STOR 664: MIDTERM EXAM
OCTOBER 19 2009

This is an open book exam. Course text, personal notes and calculator are allowed. You have 75 minutes to complete the exam. Answers should preferably be written in blue books.

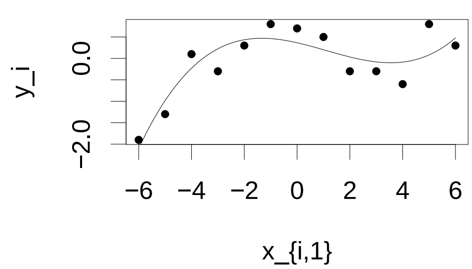
SHOW ALL WORKING: You are not expected to provide lengthy explanations or to reproduce standard results that are in the text, but you should show all working in enough detail to make clear how your answers were obtained.

The University Honor Code is in effect during this exam and you should sign the "pledge" at the front of your exam book. If you have questions, please ask the instructor.

1. Consider the following table of x and y values:

i	$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	y_i	i	$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	y_i	i	$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	y_i
1	–6	22	–216	–1.9	6	–1	–13	–1	0.8	11	4	2	64	–0.6
2	–5	11	–125	–1.3	7	0	–14	0	0.7	12	5	11	125	0.8
3	–4	2	–64	0.1	8	1	–13	1	0.5	13	6	22	216	0.3
4	–3	–5	–27	–0.3	9	2	–10	8	–0.3					
5	–2	–10	–8	0.3	10	3	–5	27	–0.3					

Note that $x_{i,2} = x_{i,1}^2 - 14$, $x_{i,3} = x_{i,1}^3$. The centering ensures that $\sum_i x_{i,1} = \sum_i x_{i,2} = \sum_i x_{i,3} = 0$. We also have: $\sum_i y_i = -1.2$, $\sum_i y_i x_{i,1} = 19.4$, $\sum_i y_i x_{i,2} = -65.4$, $\sum_i y_i x_{i,3} = 687.8$, $\sum_i x_{i,1}^2 = 182$, $\sum_i x_{i,2}^2 = 2002$, $\sum_i x_{i,3}^2 = 134342$, $\sum_i x_{i,1}x_{i,2} = 0$, $\sum_i x_{i,1}x_{i,3} = 4550$, $\sum_i x_{i,2}x_{i,3} = 0$. The problem is the cubic regression $y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \epsilon_i$, as illustrated by the inset figure and fitted curve.



(a) Describe how the problem can be written in the form $Y = X\beta + \epsilon$, and write down explicitly the matrices $X^T X$ and $(X^T X)^{-1}$, as well as $X^T Y$. (Note: You do not require any special computing for $(X^T X)^{-1}$. This is intended to be a simple hand computation!)

(b) Write down the estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ and their standard errors. You may assume that the estimated value of the standard deviation of $\epsilon_1, ..., \epsilon_{13}$ is $s = 0.451$.

(c) A conjecture has been made that in the case $i = 7$ (for which $x_{i,1} = x_{i,3} = 0$, $x_{i,2} = -14$) the expected value of y_i is 0. Formulate this statement as a hypothesis about the parameters β_0, β_1, β_2, β_3, and test that hypothesis.

Note: A table of percentage points of the t distribution is reproduced at the end of this exam (other side of page).

TURN OVER....

2. Consider the following one-way ANOVA experiment with four groups and two observations per group: $Y = X\beta + \epsilon$ where

$$Y = \begin{pmatrix} y_{1,1} \\ y_{1,2} \\ y_{2,1} \\ y_{2,2} \\ y_{3,1} \\ y_{3,2} \\ y_{4,1} \\ y_{4,2} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_{1,1} \\ \epsilon_{1,2} \\ \epsilon_{2,1} \\ \epsilon_{2,2} \\ \epsilon_{3,1} \\ \epsilon_{3,2} \\ \epsilon_{4,1} \\ \epsilon_{4,2} \end{pmatrix},$$

with $\epsilon_{1,1}, ..., \epsilon_{4,2}$ independent $N[0, \sigma^2]$. Also, you may assume without proof that the unrestricted OLS estimates of β are $\hat{\beta}_i = \bar{y}_{i\cdot} = \frac{1}{2}(y_{i,1} + y_{i,2})$ for $i = 1, 2, 3, 4$.

(a) Consider the hypothesis

$$H_0: \ \beta_1 + \beta_2 \ = \ 5, \ \beta_1 + 2\beta_3 \ = \ 9,$$

against the alternative H_1 that is the negation of H_0. If we write $\hat{\beta}_i^{(0)}, \ i = 1, 2, 3, 4$ to be the OLS estimates of $\beta_i, \ i = 1, 2, 3, 4$ under H_0, show that $\hat{\beta}_i^{(0)}, \ i = 1, 2, 3, 4$ are derived from $\hat{\beta}_i, \ i = 1, 2, 3, 4$, through the formulas

$$\begin{aligned} \hat{\beta}_1^{(0)} &= \frac{1}{9}(4\hat{\beta}_1 - 4\hat{\beta}_2 - 2\hat{\beta}_3 + 29), \\ \hat{\beta}_2^{(0)} &= \frac{1}{9}(-4\hat{\beta}_1 + 4\hat{\beta}_2 + 2\hat{\beta}_3 + 16), \\ \hat{\beta}_3^{(0)} &= \frac{1}{9}(-2\hat{\beta}_1 + 2\hat{\beta}_2 + \hat{\beta}_3 + 26), \\ \hat{\beta}_4^{(0)} &= \hat{\beta}_4. \end{aligned}$$

(b) Suppose we want to calculate the power of the F test when the true values are $\beta_1 + \beta_2 = 9, \ \beta_1 + 2\beta_3 = 5$, and $\sigma^2 = 8$. The answer will be given by the Pearson-Hartley formula for specific values of $\phi, \ \alpha, \ \nu_1$ and ν_2. Find the numerical values for $\phi, \ \nu_1$ and ν_2.

For Question 1: Table of percentage points of t distribution:

	One-sided tail probability								One-sided tail probability						
df	0.2	0.1	0.05	0.025	0.01	0.005	0.001	df	0.2	0.1	0.05	0.025	0.01	0.005	0.001
1	1.376	3.078	6.314	12.706	31.821	63.657	318.309	8	0.889	1.397	1.860	2.306	2.896	3.355	4.501
2	1.061	1.886	2.920	4.303	6.965	9.925	22.327	9	0.883	1.383	1.833	2.262	2.821	3.250	4.297
3	0.978	1.638	2.353	3.182	4.541	5.841	10.215	10	0.879	1.372	1.812	2.228	2.764	3.169	4.144
4	0.941	1.533	2.132	2.776	3.747	4.604	7.173	11	0.876	1.363	1.796	2.201	2.718	3.106	4.025
5	0.920	1.476	2.015	2.571	3.365	4.032	5.893	12	0.873	1.356	1.782	2.179	2.681	3.055	3.930
6	0.906	1.440	1.943	2.447	3.143	3.707	5.208	13	0.870	1.350	1.771	2.160	2.650	3.012	3.852
7	0.896	1.415	1.895	2.365	2.998	3.499	4.785	∞	0.842	1.282	1.645	1.960	2.326	2.576	3.090

SKETCH SOLUTIONS

1. (a)

$$(X^TX) = \begin{pmatrix} 13 & 0 & 0 & 0 \\ 0 & 182 & 0 & 4550 \\ 0 & 0 & 2002 & 0 \\ 0 & 4550 & 0 & 134342 \end{pmatrix}, \quad X^TY = \begin{pmatrix} -1.2 \\ 19.4 \\ -65.4 \\ 687.8 \end{pmatrix},$$

$$(X^TX)^{-1} = \begin{pmatrix} \frac{1}{13} & 0 & 0 & 0 \\ 0 & \frac{134342}{\Delta} & 0 & -\frac{4550}{\Delta} \\ 0 & 0 & \frac{1}{2002} & 0 \\ 0 & -\frac{4550}{\Delta} & 0 & \frac{182}{\Delta} \end{pmatrix} = \begin{pmatrix} 0.07692 & 0 & 0 & 0 \\ 0 & 0.03585 & 0 & -0.001214 \\ 0 & 0 & 0.0004995 & 0 \\ 0 & -0.001214 & 0 & 0.00004856 \end{pmatrix},$$

where $\Delta = 182 \times 134342 - 4550^2 = 3747744$. **[20 points]**

(b) $\hat{\beta}_0 = -0.0923,\ \hat{\beta}_1 = -0.1396,\ \hat{\beta}_2 = -0.0326,\ \hat{\beta}_3 = 0.00985$; standard errors 0.125, 0.085, 0.010, 0.0031. **[15 points]**

(c) $H_0:\ \beta_0 - 14\beta_2 = 0$. But $\hat{\beta}_0 - 14\hat{\beta}_2 = 0.365$, standard error $s\sqrt{0.07692 + 14^2 \times 0.0004995} = 0.189$. The t ratio is $\frac{.365}{.189} = 1.93$. From the t table with df=9, the one-sided tail probability lies between .025 and .05 (the exact one-sided p-value is about .041). So with a 2-sided test at $\alpha = .05$, we would accept the null hypothesis. Conclusion: there's not enough evidence to reject the hypothesis (though it's close). (*Note:* The theory underlying this question is Section 3.3.3. You don't need to invoke the general theory of F tests.) **[15 points]**

2. (a) Write the hypothesis in the form $C\beta = h$ where $C = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 \end{pmatrix}$, $h = \begin{pmatrix} 5 \\ 9 \end{pmatrix}$. With slight adaptation of equation (3.33) of the text, we write

$$\hat{\beta}^{(0)} = \hat{\beta} - (X^TX)^{-1}C^T \left\{ C(X^TX)^{-1}C^T \right\}^{-1} \left\{ C\hat{\beta} - h \right\},$$

where for example $C^T(X^TX)^{-1}C = \frac{1}{2}\begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}$, $\{C^T(X^TX)^{-1}C\}^{-1} = \frac{2}{9}\begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix}$. After some manipulation,

$$\begin{pmatrix} \hat{\beta}_1^{(0)} \\ \hat{\beta}_2^{(0)} \\ \hat{\beta}_3^{(0)} \\ \hat{\beta}_4^{(0)} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} - \frac{1}{9} \begin{pmatrix} 5\hat{\beta}_1 + 4\hat{\beta}_2 + 2\hat{\beta}_3 - 29 \\ 4\hat{\beta}_1 + 5\hat{\beta}_2 - 2\hat{\beta}_3 - 16 \\ 2\hat{\beta}_1 - 2\hat{\beta}_2 + 8\hat{\beta}_3 - 26 \\ 0 \end{pmatrix}$$

which reduces to the form given. **[30 points]**

(b) Since $n = 8,\ p = 4,\ q = 2$, the degrees of freedom of the F test are $\nu_1 = 2,\ \nu_2 = 4$.

Applying formula (3.42) of the text,

$$\begin{aligned}
\sigma^2\delta^2 &= \frac{2}{9}\begin{pmatrix} -4 & 4 \end{pmatrix}\begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} -4 \\ 4 \end{pmatrix} \\
&= 32
\end{aligned}$$

so with $\sigma^2 = 8$, we have $\delta = 2$ and so $\phi = \frac{\delta}{\sqrt{1+\nu_1}} = \frac{2}{\sqrt{3}} = 1.155$. **[20 points]**

Alternate solution to (a). Under H_0, $E(Y) = \begin{pmatrix} \beta_1 & \beta_1 & 5-\beta_1 & 5-\beta_1 & \frac{9-\beta_1}{2} & \frac{9-\beta_1}{2} & \beta_4 & \beta_4 \end{pmatrix}^T$. Define $\tilde{Y} = Y - \begin{pmatrix} 0 & 0 & 5 & 5 & \frac{9}{2} & \frac{9}{2} & 0 & 0 \end{pmatrix}^T$, then $E(\tilde{Y}) = \tilde{X}\begin{pmatrix} \beta_1 \\ \beta_4 \end{pmatrix}^T$ where $\tilde{X}^T = \begin{pmatrix} 1 & 1 & -1 & -1 & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$. We have $\tilde{X}^T\tilde{X} = \begin{pmatrix} \frac{9}{2} & 0 \\ 0 & 2 \end{pmatrix}$, $(\tilde{X}^T\tilde{X})^{-1} = \begin{pmatrix} \frac{2}{9} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$, from which we easily calculate $\hat{\beta}_i^{(0)},\ i = 1, 4$ as given. The remaining cases then follow from $\hat{\beta}_2^{(0)} = 5 - \hat{\beta}_1^{(0)},\ \hat{\beta}_3^{(0)} = \frac{1}{2}(9 - \hat{\beta}_1^{(0)})$.

Alternate solution to (b). Choose some set of β values consistent with H_1, e.g. $\beta_1 = 5,\ \beta_2 = 4,\ \beta_3 = \beta_4 = 0$ (by the general theory, the exact choice does not matter). Now assume these are the true data values, hence also the values of $\hat{\beta}_i,\ i = 1, 2, 3, 4$. Using the answer to (a), calculate $\hat{\beta}_1^{(0)} = \frac{11}{3}$, $\hat{\beta}_2^{(0)} = \frac{4}{3}$, $\hat{\beta}_3^{(0)} = \frac{8}{3}$, $\hat{\beta}_4^{(0)} = 0$. The RSS under H_0 is $2\sum_{i=1}^{4}(\hat{\beta}_i - \hat{\beta}_i^{(0)})^2 = 32$. By the substitution rule, this is also $\sigma^2\delta^2$. The rest is as above.