STAT 3008 Applied Regression Analysis
Midterm
2:45-4:15. Monday, 21 Oct 2013

Name: _____   **Major:** _____

1. (20) Fill in the missing values in the following tables of regression output from a data set of size 100.

ANOVA Table				
Source	Sum of Squares	d.f.	Mean Square	F-statistics
Regression				
Residuals				
Total				

Coefficient Table				
Variable	Coefficient	s.e.	t-statistics	p-value
Constant	0.5854			0.2188
X		0.4927		
$n=$	$R^2=0.03294$	$\hat{\sigma} = 4.714$		

p.s.

i) It is found in R that $2 * (1 - pt(1.238, 98)) = 0.2188$.

ii) $\hat{\beta}_1$ is positive.

You may list your steps in the below area.

2. (50) There are 8 students in the class STAT 3009. For each student, the GPA (out of 4), IQ, and averaged studying-hour per day are recorded.

Student	1	2	3	4	5	6	7	8
GPA (Y)	1.70	2.44	3.585	1.52	3.94	2.22	2.88	1.84
IQ (X_1)	150	180	120	140	130	90	110	130
Hour (X_2)	1	2	4.5	0	5	3	4	2

It is found that if

$$X = \begin{pmatrix} 1 & 150 & 1^2 \\ 1 & 180 & 2^2 \\ \vdots & \vdots & \vdots \\ 1 & 130 & 2^2 \end{pmatrix}, \text{then } (X'X)^{-1} = \begin{pmatrix} 5.237 & -0.03454 & -0.05841 \\ -0.03454 & 0.0002408 & 0.0002958 \\ -0.05841 & 0.0002958 & 0.001976 \end{pmatrix}$$

Let $\hat{e}_{Z_1|Z_2}$ be the residual vector of computed from regressing Z_1 against Z_2. Denote $X_3 = X_2^2$, $E_i = \hat{e}_{Y|X_i}$ and $E_{ij} = \hat{e}_{X_i|X_j}$ for $i \neq j$ and $i, j = \{1, 2, 3\}$. For any vectors $U = (u_1, \ldots, u_n)$ and $V = (v_1, \ldots, v_n)$, define $SUV = \sum_{i=1}^{n} u_i v_i$. We have

SYY	SX_1X_1	SX_2X_2	SX_3X_3	SE_1E_1	SE_2E_2	SE_3E_3
56.14	142900	79.25	1405.1	5.254	0.678	3.418
$SE_{12}E_{12}$	$SE_{13}E_{13}$	$SE_{21}E_{21}$	$SE_{31}E_{31}$	$SE_{23}E_{23}$	$SE_{32}E_{32}$	SE_1E_{13}
4013.1	4152.0	16.93	505.99	1.56	44.63	62.88
SYX_1	SYX_2	SYX_3	SX_1X_2	SX_1X_3	SX_2X_3	SE_3E_{31}
2605.2	64.27	255.98	2670	9640	324.1	4.99
SE_1E_{21}	SE_1E_{31}	SE_2E_{12}	SE_2E_{32}	SE_3E_{13}	SE_3E_{23}	SE_1E_{12}
9.106	51.19	35.86	3.879	33.33	0.038	64.42

i) (5 marks) Suppose that you want to draw a graph to represent the relationship between Y and X_1 with the effect of X_2 removed. What are the x and y axis of that graph? What's the name of this graph?

ii) (10 marks) Suppose that the residuals of the regressions
$Y = \alpha_0 + \alpha_1 X1 + e$ and $X2 = \alpha_0 + \alpha_1 X1 + e$ are

$$(-0.682, 0.271, 0.989, -0.933, 1.42, -0.589, 0.213, -0.685) \text{ and}$$
$$(-1.128, 0.768, 1.48, -2.42, 2.28, -0.919, 0.678, -0.725),$$

respectively. Sketch a scatterplot for the relationship between Y and $X2$. Is a linear regression appropriate? What should we do?

iii) (10 marks) For the regression $\hat{e}_{Y|X1} = \eta_0 + \eta_1 \hat{e}_{X_2^2|X_1} + e$, find the estimate of η_0 and η_1 and and find their standard errors.

iv) (10 marks) Give all the estimates of the regression $Y = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2^2 + e$. Let $\hat{\gamma}_2$ be the estimate of γ_2. Find the standard error of $\hat{\gamma}_2$.

v) (5 marks) Compare between the standards errors of $\hat{\gamma}_2$ and $\hat{\eta}_1$. Are they the same?

vi) (10 marks) Bosco's IQ is 200 and he does no revision at all. Using the model in iv), find a 90% prediction interval for his GPA.

3. (10 marks). For the regression $Y = X\beta + e$, $e \sim N(0, \sigma^2)$, let $\hat{y}_i = X_i\hat{\beta}$ be the fitted value of the i-th observation, $i = 1, \ldots, n$. Let X be a $n \times p$ matrix.

(a) Find $E(\sum_{i=1}^{n} \hat{y}_i^2)$ in terms of X, β, p and σ^2.

(b) Find $E(\sum_{i=1}^{n} (\hat{y}_i - \bar{y})(y_i - \hat{y}_i))$.

4. (10 marks). Consider the regression $Y = X\beta + e$, where X is $n \times (p+3)$ with the first column being constants 1. Express the F-statistics of the overall analysis of variance in terms of the coefficient of the determination. (Express A in terms of B means writing $A = f(B)$ for some function f.)

5. (10 marks). In searching for the estimates of the regression coefficient β, we differentiate the RSS and solve for system of equations. Will there be more than one solutions? Will the solution be the maximizer of the RSS instead of minimizer?