Ph.D. Prelim Exam, Spring 2002 Linear Models: Page 1 of 1

Suppose that $\underline{y}$ is an observable random vector that follows the Gauss-Markov model

$$\underline{y} = X_1\underline{\beta}_1 + X_2\underline{\beta}_2 + \underline{\epsilon}, \tag{1}$$

where $\underline{\beta}_1$ and $\underline{\beta}_2$ are $p \times 1$ and $q \times 1$ vectors of unknown parameters, respectively, $(X_1, X_2) = X$ is a given $n \times (p+q)$ matrix of rank $p+q$, and $\underline{\epsilon} \sim N(\underline{0}, \sigma^2 I)$. Here σ^2 is an unknown parameter with $\sigma^2 > 0$. Let $\widehat{\underline{\beta}}_1$ and $\widehat{\underline{\beta}}_2$ be the least squares estimators of $\underline{\beta}_1$ and $\underline{\beta}_2$, respectively; let $\underline{\beta} = (\underline{\beta}_1', \underline{\beta}_2')'$ and $P_{X_2} = X_2(X_2'X_2)^{-1}X_2'$.

(a) Show that $X_1'(I - P_{X_2})X_1$ is a nonsingular matrix.

(b) Show that $\widehat{\underline{\beta}}_1 = [X_1'(I - P_{X_2})X_1]^{-1}X_1'(I - P_{X_2})\underline{y}$.

(c) Find the distribution of $\widehat{\underline{\beta}}_1$.

(d) Show that $\widehat{\underline{\beta}}_1$ and $\widehat{\underline{\beta}}_2$ are independent if and only if $X_1'X_2 = 0$.

(e) Show that $(X_1'X_1)^{-1}X_1'\underline{y}$ is the least squares estimator of $\underline{\beta}_1$ if and only if $X_1'X_2 = 0$.

(f) Show that $\widehat{\underline{\beta}}_1$ is the least squares estimator of $\underline{\beta}_1$ for the Gauss-Markov model $\underline{y} = X_1\underline{\beta}_1 + \underline{\epsilon}$ if and only if $X_1'X_2\widehat{\underline{\beta}}_2 = \underline{0}$.

(g) Derive a size-γ F-test of the null hypothesis $H_0 : \underline{\beta}_2 = \underline{0}$ versus the alternative hypothesis $H_a : \underline{\beta}_2 \neq \underline{0}$. Give, under H_a, the noncentrality parameter of the test statistic and the expected values of its numerator and denominator.

(h) Construct a $(1-\gamma)$ confidence set for the ratio $\underline{a}'\underline{\beta}/\underline{b}'\underline{\beta}$ of two estimable functions $\underline{a}'\underline{\beta}$ and $\underline{b}'\underline{\beta}$.

(i) Suppose that $\underline{\beta}_2$ in Model (1) is a random vector with $\mathrm{Cov}(\underline{\beta}_2, \underline{\epsilon}) = 0$ and $\underline{\beta}_2 \sim N(\underline{0}, \sigma^2 D)$, where D is a given positive definite matrix. Suppose further that all the other assumptions made for the model are unchanged. Find the best linear unbiased predictor (BLUP) of $\underline{\beta}$ and the distribution of this predictor.

(j) Give a sequence of positive definite matrices D_k, $k = 1, 2, \ldots$, such that $\widetilde{\underline{\beta}}^{(k)} \to \widehat{\underline{\beta}}$ as $k \to \infty$, where $\widetilde{\underline{\beta}}^{(k)}$ is the BLUP of $\underline{\beta}$ for the model in part (i) with D replaced by D_k.

Ph.D. Prelim Exam Spring 2002

Solutions Linear Models

(a). Since $X=(X_1, X_2)$ is an $n \times (p+q)$ matrix of rank $p+q$, $X_1'X_1$, $X_2'X_2$ and $X'X = \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}$ are nonsingular matrices. Thus

$$\begin{pmatrix} X_1'(I-P_{X_2})X_1 & 0 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} = \begin{pmatrix} I & -X_1'X_2(X_2'X_2)^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}$$

is a nonsingular matrix. This implies that $X_1'(I-P_{X_2})X_1$ is nonsingular.

(b). The normal equations $X'X\hat{\beta} = X'\underline{y}$ can be written as $\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} X_1'\underline{y} \\ X_2'\underline{y} \end{pmatrix}$, or

$$X_1'X_1\hat{\beta}_1 + X_1'X_2\hat{\beta}_2 = X_1'\underline{y} \qquad (*)$$
$$X_2'X_1\hat{\beta}_1 + X_2'X_2\hat{\beta}_2 = X_2'\underline{y}. \qquad (**)$$

It follows from (**) that $\hat{\beta}_2 = (X_2'X_2)^{-1}X_2'(\underline{y} - X_1\hat{\beta}_1)$. Then (*) becomes $X_1'X_1\hat{\beta}_1 + X_1'P_{X_2}(\underline{y} - X_1\hat{\beta}_1) = X_1'\underline{y}$. Solving this equation for $\hat{\beta}_1$ yields

$$\hat{\beta}_1 = [X_1'(I-P_{X_2})X_1]^{-1}X_1'(I-P_{X_2})\underline{y}.$$

(c). Note that $\hat{\beta}_1$ is an unbiased estimator of β_1. Thus $E\hat{\beta}_1 = \beta_1$. Furthermore, $\hat{\beta}_1$ is normally distributed and $Var(\hat{\beta}_1) = \sigma^2[X_1'(I-P_{X_2})X_1]^{-1}X_1'(I-P_{X_2})X_1[X_1'(I-P_{X_2})X_1]^{-1}$ $= \sigma^2[X_1'(I-P_{X_2})X_1]^{-1}$. Thus $\hat{\beta}_1 \sim N(\beta_1, \sigma^2[X_1'(I-P_{X_2})X_1]^{-1})$.

(d) Denote the symmetric matrix $\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1}$ by $\begin{pmatrix} A & C \\ C' & B \end{pmatrix}$, where A and B are $p\times p$ and $q\times q$ matrices, respectively. Then $\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \sigma^2\begin{pmatrix} A & C \\ C' & B \end{pmatrix}\right)$.

Therefore, $\hat{\beta}_1$ and $\hat{\beta}_2$ are independent $\Leftrightarrow C=0$

$\Leftrightarrow \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \Leftrightarrow \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} = \begin{pmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{pmatrix}$

$\Leftrightarrow X_1'X_2 = 0.$

(e) If $X_1'X_2 = 0$, then $\hat{\beta}_1 = [X_1'(I-P_{X_2})X_1]^{-1}X_1'(I-P_{X_2})\underline{y}$ $= (X_1'X_1)^{-1}X_1'\underline{y}$, since $P_{X_2}X_1 = 0$. That is, $(X_1'X_1)^{-1}X_1'\underline{y}$ is the least squares estimator of β_1.

If $(X_1'X_1)^{-1}X_1\underline{y}$ is the least squares estimator of β_1, then $E[(X_1'X_1)^{-1}X_1\underline{y}] = (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2)$ $= \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 \equiv \beta_1$ since the estimator is an unbiased estimator of β_1. This implies that $(X_1'X_1)^{-1}X_1'X_2 = 0$, that is, $X_1'X_2 = 0$.

(f) Note that $\hat{\beta}_1$ is the least squares estimator of β_1 for $\underline{y} = X_1\beta_1 + \underline{\varepsilon}$ $\Leftrightarrow \hat{\beta}_1 = (X_1'X_1)^{-1}X_1'\underline{y}$

$\Leftrightarrow [X_1'(I-P_{X_2})X_1]^{-1}X_1'(I-P_{X_2})\underline{y} = (X_1'X_1)^{-1}X_1'\underline{y}$

$\Leftrightarrow X_1'(I-P_{X_2})\underline{y} = [X_1'(I-P_{X_2})X_1]\cdot(X_1'X_1)^{-1}X_1'\underline{y}$

$\Leftrightarrow X_1'P_{X_2}(I-P_{X_1})\underline{y} = \underline{0}$

$\Leftrightarrow X_1'X_2(X_2'X_2)^{-1}[X_2'(I-P_{X_1})X_2]\hat{\underline{\beta}}_2 = \underline{0}$

(since $\hat{\underline{\beta}}_2 = [X_2'(I-P_{X_1})X_2]^{-1}X_2'(I-P_{X_1})\underline{y}$ by the same method for deriving $\hat{\beta}_1$ as in part (b))

$\Leftrightarrow [X_1'(I-P_{X_2})X_1]\cdot(X_1'X_1)^{-1}(X_1'X_2\hat{\beta}_2) = \underline{0}$

$\Leftrightarrow X_1'X_2\hat{\beta}_2 = \underline{0}$ (since $X_1'(I-P_{X_2})X_1$ is nonsingular by part (a)).

(g) Note that $\hat{\beta}_2 \sim N(\beta_2, \sigma^2[X_2'(I-P_{X_1})X_2]^{-1})$ and $\frac{(n-p-q)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p-q)$, where $\hat{\sigma}^2 = \frac{\underline{y}'(I-P_X)\underline{y}}{n-p-q}$

Since $\hat{\beta}_2$ and $\hat{\sigma}^2$ are independent,

$$\frac{\hat{\beta}_2'X_2'(I-P_{X_1})X_2\hat{\beta}_2}{q\hat{\sigma}^2} \sim F_{q,n-p-q}$$ under H_0.

Thus a size-r F test is to reject H_0 if

$$\frac{\hat{\beta}_2'X_2'(I-P_{X_1})X_2\hat{\beta}_2}{q\hat{\sigma}^2} > F_{r:q,n-p-q}.$$

The noncentrality parameter of the F statistic under Ha is $\frac{1}{2\sigma^2}\beta_2'X_2'(I-P_{X_1})X_2\beta_2$ The expected values of the numerator and denominator of the test statistic are $q\sigma^2 + \beta_2'X_2'(I-P_{X_1})X_2\beta_2$ and $q\sigma^2$, respectively.

(h) Let $\phi = \underline{a}'\underline{\beta}/\underline{b}'\underline{\beta}$ and $L = (\underline{a}-\phi\underline{b})'\hat{\underline{\beta}}$, where $\hat{\underline{\beta}}$ is the least squares estimator of $\underline{\beta}$. Note that L is normally distributed with $E(L)=0$ and $\sigma_L^2 \equiv Var(L) = (\underline{a}-\phi\underline{b})'(X'X)^{-1}(\underline{a}-\phi\underline{b})\cdot\sigma^2$. Since $\hat{\sigma}^2 \equiv \underline{y}'(\underline{I}-P_X)\underline{y}/(n-p-q)$ and $\hat{\underline{\beta}}$ are independent, we have that

$$T = \frac{L/\sigma_L}{\sqrt{\hat{\sigma}^2/\sigma^2}} \sim t_{n-p-q}.$$

Hence a $(1-\gamma)$ confidence set for ϕ is given by

$$T^2 \le (t_{\gamma/2:n-p-q})^2 = F_{\gamma:1,n-p-q},$$

or by $(\underline{a}-\phi\underline{b})'\left[\hat{\underline{\beta}}\hat{\underline{\beta}}' - F_{\gamma:1,n-p-q}\cdot\hat{\sigma}^2(X'X)^{-1}\right](\underline{a}-\phi\underline{b}) \le 0$.

(i) The BLUP $\tilde{\underline{\beta}}$ of $\underline{\beta}$ is given by the solution to the mixed-model equations $\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & D^{-1}+X_2'X_2 \end{pmatrix}\begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{pmatrix} = \begin{pmatrix} X_1'\underline{y} \\ X_2'\underline{y} \end{pmatrix}$.

Thus, $\begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{pmatrix} = \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & D^{-1}+X_2'X_2 \end{pmatrix}^{-1}\begin{pmatrix} X_1\underline{y} \\ X_2\underline{y} \end{pmatrix}$. The distribution of $\tilde{\underline{\beta}}-\underline{\beta}$ is $N\left(\underline{0}, \sigma^2\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & D^{-1}+X_2'X_2 \end{pmatrix}^{-1}\right)$.

(j) Let $D_k = k\cdot I$, $k=1,2,\cdots$ Then $\tilde{\underline{\beta}}^{(k)} \to \hat{\underline{\beta}}$ as $k\to+\infty$, where $\tilde{\underline{\beta}}^{(k)}$ is the BLUP of $\underline{\beta}$ for the model with $\beta_2 \sim N(\underline{0}, \sigma^2 k I)$.

1. Commercial operations must sometimes store grain for long periods of time before it is used to make food products. The quality of stored grain diminishes as storage time increases. The loss of quality can be slowed by treating grain with a preservative. An experiment was conducted to compare the effectiveness of four grain preservatives A, B, C, and D.

   A total of 24 sacks of grain were randomly sampled from a much larger quantity of grain. Each of the four preservatives was independently applied to 6 sacks randomly selected from the 24 so that each sack was treated with exactly one of the 4 preservatives. After treatment with the preservatives, the grain sacks were stored in 6 coolers. Four grain sacks were randomly assigned to each of the 6 coolers subject to the constraint that all four preservatives were represented within any given cooler. Three temperatures (10°, 20°, and 30° C) were assigned to the 6 coolers in a completely randomized fashion with 2 coolers set at each of the 3 temperatures. After 18 months of storage at the assigned temperature, two subsamples were randomly selected from each sack of grain. A destructive measurement procedure was applied to each subsample to obtain a grain quality score for each subsample. The following model was fit to the 48 grain quality scores.

$$Y_{ijkl} = \mu_{ik} + C_{j(i)} + (CP)_{jk(i)} + e_{ijkl}. \tag{1}$$

   Here Y_{ijkl} denotes the lth measurement of grain quality from the sack treated with preservative k and stored in the jth cooler set at the ith temperature. The parameter μ_{ik} is the unknown mean grain quality score associated with temperature i and preservative k. The remaining model terms are random effect terms corresponding to (in order of appearance from left to right) coolers nested within temperatures, the interactions between preservatives and coolers nested within temperatures, and error. Model assumptions include independence of all random effects and

$$C_{j(i)} \sim N(0, \sigma_C^2) \qquad (CP)_{jk(i)} \sim N(0, \sigma_{CP}^2) \qquad e_{ijkl} \sim N(0, \sigma^2).$$

   It is customary to write $\mu_{ik} = \mu + \tau_i + \pi_k + (\tau\pi)_{ik}$ where

$$\mu = \bar{\mu}_{\cdot\cdot}, \quad \tau_i = \bar{\mu}_{i\cdot} - \mu, \quad \pi_k = \bar{\mu}_{\cdot k} - \mu, \quad (\tau\pi)_{ik} = \mu_{ik} - \bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot k} + \mu.$$

   The parameters τ_i, π_k, $(\tau\pi)_{ik}$ are fixed effects corresponding, respectively, to the effect of the ith temperature, the effect of the kth preservative, and the interaction effect for the ith temperature/jth preservative combination. The sums of squares and expected mean squares associated with the model terms are provided below.

Source	Sum of Squares	Expected Mean Squares
temperature	179.7	$8\sum_{i=1}^{3}\tau_i^2 + 8\sigma_C^2 + 2\sigma_{CP}^2 + \sigma^2$
cooler(temperature)	145.4	$8\sigma_C^2 + 2\sigma_{CP}^2 + \sigma^2$
preservative	1115.8	$4\sum_{k=1}^{4}\pi_k^2 + 2\sigma_{CP}^2 + \sigma^2$
temperature*preservative	7.0	$\frac{2}{3}\sum_{i=1}^{3}\sum_{k=1}^{4}(\tau\pi)_{ik}^2 + 2\sigma_{CP}^2 + \sigma^2$
cooler*preservative(temperature)	10.2	$2\sigma_{CP}^2 + \sigma^2$
error	26.4	σ^2

   One or more of the following parts may require the use of Satterthwaite's method for approximating degrees of freedom associated with a linear combination of mean squares. The approximate degrees of freedom associated with a linear combination of mean squares of the form

$$\sum_{i=1}^{I} a_i \text{MS}_i \qquad \text{is given by} \qquad \frac{\left(\sum_{i=1}^{I} a_i \text{MS}_i\right)^2}{\sum_{i=1}^{I}(a_i \text{MS}_i)^2/df_i}$$

   where df_i denotes the degrees of freedom associated with the mean square MS_i.

(a) Provide degrees of freedom corresponding to each sum of squares in the ANOVA table.

(b) Derive the expected mean square for temperature.

(c) Compute an F-statistic for testing the significance of temperature main effects.

(d) Compute an F-statistic for testing the significance of preservative main effects.

(e) Compute the standard error of the estimated mean grain quality score for grain treated with preservative B and stored at 20° C.

(f) Give the degrees of freedom associated with the standard error in part (e).

(g) Compute the standard error of the estimated mean grain quality score for grain stored at 20° C.

(h) Give the degrees of freedom associated with the standard error in part (g).

(i) Write down the covariance between the grain quality scores of two subsamples taken from the same sack. You do not need to estimate the covariance. Just give an expression for the covariance in terms of the model parameters.

(j) Write down the covariance between the grain quality scores of two subsamples taken from two different sacks within the same cooler. Give an expression for the covariance in terms of the model parameters and provide an estimate of this covariance.

2. Question 1 involves the analysis of 48 grain quality scores consisting of 2 measurements on each of 24 sacks of grain. Now suppose that the experiment was conducted with 48 sacks of grain with only a single measurement for each sack. There would still be two measurements for each preservative in each cooler, but these two measurements would come from two sacks independently treated with the preservative rather than from a single sack.

   (a) Explain why model (1) may not be the best choice for the data from 48 sacks.

   (b) Write down an alternative model for the analysis of the data from 48 sacks.

   (c) Suppose the sums of squares reported in question 1 are actually the sums of squares obtained by fitting model (1) to the data from 48 sacks. Show how you would use the sums of squares from the fit of model (1) to construct F-statistics for testing (i) for differences among the temperatures and (ii) for differences among the preservatives, given that the data consist of a single grain quality measure on each of 48 sacks.

3. Suppose we are again concerned about studying the two factors storage temperature (10°, 20°, and 30° C) and preservative (A, B, C, and D). As in question 1, we will work with 24 sacks of grain. Suppose, unlike in question 1, that the 24 sacks were obtained as follows. Eight lots of grain were randomly sampled from a large quantity of grain. The 4 preservatives were assigned to the 8 lots in a balanced and completely randomized fashion. After each lot was independently treated with its assigned preservative, each lot was randomly divided into 3 sacks of grain. One measurement of grain quality is to be obtained for each sack after 18 months of storage.

   (a) Suppose that the 3 temperatures have been randomly assigned to the 6 coolers exactly as in question 1. Explain how you would randomly assign the 24 sacks of grain to the 6 coolers. You might draw a picture indicating the locations of the sacks of grain within coolers and the lots from which each sack originated.

   (b) Suppose the experiment will be carried out as you have designed it in part (a). Write down a potentially appropriate model for the data and explain the reasons for your choice.

1. (a) The degrees of freedom are provided below. Note that the df for error can be obtained by subtraction or by recognizing the error as subsample nested within sack.

```
Source                                                                  DF
--------------------------------------------------------------------------
temperature                                                    I-1=3-1= 2
cooler(temperature)                                       I(J-1)=3(2-1)= 3
preservative                                                   K-1=4-1= 3
temperature*preservative                          (I-1)(K-1)=(3-1)(4-1)= 6
cooler*preservative(temperature)  I(J-1)(K-1)=3(2-1)(4-1)= 9
error                                               IJK(L-1)=(2-1)3*2*4=24
```

(b)

$$
\begin{aligned}
&E\left[\frac{1}{I-1}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L}(\bar{Y}_{i\cdots}-\bar{Y}_{\cdots\cdot})^2\right] \\
&= \frac{1}{I-1}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L}E\left[(\tau_i+\bar{C}_{\cdot(i)}-\bar{C}_{\cdot(\cdot)}+\overline{CP}_{\cdot\cdot(i)}-\overline{CP}_{\cdot\cdot(\cdot)}+\bar{e}_{i\cdots}-\bar{e}_{\cdots\cdot})^2\right] \\
&= \frac{1}{I-1}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L}\left[\tau_i^2+\mathrm{Var}(\bar{C}_{\cdot(i)}-\bar{C}_{\cdot(\cdot)}+\overline{CP}_{\cdot\cdot(i)}-\overline{CP}_{\cdot\cdot(\cdot)}+\bar{e}_{i\cdots}-\bar{e}_{\cdots\cdot})\right] \\
&= \frac{JKL}{I-1}\sum_{i=1}^{I}\tau_i^2+\frac{IJKL}{I-1}\mathrm{Var}(\bar{C}_{\cdot(i)}-\bar{C}_{\cdot(\cdot)}+\overline{CP}_{\cdot\cdot(i)}-\overline{CP}_{\cdot\cdot(\cdot)}+\bar{e}_{i\cdots}-\bar{e}_{\cdots\cdot})
\end{aligned}
$$

Now note that

$$
\begin{aligned}
\mathrm{Var}(\bar{C}_{\cdot(i)}-\bar{C}_{\cdot(\cdot)}) &= \frac{\sigma_C^2}{J}+\frac{\sigma_C^2}{IJ}-2\frac{\sigma_C^2}{IJ}=\frac{(I-1)\sigma_C^2}{IJ} \\
\mathrm{Var}(\overline{CP}_{\cdot\cdot(i)}-\overline{CP}_{\cdot\cdot(\cdot)}) &= \frac{\sigma_{CP}^2}{JK}+\frac{\sigma_{CP}^2}{IJK}-2\frac{\sigma_{CP}^2}{IJK}=\frac{(I-1)\sigma_{CP}^2}{IJK} \\
\mathrm{Var}(\bar{e}_{i\cdots}-\bar{e}_{\cdots\cdot}) &= \frac{\sigma^2}{JKL}+\frac{\sigma^2}{IJKL}-2\frac{\sigma^2}{IJKL}=\frac{(I-1)\sigma^2}{IJKL}
\end{aligned}
$$

The result follows.

(c) $F=\frac{179.7/2}{145.4/3}\approx 1.85$

(d) $F=\frac{1115.8/3}{10.2/9}\approx 328.18$

(e)

$$
\mathrm{Var}(\bar{Y}_{i\cdot k\cdot})=\mathrm{Var}(\bar{C}_{\cdot(i)}+\overline{CP}_{\cdot k(i)}+\bar{e}_{i\cdot k\cdot})=\frac{2\sigma_C^2+2\sigma_{CP}^2+\sigma^2}{4}.
$$

The variance is estimated by

$$
\frac{\mathrm{MS}_{C(T)}+3\mathrm{MS}_{CP(T)}}{16}=\frac{(145.4/3)+3(10.2/9)}{16}\approx 3.24
$$

The standard error is approximately $\sqrt{3.24}=1.8$

(f) The approximate degrees of freedom are

$$\frac{(\frac{1}{16}\mathrm{MS}_{C(T)} + \frac{3}{16}\mathrm{MS}_{CP(T)})^2}{(\frac{1}{16}\mathrm{MS}_{C(T)})^2/3 + (\frac{3}{16}\mathrm{MS}_{CP(T)})^2/9} \approx 3.43.$$

(g)

$$\mathrm{Var}(\bar{Y}_{i\cdots}) = \mathrm{Var}(\bar{C}_{\cdot(i)} + \overline{CP}_{\cdot\cdot(i)} + \bar{e}_{i\cdots}) = \frac{8\sigma_C^2 + 2\sigma_{CP}^2 + \sigma^2}{16}.$$

The variance is estimated by

$$\mathrm{MS}_{C(T)}/16 \approx 3.029.$$

The standard error is approximately $\sqrt{3.029} = 1.74$

(h) The degrees of freedom are the degrees of freedom for cooler nested within temperatures (i.e., 3).

(i) $\mathrm{Cov}(Y_{ijk1}, Y_{ijk2}) = \mathrm{Var}(C_{j(i)} + (CP)_{jk(i)}) = \sigma_C^2 + \sigma_{CP}^2$

This covariance is estimated by

$$\left[\left(\frac{145.4}{3}\right) + 3\left(\frac{10.2}{9}\right) - 4\left(\frac{26.4}{24}\right)\right]/8 = 5.9\bar{3}.$$

(j) $\mathrm{Cov}(Y_{ijkl}, Y_{ijk'l}) = \mathrm{Var}(C_{j(i)}) = \sigma_C^2$
This covariance is estimated by

$$\left[\left(\frac{145.4}{3}\right) - \left(\frac{10.2}{9}\right)\right]/8 = 5.91\bar{6}.$$

2. (a) With model (1), the correlation between two sacks within a given cooler is greater if the two sacks have been given the same treatment than if the two sacks have been given different treatments. (See parts (i) and (j) of question 1.) This does not seem appropriate because all the sacks within a cooler have been treated independently. The pairwise correlations between any two sacks in a given cooler would usually be assumed to be constant. Any additional similarity between sacks given the same treatment is accounted for by the mean structure.

(b) Just drop the $(CP)_{jk(i)}$ terms from model (1).

(c) The test for differences among temperatures is exactly as before. The denominator for the preservative test becomes $(10.2 + 26.4)/(9 + 24) \approx 1.109$.

3. (a) Randomly divide the 6 coolers into 2 blocks of 3 coolers so that each temperature is represented exactly once in each block. Randomly divide the 8 lots of grain into 2 blocks of 4 lots so that each preservative is represented by exactly one lot in each block of lots. Randomly match each block of coolers with a block of lots. Assign the 12 sacks of grain from a block of lots to its corresponding block of 3 coolers so that each preservative is represented exactly once in each cooler.

(b)

$$Y_{ijk} = \mu + R_i + \tau_j + (RT)_{ij} + \pi_k + (RP)_{ik} + (\tau\pi)_{jk} + e_{ijk}$$

Here R_i denotes the effect associated with the ith "replication" of the experiment. (Call a pairing of a block of lots with a block of coolers one replication of the experiment. There are two replications in this case. The replications may be viewed as fixed or random. The decision has no bearing on the inferences of interest in this case.) The notation for the fixed effects is the same as that used previously. The interaction of each fixed term with replication forms the three error terms (rep-by-temp for testing temperature, rep-by-preservative for testing preservative, and rep-by-temp-by-preservative (e_{ijk}) for testing temp-by-preservative interaction). These rep-by-temp and rep-by-preservative terms allow for correlation between the grain quality measures from sacks in the same cooler and sacks from the same lot, respectively. A break down of the degrees of freedom is provided below.

SOURCE	DF
rep	1
temp	2
temp*rep	2
pres	3
pres*rep	3
temp*pres	6
temp*pres*rep	6

This problem describes a series of modeling questions related to the following basic problem. Economists would like to find out how often individuals use resources like state parks and how that might change if entry prices were increased. This information is useful for policymakers in deciding how much to charge as a usage fee. Surveys of consumers are the main source of data in such studies.

As an example suppose we wish to study the behavior of the public with respect to the Imaginary State Park (ISP) in Iowa. A survey of Iowa residents is carried out. Respondents provide the number of visits to the ISP over the last 12 months (let's call this y_i for respondent i) and a number of other variables (the self-reported cost of such trips, age of respondent, family size, etc.). We think of these other variables as covariates, let's call the vector of such variables for the ith respondent x_i.

1. The basic linear regression model can be used for a starting point in such analyses. Assuming there are n survey respondents, we have:

$$Y = X\beta + \epsilon$$

where Y is $n \times 1$ vector of responses, X is $n \times p$ matrix of explanatory variables, β is $p \times 1$ vector of parameters (including the intercept) and $\epsilon \sim N(0, \sigma^2 I)$ is the $n \times 1$ vector of disturbances or errors.

(a) Suppose the n survey respondents come from J different regions. Set up a single regression model that allows for separate vectors of regression coefficients (β_j) in each region. Be sure to define any new variables you introduce.

(b) We wish to test the hypothesis that all regions have the same population regression coefficients. Describe how you would carry out a statistical test of this hypothesis.

(c) Assume the hypothesis tests fails to reject the hypothesis of equal regression coefficients. We may still be concerned about the effect of region in that it may not be appropriate to model the respondents from the same region as independent of each other. Describe how you might check for such non-independence.

(d) One way to address a lack of independence is to take the regression coefficients for each region β_j to be a draw from a $N(\beta_o, \Sigma_\beta)$ population, that is to treat the regression coefficients in region j as random effects. Define Y_j as the response vector for those individuals in region j. Compute the marginal distribution for Y_j (integrating out β_j) thereby show that the use of random effects allows for dependence in the distribution of Y_j.

2. For the remainder of the problem (this part and the next one) assume that we can use a single vector of regression coefficients for all respondents (that is you can ignore region). A limitation of the linear model used in part 1 is that the response y_i is restricted to be nonnegative (note that it an be zero).

   (a) One alternative is known as a Tobit model. The Tobit model assumes $y_i = 0$ if $(x_i^t\beta) + \epsilon_i < 0$ and $y_i = x_i^t\beta + \epsilon_i$ otherwise. Assuming ϵ_i is a normal random variable (as in part 1), write down the joint distribution for the vector of responses Y.

   (b) A second alternative is to model y_i as a Poisson random variable with mean λ_i and assume $\log \lambda_i = x_i^t\beta$. Write down the joint distribution for the vector of responses in this case.

   (c) Which of the two solutions seems best in the present context? Explain why.

   (d) For the solution you think best, describe the statistical method you would use to fit the model. Also, describe how you would obtain estimated standard errors of the estimated coefficients.

3. A recent study of this type obtained the data by interviewing visitors to the park. There are a couple of potential difficulties with collecting data in this way.

   (a) First, this design implies that all of the survey respondents have at least one visit! Explain how you would construct a probability model for Y to accommodate this aspect of the design.

   (b) The second difficulty is that people with many visits to the park are more likely to be included in the sample than those with few (or no) visits. This means the sample is not representative of the population. Discuss how you might address this problem.

(1.a) Define $z_{ij} = 1$ if individual i comes from region j and zero otherwise. Then the response for a single individual can be written as $y_i = \sum_{j=1}^{J} z_{ij} x_i^t \beta_j + \epsilon_i$. In matrix form we can achieve this by creating a new X matrix and a new β vector. Each row ($1 \times Jp$ in length) in the X matrix contains J sets of p columns; all the columns in a row take the value zero except for the set of columns corresponding to the correct region for that row. The new β vector has all of the region's coefficient vectors concatenated into a single vector.

(1.b) This is easily done using traditional regression methods. Fit the reduced model (a single set of regression coefficients as in the original problem set up) and the full model as in (1.a). Then $F = (SSE(reduced) - SSE(full))/((J-1)p)/MSE(full)$ can be used to test the null hypothesis; under the null hypothesis it has an $F_{(J-1)p,n-Jp}$ distribution.

(1.c) Examine the distribution of residuals for each region and look for patterns across the different regions.

(1.d) Let X_j be the relevant part of the X matrix. Then we have $Y_j \sim N(X_j\beta_j, \sigma^2 I)$ and $\beta_j \sim N(\beta_o, \Sigma_\beta)$. The marginal distribution of Y_j is then $N(X_j\beta_o, X_j\Sigma_\beta X_j^t + \sigma^2 I)$. This last covariance matrix is not necessarily diagonal thus we have introduced some intra-region correlation.

(2.a) With the normal assumption, we find $\Pr(y_i = 0) = \Phi(-x_i^t\beta/\sigma)$. Then

$$p(Y|\beta,\sigma) = \prod_{i=1}^{n} \Phi(-x_i^t\beta/\sigma)^{I_{(y_i=0)}} \left[\phi((y_i - x_i^t\beta)/\sigma)/\sigma\right]^{I_{(y_i>0)}}$$

where I_A is an indicator for the event A, ϕ is the standard normal density, and Φ is the standard normal cumulative distribution function.

(2.b)

$$p(Y|\beta) = \prod_{i=1}^{n} \left(\frac{e^{-e^{x_i^t\beta}} e^{(x_i^t\beta)y_i}}{y_i!} \right)$$

(2.c) Since the number of visits is an integer the Poisson model seems more appropriate than the Tobit model.

(2.d) Both models are generally fit using the method of maximum likelihood. The estimated standard errors would be obtained by evaluating the inverse of the observed Fisher information matrix (inverse of the negative second derivative matrix of the likelihood evaluated at the MLE).

(3.a) This is a truncation or censoring problem. One way to address this is by starting with the Poisson model of (2.b). We replace the Poisson density by the truncated Poisson with no zero values. The pdf for this random variable is $e^{-\lambda}\lambda^{y}/(y_i!(1-e^{-\lambda}))$. Note the mean is now $\lambda/(1-e^{-\lambda})$; I would still be inclined to take $\log\lambda = x^{t}\beta$. It is also possible to modify the Tobit model of (2.a) to accommodate the absence of zero values.

(3.b) Wow, that's a hard one! One idea is to perhaps get some information about the distribution of the number of visits among the population and try to use that information to weight the sample cases to provide a reasonable population inference.

A university researcher is interested in studying the environmental health of natural ponds in Maine. So far, she has collected data on 30 ponds, including the following variables:

- an index of ecological health, on a scale 0–100 (Y)
- surface area of the pond (X_1)
- distance to nearest urbanized area (X_2).

She decides to use regression to study relationships between the variables and develop a predictive ecological health model.

a) The researcher runs an Ordinary Least Squares (OLS) regression on these data and generates parameter estimates (partial S-Plus output shown below).

```
Call: lm(formula = Y ~ X1 + X2)

Coefficients:
               Value Std. Error t value Pr(>|t|)
(Intercept)  1.6544 21.3689     0.0774  0.9389
         X1  3.3076  2.0213      1.6363  0.1134
         X2  1.9951  0.8324      2.3967  0.0237

Residual standard error (MSE): 26.8 on A degrees of freedom
Multiple R-Squared: 0.3105
F-statistic: 6.081 on B and C degrees of freedom, p-value is 0.0066
```

Write down the statistical model that is required for the OLS parameter estimators to be Best Linear Unbiased Estimators (BLUE), and provide conditions for the estimates to be unique.

b) The degrees of freedom in the above S-Plus output (indicated by `A, B, C`) appear to be missing. Provide the missing values.

c) A residual plot (not shown) clearly indicates that as the pond surface area (X_1) increases, the variability in the errors increases. What is the effect of this finding on the properties of the OLS parameter estimators? What is its effect on the estimators of the standard errors of the parameter estimators?

d) Assume now that the variance of the model errors is of the form $\sigma^2 X_1$, with σ^2 an *unknown* parameter. The researcher performs a Weighted Least Squares (WLS) regression, using the values $1/X_1$ as the weights for each observation (partial S-Plus output shown below).

```
Call: lm(formula = Y ~ X1 + X2, weights = 1/X1)

Coefficients:
                Value Std. Error  t value Pr(>|t|)
(Intercept)  -2.3110  17.2382    -0.1341   0.8943
         X1   3.5860   1.7509     2.0480   0.0504
         X2   2.1030   0.8246     2.5503   0.0167

Residual standard error: 7.811 on A degrees of freedom
Multiple R-Squared: 0.3749
F-statistic: 8.097 on B and C degrees of freedom, p-value is 0.00176
```

Are the WLS parameter estimators BLUE in this case?

e) Propose an unbiased estimator for the variance parameter σ^2 (call it $\hat{\sigma}^2$) under the heteroskedastic model described in (c), and show that it is unbiased. Write down an unbiased estimator for the variance-covariance matrix of the parameter estimators under this model (call it $\hat{V}_\beta$).

f) The researcher had an initial hypothesis that this model would have the following restriction on the parameter values: $\beta_1 + \beta_2 = \beta_0$. Write down a formal hypothesis test for this hypothesis and propose a test statistic. Write down any additional assumptions on the model that are required for this test statistic to have a known distribution under the null hypothesis, and specify that distribution. Explain carefully how you would calculate this statistic for these data.

g) Suppose that

$$\hat{V}_\beta = \begin{bmatrix} 38.045 & -3.447 & -0.072 \\ -3.447 & 0.393 & -0.063 \\ -0.072 & -0.063 & 0.087 \end{bmatrix}$$

Perform the test described in (e) at the 95% confidence level, and state your conclusion.

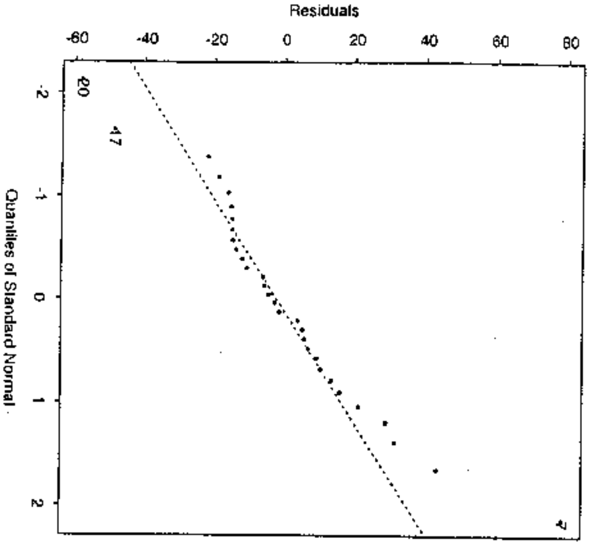
h) The researcher is concerned that the linear model assumed so far might not hold for all the ponds. In order to check for non-linearities in the effect of surface area (X_1), she wants to split the effect of X_1 into two parts: the effect for ponds with surface area larger than 10, and those with surface area smaller than or equal to 10. Write down a regression model that will make it possible to capture this type of differential effects. Then, write down a hypothesis for testing the significance of this non-linearity, and propose a test statistic. State any assumptions you would need in order to allow specification of the distribution of the test statistic, and state that distribution.

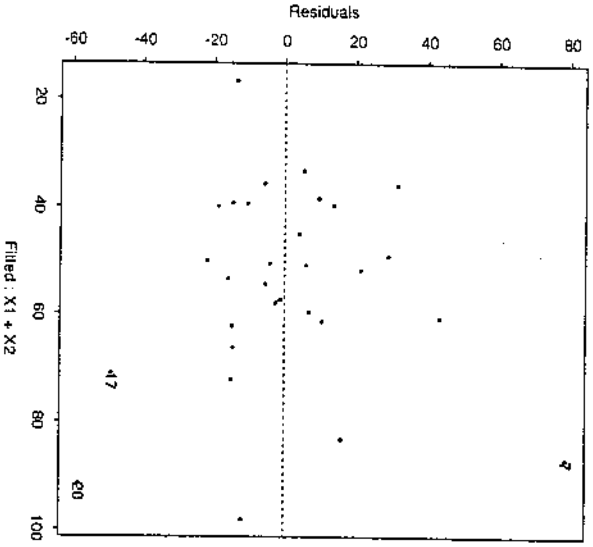
i) The normal quantile residual plot and other residual plots (on separate page) indicate some departure from the normal distribution. Discuss at least three possible solutions to achieve residuals that are closer to being normally distributed in the context of these data (be specific).

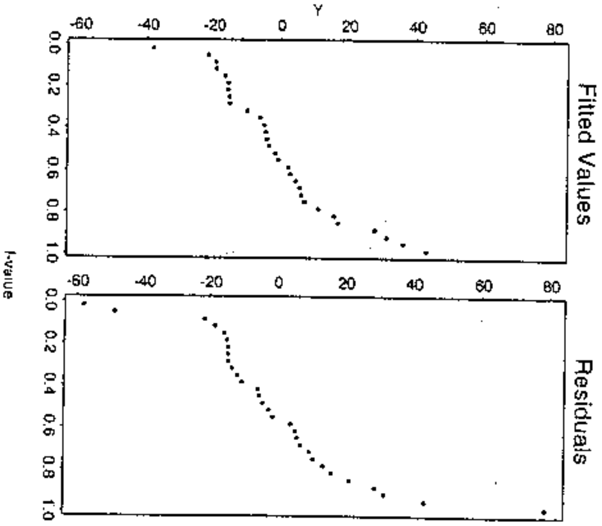
j) The researcher is confident that variable X_2 is measured with reasonable precision, but worries that X_1 might be subject to some error. To study the effect of error in X_1 on the parameter estimators, assume the following model for X_1:

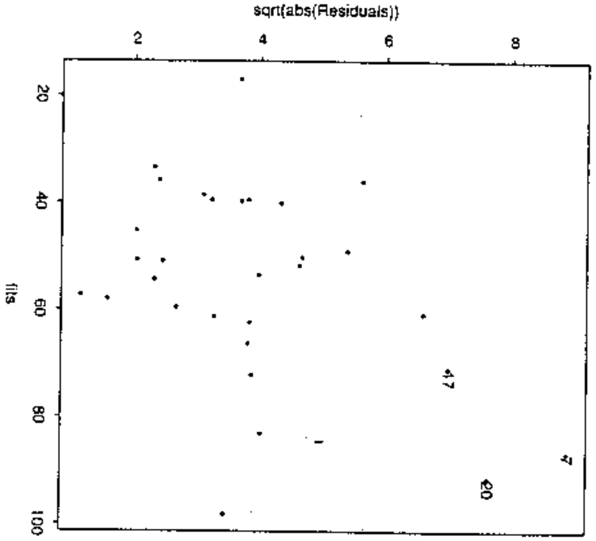
$$X_{1i} = Z_{1i} + \delta_i,$$

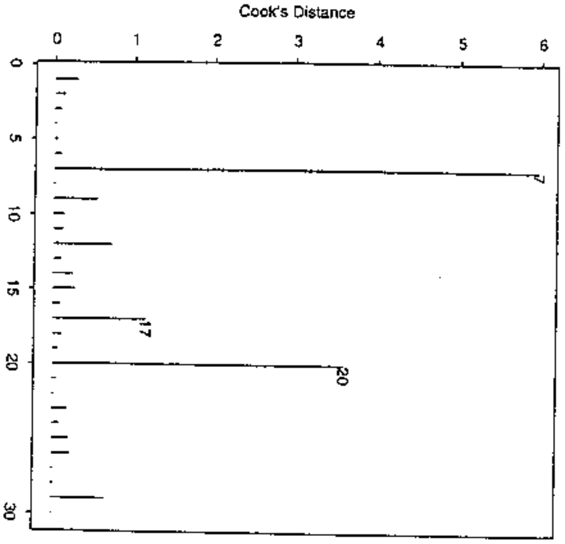
where the Z_{1i} are the true pond surface areas and the δ_i are independent and identically distributed errors with mean 0 and common variance $\tau^2 > 0$. The δ_i are assumed independent of the regression model errors ε_i. Hence, the model generating the pond ecological health index Y is a linear model in the variables Z_1 and X_2, but the *observed* data include only Y, X_1, X_2, not Z_1. Show that both WLS and OLS produce biased estimators of the model parameters in this situation.

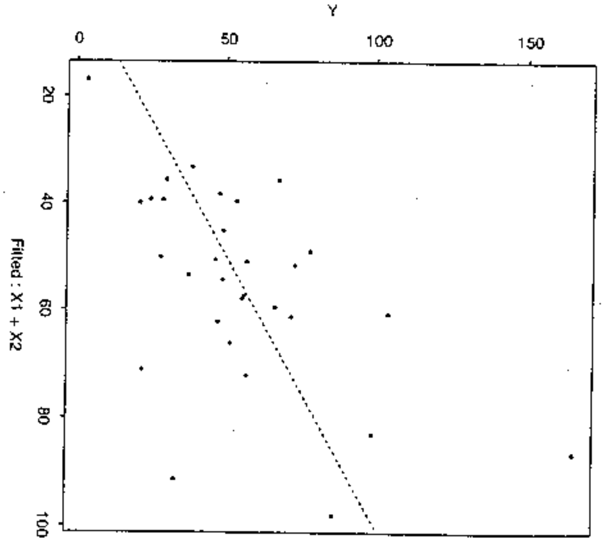

Residuals
-60
-40
-20
0
20
40
60
80
Quantiles of Standard Normal
-2
-1
0
1
2
20
47
7


Residuals
-60
-40
-20
0
20
40
60
80
Fitted : X1 + X2
20
40
60
80
100
47
20
7


Y
-60
-40
-20
0
20
40
60
80
0.0
0.2
0.4
0.6
0.8
1.0
Fitted Values
f-value
Residuals


sqrt(abs(Residuals))
2
4
6
8
fits
20
40
60
80
100
47
20
7


Cook's Distance
0
1
2
3
4
5
6
0
5
10
15
20
30
7
17
20


Y
0
50
100
150
Fitted : X1 + X2
20
40
60
80
100

Prelim 2002: Methods III SOLUTIONS

A university researcher is interested in studying the environmental health of natural ponds in Maine. So far, she has collected data on 30 ponds, including the following variables:

- an index of ecological health, on a scale 0–100 (Y)
- surface area of the pond (X_1)
- distance to nearest urbanized area (X_2)

She decides to use regression to study relationships between the variables and develop a predictive ecological health model.

a) The researcher runs an Ordinary Least Squares (OLS) regression on these data and generates parameter estimates (partial S-Plus output shown below).

```
Call: lm(formula = Y ~ X1 + X2)

Coefficients:
               Value Std. Error t value Pr(>|t|)
(Intercept)  1.6544 21.3689      0.0774  0.9389
         X1  3.3076  2.0213      1.6363  0.1134
         X2  1.9951  0.8324      2.3967  0.0237

Residual standard error (MSE): 26.8 on A degrees of freedom
Multiple R-Squared: 0.3105
F-statistic: 6.081 on B and C degrees of freedom, p-value is 0.0066
```

Write down the statistical model that is required for the OLS parameter estimators to be Best Linear Unbiased Estimators (BLUE), and provide conditions for the estimates to be unique.

Answer:

The linear model (in vector notation)

$$Y = X^T\beta + \varepsilon$$
$$\varepsilon \ distr.\ F(0, \sigma^2 I)$$

for some distribution F (normality not required).

The estimates will be unique if $X^T X$ is invertible.

b) The degrees of freedom in the above S-Plus output (indicated by `A, B, C`) appear to be missing. Provide the missing values. Write down the hypothesis test corresponding to the F-statistic in the last line of the S-Plus output, and state its conclusion at the 95% confidence level.

Answer:
The MSE has $n - p$ degrees of freedom, or 27 in this case.

The F-statistic has $p - 1$ and $n - p$ degrees of freedom, or 2 and 27. That statistic is used to test the hypothesis

$$H_0 : \beta_1 = \beta_2 = 0 \quad \text{vs.} \quad H_a : \text{at least one of both parameters} \neq 0$$

In this case, the p-value is much smaller than 5%, so we reject H_0 at that confidence level.

c) A residual plot (not shown) clearly indicates that as the pond surface area (X_1) increases, the variability in the errors increases. What is the effect of this finding on the properties of the OLS parameter estimators? What is its effect on the estimators of the standard errors of the parameter estimators?

Answer:
The OLS parameter estimators remain unbiased. However, their variance is no longer the smallest possible, so that the estimator is no longer BLUE. The estimators of the standard errors of the parameters are biased.

d) Assume now that the variance of the model errors is of the form $\sigma^2 X_1$, with σ^2 an *unknown* parameter. The researcher performs a Weighted Least Squares (WLS) regression, using the values $1/X_1$ as the weights for each observation (partial S-Plus output shown below).

```
Call: lm(formula = Y ~ X1 + X2, weights = 1/X1)

Coefficients:
                Value Std. Error  t value Pr(>|t|)
(Intercept)  -2.3110  17.2382    -0.1341   0.8943
         X1   3.5860   1.7509     2.0480   0.0504
         X2   2.1030   0.8246     2.5503   0.0167

Residual standard error: 7.811 on A degrees of freedom
Multiple R-Squared: 0.3749
F-statistic: 8.097 on B and C degrees of freedom, p-value is 0.00176
```

Write down the assumed regression model. Are the WLS parameter estimators BLUE in this case?

Answer:

The linear model (in vector notation)

$$Y = X^T\beta + \varepsilon$$
$$\varepsilon \; distr. \; F(0, \sigma^2 W)$$

for some distribution F (normality not required), and $W = \text{diag}\{X_1\}$. The variance is known up to a proportionality constant, so the WLS estimator is BLUE.

e) Propose an unbiased estimator for the variance parameter σ^2 (call it $\hat{\sigma}^2$) under the heteroskedastic model described in (c), and show that it is unbiased. Write down an unbiased estimator for the variance-covariance matrix of the parameter estimators under this model (call it $\hat{V}_\beta$).

Answer:

Use

$$\hat{\sigma}^2 = (Y - X^T\hat{\beta}_{WLS})^T W^{-1}(Y - X^T\hat{\beta}_{WLS})/(n-p)$$

and show that it is unbiased by direct calculation of the denominator. The unbiased estimator for $\text{Var}(\hat{\beta}_{WLS})$ is

$$\hat{V}_\beta = \hat{\sigma}^2(X^T W^{-1} X)^{-1}.$$

f) The researcher had an initial hypothesis that this model would have the following restriction on the parameter values: $\beta_1 + \beta_2 = \beta_0$. Write down a formal hypothesis test for this hypothesis and propose a test statistic. Write down any additional assumptions on the model that are required for this test statistic to have a known distribution under the null hypothesis, and specify that distribution. Explain carefully how you would calculate this statistic for these data.

Answer:

Let $c = (1, -1, -1)^T$. The hypothesis test is set up as

$$H_0 : c^T\beta = 0 \quad \text{vs.} \quad H_a : \text{the equality does not hold}$$

The test statistic can be written as

$$F = \frac{c^T\hat{\beta}(c^T(X^T W^{-1} X)^{-1} c)^{-1} c^T\hat{\beta}}{\hat{\sigma}^2}.$$

With a sample size of only 30, the model errors have to be assumed to be normal for this F-statistic to have be distributed F. Under that assumption, the degrees of freedom are (1,27).

Note: possible alternative answer: define the F-statistic with the same denominator but with the numerator written as the difference in normalized sums of squares between a full and a reduced model. Specifically, the numerator is $MSE_{dif} = (SSE_{full} - SSE_{red})/1$ with

$$\begin{aligned} SSE_{full} &= (Y - X^T\hat{\beta}_{WLS})^T W^{-1}(Y - X^T\hat{\beta}_{WLS}) \\ SSE_{full} &= (Y - X^T\hat{\beta}_{red})^T W^{-1}(Y - X^T\hat{\beta}_{red}) \end{aligned}$$

and the $\hat{\beta}_{red}$ are obtained by a separate regression. That regression would have to be a constrained least squares procedure that satisfies the hypothesis constaint. This answer will not work when trying to solve the next part, however...

g) Suppose that

$$\hat{V}_\beta = \begin{bmatrix} 38.045 & -3.447 & -0.072 \\ -3.447 & 0.393 & -0.063 \\ -0.072 & -0.063 & 0.087 \end{bmatrix}$$

Perform the test described in (e) at the 95% confidence level, and state your conclusion.

Answer:

We have $\hat{\sigma}^2 = 61.008$ ($= 7.811^2$, from the S-Plus output). Hence, the rest is staightforward calculation based on the provided variance-covariance matrix. Note that $\hat{V}_\beta = (X^T W^{-1} X)^{-1}\hat{\sigma}^2$, so that we can rewrite the F-statistic as

$$F = (c^T\hat{\beta})^2(c^T\hat{V}_\beta c)^{-1} = 8^2(45.437)^{-1} = 1.4085$$

This corresponds to a p-value of approximately 0.75, so we cannot reject H_0.

h) The researcher is concerned that the linear model assumed so far might not hold for all the ponds. In order to check for non-linearities in the effect of surface area (X_1), she wants to split the effect of X_1 into two parts: the effect for ponds with surface area larger than 10, and those with surface area smaller than or equal to 10. Write down a regression model that will make it possible to capture this type of differential effects. Then, write down a hypothesis for testing the significance of this non-linearity, and propose a test statistic. State any assumptions you would need in order to allow specification of the distribution of the test statistic, and state that distribution.

Answer:

Define new variables $X_{11} = X_1 * 1_{(X_1 \leq 10)}$ and $X_{12} = X_1 - X_{11}$, and define a correspondingly expanded regression model. The hypothesis we are interested in for this case is

$$H_0 : \beta_{11} = \beta_{12} \quad \text{vs.} \quad H_a : \text{the equality does not hold}$$

The test statistic can be constructed using either possible answer from part (e). It will have an F distribution with (1,26) degrees of freedom, if the model errors are normally distributed.

i) The normal quantile residual plot and other residual plots (on separate page) indicate some departure from the normal distribution. Discuss at least three possible solutions to achieve residuals that are closer to being normally distributed in the context of these data (be specific).

Answer:
I would expect a discussion including the following possible solutions: Box-Cox transformation of the Y, complete removal of the outliers (since it appears to be only a small number of points), alternative variance specifications, additional covariates, other...

j) The researcher is confident that variable X_2 is measured with reasonable precision, but worries that X_1 might be subject to some error. To study the effect of error in X_1 on the parameter estimators, assume the following model for X_1:

$$X_{1i} = Z_{1i} + \delta_i,$$

where the Z_{1i} are the true pond surface areas and the δ_i are independent and identically distributed errors with mean 0 and common variance $\tau^2 > 0$. The δ_i are assumed independent of the regression model errors ε_i. Hence, the model generating the pond ecological health index Y is a linear model in the variables Z_1 and X_2, but the *observed* data include only Y, X_1, X_2, not Z_1. Show that both WLS and OLS will generally produce biased estimators of the model parameters in this situation.

Answer:
This is the classical measurement setup. If the OLS estimator is calculated using X_1 instead of Z_1, the parameter estimator is defined as

$$\begin{aligned}\hat{\beta} &= (X^TX)^{-1}X^TY \\ &= (X^TX)^{-1}X^T(X\beta + \varepsilon - \delta\beta_1) \\ &= \beta + (X^TX)^{-1}X^T\varepsilon - (X^TX)^{-1}X^T\delta\beta_1\end{aligned}$$

The expectation of this quantity is therefore

$$\begin{aligned}\mathrm{E}(\hat{\beta}) &= \beta + \mathrm{E}((X^TX)^{-1}X^T\varepsilon) - \mathrm{E}((X^TX)^{-1}X^T\delta)\beta_1 \\ &= \beta - \mathrm{E}((X^TX)^{-1}X^T\delta)\beta_1\end{aligned}$$

with the last term not equal to 0, since $\mathrm{E}(X^T\delta) = (0, n\tau^2, 0)^T$, so that $\mathrm{E}((X^TX)^{-1}X^T\delta)$ $\neq 0$ (none of the three parameters is unbiased). The same reasoning applies to WLS.