Stat 511 Exam1

February 24, 2009
Prof. Vardeman

I have neither given nor received unauthorized assistance on this exam.

____
Name

____
Name Printed

1. Consider a segmented simple linear regression problem in one variable, x. In particular, suppose that $n=6$ values of a response y are related to values $x=0,1,2,3,4,5$ by a Gauss-Markov normal linear model $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$ for

$$\mathbf{Y}=\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix},\ \mathbf{X}=\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 1 \\ 1 & 4 & 2 \\ 1 & 5 & 3 \end{pmatrix},\ \boldsymbol{\beta}=\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix},\ \text{and}\ \boldsymbol{\varepsilon}=\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}$$

Values of x are in the second column of the model matrix. This model allows the linear form $y \approx \beta_0+\beta_1 x$ for $x \le 2$ and the linear form $y \approx \beta_0 + 2\beta_1 + (\beta_1+\beta_2)(x-2)$ for $x \ge 2$. *Notice* that there is continuity of these forms at $x=2$.

10 pts

a) This a full rank model. **Argue** carefully that this is the case.

Here $(\mathbf{X}'\mathbf{X})^{-1}=\begin{pmatrix} .825 & -.474 & .526 \\ -.474 & .421 & -.579 \\ .526 & -.579 & .921 \end{pmatrix}$ and for $\mathbf{Y}'=(0,2,4,3,1,0)$, $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}=\begin{pmatrix} -.018 \\ 2.053 \\ -3.447 \end{pmatrix}$ and $SSE=.202$.

10 pts

b) **Is there definitive evidence** that a simpler model $y \approx \beta_0+\beta_1 x\ \ \forall x$ is inadequate here? **Explain**.

10 pts

c) Tomorrow a total of 3 new observations are to be drawn from this model at, respectively, $x=1,2,$ and 3. Call these $y_1^*, y_2^*,$ and y_3^*. The quantity $\left(y_3^*-y_2^*\right)-\left(y_2^*-y_1^*\right)=y_3^*-2y_2^*+y_1^*$ is an empirical measure of change in slope of mean y as a function of x at $x=2$ based on these new observations. **Provide** 95% two-sided **prediction limits** for this quantity. (**Plug in completely**, but you need not do arithmetic.)

10 pts

d) **Find the value and degrees of freedom** for a t statistic for testing $\text{H}_0:\mu_{y|x=1}=\mu_{y|x=5}$ (the hypothesis that the mean responses are the same for $x=1$ and $x=5$).

$T=$ ____________ $df=$ __________________

10 pts

e) **Write out** (plug in completely so that your implied answer is numerical, but you need not do the arithmetic) **a test statistic** that you could use to test the hypothesis that $\mathrm{H}_0 : \mu_{y|x=1} = \mu_{y|x=5} = 1$. **Say** exactly what null distribution you would use.

10 pts

f) It is possible to compute $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ both for the full model specified at the beginning of this problem and for a model with $\mathbf{X}$ matrix consisting of only the first two columns of the original one. The diagonal entries of these two matrices are in the table below.

i	1	2	3	4	5	6
x_i	0	1	2	3	4	5
diagonal entry of $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ for the original $\mathbf{X}$	.825	.298	.614	.272	.298	.693
diagonal entry of $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ for the reduced $\mathbf{X}$	.524	.295	.181	.181	.295	.524

Compare the two patterns above and **say** why (in the context provided at the beginning of this problem) they "make sense."

10 pts

2. Suppose that $\mathbf{Y}$ is $\text{MVN}_n\left(\boldsymbol{\mu},\sigma^2\mathbf{I}\right)$ and that $\mathbf{A},\mathbf{B},$ and $\mathbf{C}$ are symmetric $n\times n$ matrices with $\mathbf{AB}=\mathbf{0}$, $\mathbf{AC}=\mathbf{0}$, and $\mathbf{BC}=\mathbf{0}$. **Argue carefully** that the three random variables $\mathbf{Y}'\mathbf{AY}$, $\mathbf{Y}'\mathbf{BY}$, and $\mathbf{Y}'\mathbf{CY}$ are *jointly* independent.

10 pts

3. Suppose that y_{11} and y_{12} are independent $\text{N}\left(\mu_1,\eta\right)$ variables independent of y_{21} and y_{22} that are independent $\text{N}\left(\mu_2,4\eta\right)$ variables. (The η and 4η are variances.) **What is the BLUE** of $\mu_1-\mu_2$? **Explain** carefully.

5pts

4. a) For any non-zero $\mathbf{w} \in \Re^n$ the set of multiples of $\mathbf{w}$, namely $\{c\mathbf{w} \mid c \in \Re\}$, is a 1-dimensional subspace of $\Re^n$. We might call this subspace $C(\mathbf{w})$. Consider the operation of perpendicular projection onto $C(\mathbf{w})$, accomplished using the $n \times n$ projection matrix $\mathbf{P}_{\mathbf{w}}$. **Argue carefully** that for any $\mathbf{v} \in \Re^n$,

$$\mathbf{P}_{\mathbf{w}}\mathbf{v} = \left(\frac{\mathbf{v}'\mathbf{w}}{\mathbf{w}'\mathbf{w}}\right)\mathbf{w}$$

(Note that $\mathbf{P}_{\mathbf{w}}\mathbf{v} = c\mathbf{w}$ for some $c \in \Re$, and consider $c\mathbf{w}'\mathbf{w}$.)

5pts

b) In the regression context from lecture, let $\mathbf{X} = (\mathbf{1} \mid \mathbf{x}_1 \mid \mathbf{x}_2 \mid \cdots \mid \mathbf{x}_{r-1} \mid \mathbf{x}_r)$ and $\mathbf{X}_{r-1} = (\mathbf{1} \mid \mathbf{x}_1 \mid \mathbf{x}_2 \mid \cdots \mid \mathbf{x}_{r-1})$. Further, let

$$\mathbf{z}_r = \mathbf{x}_r - \mathbf{P}_{\mathbf{X}_{r-1}}\mathbf{x}_r = \left(\mathbf{I} - \mathbf{P}_{\mathbf{X}_{r-1}}\right)\mathbf{x}_r$$

Argue carefully that for any $\mathbf{v} \in C(\mathbf{X}_{r-1})$, $\mathbf{v} \perp \mathbf{z}_r$.

5pts

c) As a matter of fact, $\mathbf{P_X} - \mathbf{P}_{\mathbf{X}_{r-1}} = \mathbf{P}_{\mathbf{z}_r}$. **Argue carefully** here that $\mathbf{P_X} - \mathbf{P}_{\mathbf{X}_{r-1}}$ is symmetric and idempotent, and that $\left(\mathbf{P_X} - \mathbf{P}_{\mathbf{X}_{r-1}}\right)\mathbf{v} = \mathbf{v}$ for any $\mathbf{v} \in C(\mathbf{z}_r)$.

5pts

d) Using the facts in a)-c) **argue carefully** that

$$\hat{\mathbf{Y}} = \mathbf{P}_{\mathbf{X}_{r-1}}\mathbf{Y} + \left(\frac{\mathbf{e}'_{r-1}\mathbf{z}_r}{\mathbf{z}'_r\mathbf{z}_r}\right)\mathbf{z}_r$$

for $\mathbf{e}_{r-1} = \left(\mathbf{I} - \mathbf{P}_{\mathbf{X}_{r-1}}\right)\mathbf{Y}$. Then **say why** it is clear that the multiplier of $\mathbf{z}_r$ here is b_r^{OLS}, the ordinary least squares estimate of the regression coefficient β_r in the full original regression. **What interpretation** does this development provide for b_r^{OLS}?