STATISTICS 174: APPLIED STATISTICS

MIDTERM EXAM

OCTOBER 10, 2000

Time allowed: 75 minutes.

This is an open book exam: all course notes and the text are allowed, and you are expected to use your own calculator. Answers should preferably be written in a blue book.

The exam is expected to be your own work and no consultation during the exam is allowed. You are allowed to ask the instructor for clarification if you feel the question is ambiguous.

The three questions, numbered 1, 2 and 3, are connected, but are intended to be independent of one another, in the sense that you can do any one of them without referring to the other two. However, you are allowed to refer to your answers to one question in answering the others. In any case, my recommendation is that you do number 1 first, but this is not mandatory.

Within each of the three questions, the individual parts (a,b,c,..) are more interconnected, though if you answer later parts correctly, even though there may be errors in the earlier parts, you will still receive credit for the parts that are correct. In any case, I do not recommend that you spend more time on any individual part than the marks available would justify. The total marks for each part are indicated in square brackets (total for the whole exam: 100).

A table of the t distribution is provided. All other information you need should be already in the course notes. Please ask if you have any queries.

A two-variable regression problem features the following values of the covariates x_{i1} and x_{i2}:

i	x_{i1}	x_{i2}
1	–5	–1
2	–4	0
3	–3	1
4	–2	–1
5	–1	0
6	0	1
7	1	–1
8	2	0
9	3	1
10	4	–1
11	5	1

The model, for all parts of the exam, is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i,$$

where ϵ_i, $1 \le i \le 11$, are independent $N(0, \sigma^2)$ random variables. β_0, β_1, β_2 and σ^2 are all unknown *a priori*.

1. Find explicit algebraic expressions for the least squares estimators, $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$. [20]

2. In a particular experiment, the values of these estimators are $\hat{\beta}_0 = 4.6$, $\hat{\beta}_1 = 2.0$, $\hat{\beta}_2 = 1.5$. The error sum of squares, $\sum e_i^2$ (where $e_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}$), is 206.

(a) State an unbiased estimate of σ^2. What are the degrees of freedom? [4]

(b) What are the standard errors of $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$? [7]

(c) Give a 95% confidence interval for β_2. [7]

(d) Test the hypothesis $H_0: \beta_1 = 0$ against the alternative $H_1: \beta_1 > 0$. Use the significance level $\alpha = 0.1$. Should you accept or reject the hypothesis? [7]

(e) A new observation is taken for which $x_{i1} = 3.5$, $x_{i2} = 1$. State a 90% prediction interval for this observation. [15]

3. We wish to test $H_0: \beta_1 = \beta_2 = 0$, against the alternative that β_1 and β_2 are not both 0.

(a) Show how to construct an F test for this problem. In particular, give an explicit expression for the quantity $SSE_0 - SSE_1$. [15]

(b) Suppose the test is applied for certain specific values of β_1 and β_2 which are not both 0. In that case, the power of the F test is given by a certain noncentral F distribution. What is the noncentrality parameter δ of that distribution? (*Note:* For this part of the question, you have to assume that σ^2 is known. So assume that, i.e. the final answer will be some algebraic expression that includes σ^2 as well as other quantities.) [15]

(c) Now suppose $\beta_1 = 1$, $\beta_2 = 2$, $\sigma^2 = 7$. Also assume the test of $\beta_1 = \beta_2 = 0$ is carried out at size $\alpha = .01$. What is (the numerical value of) the power of the test in this instance? [10]

SOLUTIONS

1. First calculate

$$X=\begin{pmatrix}1&-5&-1\\1&-4&0\\1&-3&1\\1&-2&-1\\1&-1&0\\1&0&1\\1&1&-1\\1&2&0\\1&3&1\\1&4&-1\\1&5&1\end{pmatrix},\quad X^TX=\begin{pmatrix}11&0&0\\0&110&7\\0&7&8\end{pmatrix},\quad (X^TX)^{-1}=\begin{pmatrix}\frac{1}{11}&0&0\\0&\frac{8}{831}&-\frac{7}{831}\\0&-\frac{7}{831}&\frac{110}{831}\end{pmatrix}.$$

Therefore,

$$\begin{aligned}\hat\beta_0&=\frac{\sum y_i}{11},\\ \hat\beta_1&=\frac{8\sum x_{i1}y_i-7\sum x_{i2}y_i}{831},\\ \hat\beta_2&=\frac{-7\sum x_{i1}y_i+110\sum x_{i2}y_i}{831}.\end{aligned}$$

2. (a) $s^2=\frac{206}{8}=25.75$ with 8 degrees of freedom.

(b) The three standard errors are

$$\sqrt{\frac{25.75}{11}}=1.530,\qquad \sqrt{\frac{8\times 25.75}{831}}=0.498,\qquad \sqrt{\frac{110\times 25.75}{831}}=1.846.$$

(c) The two-sided .05-level rejection point for the t_8 distribution is 2.306, so the confidence interval is $1.5\pm 2.306\times 1.846=1.5\pm 4.257=(-2.757,5.757)$.

(d) A one-sided test at significance level 0.1 requires the same rejection point as a two-sided test at significance level 0.2, so the relevant t_8 rejection point is 1.397. Thus, the rejection point for this test is $1.397\times 0.498=0.696$, which is exceeded by the point estimate 2.0, so the null hypothesis is rejected.

(e) $y^*=\beta_0+\beta_1x_1^*+\beta_2x_2^*+\epsilon^*$, $\hat y^*=\hat\beta_0+\hat\beta_1x_1^*+\hat\beta_2x_2^*$, where $x_1^*=3.5,\ x_2^*=1$, so

$$y^*-\hat y^*=(\beta_0-\hat\beta_0)+(\beta_1-\hat\beta_1)x_1^*+(\beta_2-\hat\beta_2)x_2^*+\epsilon^*,$$

and the variance of that, taking into account the correlation between $\hat{\beta}_1$ and $\hat{\beta}_2$, is

$$\sigma^2\left(\frac{1}{11}+\frac{8(x_1^*)^2}{831}+\frac{110(x_2^*)^2}{831}-\frac{14x_1^*x_2^*}{831}+1\right)=1.282\sigma^2.$$

Therefore, the prediction standard error is $\sqrt{1.282\times 25.75}=5.75$, the point prediction is $\hat{y}^*=4.6+2\times 3.5+1.5=13.1$, and a 90% (two-sided) prediction interval is $13.1\pm(1.860\times 5.75)=13.1\pm 10.7=(2.4,23.8)$.

3. (a) The basic F test is

$$\frac{SSE_0-SSE_1}{2}\cdot\frac{8}{SSE_1}\sim F_{2,8} \text{ under } H_0.$$

Here,

$$\begin{aligned}
SSE_1 &= \sum(y_i-\hat{\beta}_0-\hat{\beta}_1x_{i1}-\hat{\beta}_2x_{i2})^2,\\
SSE_0 &= \sum(y_i-\bar{y})^2,\\
SSE_0-SSE_1 &= \hat{\beta}_1^2\sum x_{i1}^2+\hat{\beta}_2^2\sum x_{i2}^2+\hat{\beta}_1\hat{\beta}_2\sum x_{i1}x_{i2}\\
&= 110\hat{\beta}_1^2+8\hat{\beta}_2^2+14\hat{\beta}_1\hat{\beta}_2.
\end{aligned}$$

[*Note.* If we write $e_i=y_i-\hat{\beta}_0-\hat{\beta}_1x_{i1}-\hat{\beta}_2x_{i2}$, $e=(\,e_1\quad e_2\quad\ldots\quad e_{11}\,)^T$, then it follows from the basic theory that $X^Te=X^T(Y-X\hat{\beta})=0$, so $\sum e_ix_{i1}=\sum e_ix_{i2}=0$. The expansion for SSE_0-SSE_1 follows from $SSE_0=\sum\{e_i+(\hat{\beta}_1x_{i1}+\hat{\beta}_2x_{i2})\}^2=\sum e_i^2+\sum(\hat{\beta}_1x_{i1}+\hat{\beta}_2x_{i2})^2$ since the cross-product is 0. It wasn't necessary to give either this argument or the precise formula to gain credit for the question, provided some reasonably explicit expression was given.]

(b) Under the substitution rule, or directly from the $(h-h')^T\{C(X^TX)^{-1}C^T\}^{-1}(h-h')$ formula,

$$\delta^2=\frac{110\beta_1^2+8\beta_2^2+14\beta_1\beta_2}{\sigma^2}.$$

[For the second method: since $C=\begin{pmatrix}0&1&0\\0&0&1\end{pmatrix}$,

$$C(X^TX)^{-1}C=\begin{pmatrix}\frac{8}{831}&-\frac{7}{831}\\[1ex]-\frac{7}{831}&\frac{110}{831}\end{pmatrix},\quad \{C(X^TX)^{-1}C\}^{-1}=\begin{pmatrix}110&7\\7&8\end{pmatrix},$$

from which the stated result quickly follows.]

(c) With the given numbers, $\delta=\sqrt{\frac{170}{7}}=4.93$, $\phi=\frac{\delta}{\sqrt{3}}=2.85$, and from the charts of the noncentral F distribution, the power of the test is about .76.