Stat 511 Exam1

February 24, 2009
Prof. Vardeman

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

KEY

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 \( Y = X\beta + \varepsilon \) for

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

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 \( x \leq 2 \) and the linear form \( y \approx \beta_0 + 2\beta_1 + (\beta_1 + \beta_2)(x - 2) \) for \( x \geq 2 \). 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.

**Rank is the number of linearly independent columns or rows of X. Consider the 1st, 2nd and 4th rows of the X matrix. Placing these into a 3x3 matrix, one has**

\[
\begin{pmatrix}
1 & 0 & 0 \\
1 & 1 & 0 \\
1 & 3 & 1 \\
\end{pmatrix}
\]

**and this matrix has determinant 1 \( \neq 0 \). That is, it is non-singular i.e. has rank = 3. So X has maximum rank.**

**Here \( X'X \)^{-1} = \begin{pmatrix}
.825 & -1.474 & .526 \\
-1.474 & .421 & -.579 \\
.526 & -.579 & .921
\end{pmatrix} \text{ and for } Y' = (0, 2, 4, 3, 1, 0), \quad (X'X)^{-1} X Y = \begin{pmatrix}
-.018 \\
2.053 \\
-3.447
\end{pmatrix} \text{ and } \quad 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.

Consider testing \( H_0: \beta_2 = 0 \) in the full model. A t-statistic for doing this is

\[
T = \frac{b_2 - 0}{\sqrt{MSE \sqrt{d_3}}} = \frac{-3.447 - 0}{\frac{.202}{6-3} \sqrt{.921}} = -13.8
\]

This is a huge value of a \( t \)-r.v. and there is thus definitive evidence that \( \beta_2 \neq 0 \) i.e. The simple linear regression model is inadequate here.
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 \( (y_3^* - y_2^* - (y_2^* - y_1^*) = y_3^* - 2y_2^* + y_1^* \) 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.)

Note that the mean of \( y_3^* - 2y_2^* + y_1^* \) is \( \beta_0 + 3\beta_1 + 3\beta_2 \)
\[ -2(\beta_0 + 2\beta_1 + 0\beta_2) + \beta_0 + \beta_1 + 0\beta_2 = \beta_2 \]
So, since
\[ \text{Var}(y_3^* - 2y_2^* + y_1^*) = 2\sigma^2 + 4\sigma^2 + \sigma^2 = 6\sigma^2 \]
we may use limits
\[ \hat{b}_{2}\text{OLS} \pm t\sqrt{MSE} \sqrt{6 + d_3} \]
\[ -3.447 \pm 3.182\sqrt{\frac{202}{3}} \sqrt{6 + 5.921} \]
\[ 2.17 \]

\[ d_3 \]

\[ \text{df} = 6 - 3 = 3 \]

---

d) **Find the value and degrees of freedom** for a \( t \) statistic for testing \( H_0 : \mu_{y|y=1} = \mu_{y|y=3} \) (the 
hypothesis that the mean responses are the same for \( x = 1 \) and \( x = 3 \)).

\[ M_{y|y=1} - M_{y|y=3} = \beta_0 + \beta_1 - (\beta_0 + 5\beta_1 + 3\beta_2) = -4\beta_1 - 3\beta_2 \]
So \[ \hat{\beta}_{OLS} = -4(2.053) - 3(-3.447) = 2.129 \] This
has standard error \[ \sqrt{MSE} \sqrt{(-4\sigma^2 - 3\sigma^2)} \]
\[ \frac{1}{\sqrt{202}} \sqrt{1.129} = .2757 \]
So \[ \frac{\hat{\beta}_{OLS}}{\text{std.err.}} = \frac{2.129}{.2757} = 7.72 \]

\[ T = 7.72 \]

\[ df = 6 - 3 = 3 \]
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 \( H_0 : \mu_{y|x=1} = \mu_{y|x=5} = 1 \). Say exactly what null distribution you would use.

We may use an \( F \) test of the testable hypothesis

\[
H_0 : \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 1 \end{pmatrix}
\]

This is based on

\[
F = \left( \frac{.202}{3} \right)
\]

\[
= \left( \frac{-0.018+5(2.053)+3(-5.997)-1}{153} \right) \left( \begin{pmatrix} .825 & .474 & .526 \\ -9.74 & .921 & -.579 \\ .526 & -.579 & .921 \end{pmatrix} \right)^{\frac{1}{2}} \left( \begin{pmatrix} 1 \\ 153 \\ 1 \end{pmatrix} \right)
\]

The reference df will be \( F_{2,3} \).

---

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

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_i )</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>diagonal entry of ( X(X'X)^{-1}X' ) for the original ( X )</td>
<td>.825</td>
<td>.298</td>
<td>.614</td>
<td>.272</td>
<td>.298</td>
<td>.693</td>
</tr>
<tr>
<td>diagonal entry of ( X(X'X)^{-1}X' ) for the reduced ( X )</td>
<td>.524</td>
<td>.295</td>
<td>.181</td>
<td>.181</td>
<td>.295</td>
<td>.524</td>
</tr>
</tbody>
</table>

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

The **reduced model pattern** is high on the "ends" and low in the middle. The **full model pattern** is high on the ends, but also has a high value at \( x = 2 \) in the "middle," exactly where the \( 2 \) linear forms are assumed to "join up"/be continuous. These are measures of "influence" of the data points on the fitting. In the **reduced case** it is plausible that the "end" data points are most influential in determining fit. In the **full model case** it is also quite plausible that the interior point where linear forms must agree (and which functions as an "end point" for both segments) will also be important.
2. Suppose that \( Y \) is \( \text{MVN}_n\left(\mu, \sigma^2 I\right) \) and that \( A, B, \) and \( C \) are symmetric \( n \times n \) matrices with \( AB = 0, AC = 0, \) and \( BC = 0. \) Argue carefully that the three random variables \( Y'AY, Y'BY, \) and \( Y'CY \) are jointly independent. Consider \( \begin{pmatrix} A & B \\ B & C \end{pmatrix} \) \( Y \). This is \( \text{MVN}_{3n} \) with covariance matrix

\[
\begin{pmatrix} A \\ B \\ C \end{pmatrix} \begin{pmatrix} A & B & C \\ B & B & B \\ C & C & C \end{pmatrix} = \sigma^2 \begin{pmatrix} A & B & C \\ B & B & B \\ C & C & C \end{pmatrix} = \sigma^2 \begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{pmatrix}
\]

and since for MVN variables \( 0 \) covariance implies independence and since for MVN variables \( 0 \) covariance implies independence (and so are functions of these). But \( Y'AY = Y'A A^{-1} AY = (AY)'A^{-1}AY \) is a function of \( AY. \) Similarly, \( Y'BY \) is a function of \( BY \) and \( Y'CY \) is a function of \( CY. \)

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

\[
\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \epsilon \quad \text{where } \epsilon \sim \text{MVN}_4(0, \eta \text{diag}(1,1,4,4))
\]

Let \( V^{-\frac{1}{2}} = \text{diag}(1,1,\frac{1}{2},\frac{1}{2}) \) and \( U = V^{-\frac{1}{2}} Y = \begin{pmatrix} y_{11} \\ y_{12} \\ \frac{1}{2} y_{21} \\ \frac{1}{2} y_{22} \end{pmatrix} \) follows a Gauss-Markov linear model with model matrix \( W = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \).

Here, OLS is BLUE, so the BLUE of \( \mu_1 - \mu_2 = (1-1) \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \) is

\[
(1-1) (W'W)^{-1} W'U = (1-1) \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{pmatrix} = (1-1) \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} y_{11} + y_{12} \\ \frac{1}{4} y_{21} + \frac{1}{4} y_{22} \end{pmatrix}
\]

\[
= \frac{1}{2} (y_{11} + y_{12}) - \frac{1}{2} (y_{21} + y_{22})
\]

Notice that this is true no matter what \( y \).
4. a) For any non-zero $w \in \mathbb{R}^n$ the set of multiples of $w$, namely $\{cw | c \in \mathbb{R}\}$, is a 1-dimensional subspace of $\mathbb{R}^n$. We might call this subspace $C(w)$. Consider the operation of perpendicular projection onto $C(w)$, accomplished using the $n \times n$ projection matrix $P_w$. Argue carefully that for any $v \in \mathbb{R}^n$,

$$P_w v = \left( \frac{v'w}{w'w} \right) w$$

(Note that $P_w v = cw$ for some $c \in \mathbb{R}$, and consider $cw'w$.)

$$C\left( \frac{w'w}{w'w} \right) = \left( P_w v \right)'w = \frac{v'w}{w'w} \cdot w = \frac{v'w}{w'w}$$

so, taking the 1st and last of these, we have

$$\frac{v'w}{w'w} = v'w \quad \text{and} \quad c = \frac{v'w}{w'w} \quad \text{i.e.} \quad P_w v = \left( \frac{v'w}{w'w} \right) v$$

b) In the regression context from lecture, let $X = (1 | x_1 | x_2 | \cdots | x_{r-1} | x_r)$ and $X_{r-1} = (1 | x_1 | x_2 | \cdots | x_{r-1})$. Further, let

$$z_r = x_r - P_{X_{r-1}} x_r = (I - P_{X_{r-1}}) x_r$$

Argue carefully that for any $v \in C(X_{r-1})$, $v \perp z_r$.

For $v \in C(X_{r-1})$

$$v' z_r = v' (z_r - P_{X_{r-1}} z_r) = v' z_r - v' P_{X_{r-1}} z_r$$

$$= v' z_r - \left( \frac{v' z_r}{z_r' z_r} \right) z_r$$

$$= v' z_r - \left( \frac{v' z_r}{z_r' z_r} \right) z_r$$

$$= 0$$
c) As a matter of fact, \( P_X - P_{X_{r-1}} = P_z \). Argue carefully here that \( P_X - P_{X_{r-1}} \) is symmetric and idempotent, and that \( (P_X - P_{X_{r-1}})v = v \) for any \( v \in C(z_r) \).

\[
(P_X - P_{X_{r-1}})(P_X - P_{X_{r-1}}) = P_X P_X - P_X P_{X_{r-1}} - P_{X_{r-1}} P_X + P_{X_{r-1}} P_{X_{r-1}}
\]

\[
= P_X - P_{X_{r-1}} - P_{X_{r-1}} + P_{X_{r-1}}
\]

\[
= P_X - P_{X_{r-1}}
\]

\[
(P_X - P_{X_{r-1}})c z_r = c (P_X - P_{X_{r-1}})z_r = c (P_X - P_{X_{r-1}})(z_r - P_{X_{r-1}} z_r)
\]

\[
= c (P_X z_r - P_{X_{r-1}} z_r - P_{X_{r-1}} z_r + P_{X_{r-1}} P_{X_{r-1}} z_r)
\]

\[
= c (z_r - P_{X_{r-1}} z_r - P_{X_{r-1}} z_r + P_{X_{r-1}} z_r)
\]

\[
= c (z_r - P_{X_{r-1}} z_r) = c z_r
\]

5pts
d) Using the facts in a)-c) argue carefully that

\[
\hat{Y} = P_{X_{r-1}}Y + \left( \frac{e_{r-1}'}{z_r' z_r} \right) z_r
\]

for \( e_{r-1} = (I - P_{X_{r-1}})Y \). Then say why it is clear that the multiplier of \( z_r \) here is \( b_r^{OLS} \), the ordinary least squares estimate of the regression coefficient \( \beta_r \) in the full original regression. What interpretation does this development provide for \( b_r^{OLS} \)?

\[
\hat{Y} = P_X Y = (P_{X_{r-1}} + (I - P_{X_{r-1}}))Y = P_{X_{r-1}} Y + P_{I-r} Y
\]

But since \( z_r \perp C(X_{r-1}) \), \( P_{z_r} Y = P_{z_r} Y + P_{z_r} P_{X_{r-1}} Y \)

\[
= P_{z_r} (Y - P_{X_{r-1}} Y)
\]

\[
= P_{z_r} (I - P_{X_{r-1}}) Y
\]

\[
= P_{z_r} e_{r-1}
\]

Now then, since \( P_{X_{r-1}} Y \) is a linear combination of only the first \( r-1 \) columns of \( X \), the coefficient of \( z_r = \frac{e_{r-1}'}{z_r' z_r} \) is the multiplier of \( z_r \) in making up \( \hat{Y} \) as a l.c. of the columns of \( X \), i.e. is \( b_r^{OLS} \). This regression coefficient is in the part of \( Y \) not in \( C(X_{r-1}) \) regressed on the part of \( z_r \) also not in \( C(X_{r-1}) \) ... it measures what remaining part of \( Y \) is "explained" by the part of \( z_r \) not already accounted for by \( X_{r-1} \).