef $= \sqrt{R^2} = \sqrt{.5714} = .7559$

2. Let $y_i = \beta x_i + e_i$, $i = 1, 2, 3$, where $\mathbf{x}^T = (1, 2, 4)$, e_1, e_2, e_3 are mutually independent and $\mathrm{E}(e_i) = 0$, $\mathrm{var}(e_i) = \sigma^2 x_i$, for all i.

a. **(10 pts)** Under the assumptions given above, provide a simple, non-vector/matrix formula for the BLUE of β in terms of $\mathbf{y} = (y_1, y_2, y_3)^T$.

$\underline{y} = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} \beta + \underline{e}$ where $E(\underline{e}) = \underline{0}$, $Var(\underline{e}) = \sigma^2 \underline{V}$ where $V = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & & 4 \end{pmatrix}$

$V^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/4 \end{pmatrix}$ $\quad \hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y$

$X^T V^{-1} \underline{y} = (1 \;\; 2 \;\; 4) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/4 \end{pmatrix} \underline{y} = (1, 1, 1)\underline{y} = y_1 + y_2 + y_3 = y_\bullet$

$(X^T V^{-1} X)^{-1} = \left[(1, 1, 1) \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} \right]^{-1} = 7^{-1} = 1/7 \quad \Rightarrow \hat{\beta} = y_\bullet / 7$

b. **(10 pts)** Suppose now that the assumptions on $\mathbf{e} = (e_1, e_2, e_3)^T$ above were wrong and instead, the truth is that $\mathrm{E}(\mathbf{e}) = \mathbf{0}$ and $\mathrm{var}(\mathbf{e}) = \sigma^2\mathbf{I}$. Under these circumstances, compute the bias and variance of your estimator $\hat{\beta}$ from part (a).

$$var(\hat{\beta}) = var\left((X^TV^{-1}X)^{-1}X^TV^{-1}y\right) =$$

$$= (X^TV^{-1}X)^{-1}X^TV^{-1}var(y)\,V^{-1}X\,(X^TV^{-1}X)^{-1}$$

$$= \frac{1}{7}\,\underline{j}_3^T\,\sigma^2 I\,\frac{1}{7}\,\underline{j}_3 = \frac{3\sigma^2}{49}$$

$$E(\hat{\beta}) = (X^TV^{-1}X)^{-1}X^TV^{-1}\underbrace{E(y)}_{X\beta} = \cancel{(X^TV^{-1}X)^{-1}X^TV^{-1}X}\,\beta = \beta$$

(unbiased)

3. **(12 pts)** Consider the classical linear model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}, \quad \mathbf{e} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$$

where $\mathbf{X}$ is $n \times (k+1)$ with $\text{rank}(\mathbf{X}) = k+1 < n$. Suppose this model is fit to data $\mathbf{y}$ yielding the ordinary least squares regression parameter estimate $\hat{\beta}$ and MSE, s^2.

What is the distribution of $(k+1)s^2/\sigma^2$ in this model? Use this distributional result to derive formulas for the lower and upper endpoints (L, U) of a $100(1-\alpha)\%$ confidence interval for σ, the error standard deviation. **Show your work!**

$$\frac{(k+1)s^2}{\sigma^2} = \frac{(n-k-1)s^2}{\sigma^2}\,\frac{(k+1)}{(n-k-1)} \sim \chi^2(n-k-1) \underset{\text{times}}{\times} \left(\frac{k+1}{n-k-1}\right)$$

$$\Rightarrow \Pr\left(\cancel{\frac{k+1}{n-k-1}}\,\chi^2_\alpha(n-k-1) \le \cancel{\frac{k+1}{n-k-1}}\,\frac{(n-k-1)s^2}{\sigma^2} \le \cancel{\frac{k+1}{n-k-1}}\,\chi^2_{1-\frac{\alpha}{2}}(n-k-1)\right) = 1-\alpha$$

$$\Rightarrow \Pr\left(\chi^2_\alpha(n-k-1) \le \frac{(n-k-1)s^2}{\sigma^2} \le \chi^2_{1-\alpha/2}(n-k-1)\right) = 1-\alpha$$

$$\Rightarrow \Pr\left(\frac{\chi^2_\alpha(n-k-1)}{(n-k-1)s^2} \le \sigma^{-2} \le \frac{\chi^2_{1-\alpha/2}(n-k-1)}{(n-k-1)\,s^2}\right) = 1-\alpha$$

$$\Rightarrow \Pr\left(\frac{(n-k-1)s^2}{\chi^2_\alpha(n-k-1)} \ge \sigma^2 \ge \frac{(n-k-1)\,s^2}{\chi^2_{1-\alpha/2}(n-k-1)}\right) = 1-\alpha$$

$$\Rightarrow \Pr\left(\underbrace{\sqrt{\frac{(n-k-1)s^2}{\chi^2_\alpha(n-k-1)}}}_{=U} \ge \sigma \ge \underbrace{\sqrt{\frac{(n-k-1)\,s^2}{\chi^2_{1-\alpha/2}(n-k-1)}}}_{=L}\right) = 1-\alpha$$

4. Consider the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad E(\mathbf{e}) = \mathbf{0}, \quad \text{var}(\mathbf{e}) = \sigma^2\mathbf{V}$$

where $\mathbf{X}$ is $n \times (k+1)$ with $\text{rank}(\mathbf{X}) = k+1 < n$ and $\mathbf{V}$ is assumed to be a positive definite matrix. The following questions are True/False (**3 pts each**).

I meant (d) to be false by writing that c'*betahat was BLUE, but only if c is in the row space of X. No such requirement on c is necessary in the full rank linear model. However, my phrasing was poor here. It reads more like, "if c is in the row space of X then BLUE-ness follows". This statement is TRUE (it wouldn't be true if we replaced "if" by "only if" which was my intention). Therefore, I accepted both TRUE and FALSE as correct answers here.

a. (True/False) If $\mathbf{V} = \mathbf{I}$, the estimator $\mathbf{c}^T\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$, has minimum variance among all unbiased estimators of $\mathbf{c}^T\boldsymbol{\beta}$.

(need normality for this result)

b. (True/False) If $\mathbf{V} = \mathbf{I}$, the estimator $s^2 = \|\mathbf{P}_{C(\mathbf{X})^\perp}\mathbf{y}\|^2/(n-k-1)$ has minimum variance among all quadratic unbiased estimators of σ^2.

c. (True/False) The estimator $\mathbf{c}^T\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{y}$ and $\hat{\mathbf{V}}$ is an estimator of $\mathbf{V}$, is the BLUE (best linear unbiased estimator) of $\mathbf{c}^T\boldsymbol{\beta}$.

Would be true if V were known and used in formula for β instead of $\hat{V}$

d. (True/False) If $\mathbf{V} = \mathbf{I}$ and $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y}$, the estimator $\mathbf{c}^T\hat{\boldsymbol{\beta}}$ has minimum variance among all linear unbiased estimators of $\mathbf{c}^T\boldsymbol{\beta}$, provided that $\mathbf{c} \in C(\mathbf{X}^T)$.

there are no such conditions on the BLUE-ness of $c^T\hat{\beta}$ in the full rank model

e. (True/False) If $\mathbf{V}$ is known and $\mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{V})$, the estimator $\mathbf{c}^T\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$, has minimum mean squared error among all unbiased estimators of $\mathbf{c}^T\boldsymbol{\beta}$.

In class on 4/15/08 I said that there was a mistake here and that the answer was really FALSE. However, that's not correct! The answer is TRUE! I just got a bit confused about this in the rush to finish class. Under normal, spherical errors the OLS estimator is not only best among linear unbiased estimators, it is also best among all unbiased estimators (p.115 of our class notes). This result extends to the GLS estimator too because the GLS estimator is the OLS estimator in a transformed model and therefore inherits all the properties of the OLS estimator.

$MSE = \text{variance} + \text{bias}^2$

bias $= 0$ here $\Rightarrow$ minimum MSE $=$ minimum variance

So this statement say $c^T\hat{\beta}$ is ~~BLUE~~, which is true

if errors are normally distributed

the minimum variance unbiased estimator