UNIVERSITY OF TORONTO SCARBOROUGH
Department of Computer and Mathematical Sciences
Midterm Test, October 2013

STAC67H3 Regression Analysis
Duration: One hour and fifty minutes

Last Name:________________________ First Name: ________________________

Student number: ________________________

Aids allowed:

- The textbook: Applied Regression Analysis by Ketner et al published by McGraw Hill
- Class notes
- A calculator (No phone calculators are allowed)

No other aids are allowed. For example you are not allowed to have any other textbook or past exams.

All your work must be presented clearly in order to get credit. Answer alone (even though correct) will only qualify for **ZERO** credit. Please show your work in the space provided; you may use the back of the pages, if necessary, but you MUST remain organized. Show your work and answer in the space provided, in ink. Pencil may be used, but then any re-grading will NOT be allowed.

There are 14 pages including this page. Please check to see you have all the pages.

Good Luck!

Question:	1	2	3	4	Total
Points:	10	20	27	8	65
Score:					

1. For the simple linear regression model with independent Normal errors,

   (a) (7 points) Show that $\bar{Y}$ and the least square estimator of the slope B_1 are uncorrelated.

   > **Solution:** $Cov(\bar{Y}, B_1) = Cov(\sum \frac{1}{n} Y_i, \sum k_i Y_i) = \sum \frac{1}{n} k_i Cov(Y_i, Y_i)$ (since Y_i's are independent) $= \sigma^2 \frac{1}{n} \sum k_i = 0$ since $\sum k_i = 0$ ■

   (b) (3 points) Are $\bar{Y}$ and B_1 independent. Give reasons for your answer.

   > **Solution:** Both $\bar{Y}$ and B_1 are linear combinations of independent Normal random variables. Thus $\bar{Y}$ and B_1 are jointly Normal. With joint Normality, uncorrelated implies independent.

2. A rocket motor is manufactured by bonding together two types of propellants, an igniter and a sustainer. The shear strength (psi) of the bond y is thought to be a linear function of the age (weeks) of the propellant x when the motor is cast. The R code below gives the summary statistics from 10 observations on these two variables.

```
> rocket=read.table("C:/Users/Mahinda/Desktop/rocket.txt", header=1)
> length(rocket$x)
[1] 10
> mean(rocket$x)
[1] 13.325
> sd(rocket$x)
[1] 7.65402
> mean(rocket$y)
[1] 2112.525
> sd(rocket$y)
[1] 301.1279
> cor(rocket$x, rocket$y)
[1] -0.9209366
```

(a) (10 points) Find the least squares estimates of the slope and intercept in the simple linear regression model.

Solution: $b_1 = \frac{SS_{XY}}{SS_{XX}} = r\frac{s_Y}{s_X} = -0.9209366 \times \frac{301.1279}{7.65402} = -36.23190224.$
$b_0 = \bar{y} - b_1\bar{x} = 2112.525 - (-36.23190224) \times 13.325 = 2595.315097$

Here is a MINITAB output (just to compare the above answers):

```
Regression Analysis: y versus x

The regression equation is
y = 2595 - 36.2 x


Predictor     Coef  SE Coef      T      P
Constant   2595.32    82.26  31.55  0.000
x          -36.232    5.421  -6.68  0.000


S = 124.472   R-Sq = 84.8%   R-Sq(adj) = 82.9%


Analysis of Variance

Source           DF       SS       MS       F       P
```

Question 2 continues on the next page...

Regression	1	692156	692156	44.67	0.000
Residual Error	8	123946	15493		
Total	9	816102			

(b) (10 points) Find an estimate of σ^2.

Solution: $SSR = b_1^2 SS_{XX} = b_1^2(n-1)s_X^2 = (-36.23190224)^2 \times (10-1) \times 7.65402^2 = 692155.966$,
$SST = (n-1)s_Y^2 = (10-1) \times 301.1279^2 = 816102.1094$,
$SSE = SST - SSR = 816102.1094 - 692155.966 = 123946.1434$,
$MSE = \frac{SSE}{n-2} = \frac{123946.1434}{10-2} = 15493.26793$

Question 2 continues on the next page…

You may continue your to answer to question 2 on this page

3. A story by James R. Hagerty entitled With Buyers Sidelined, Home Prices Slide published in the Thursday October 25, 2007 edition of the Wall Street Journal contained data on so-called fundamental housing indicators in major real estate markets across the US. The author argues that the prices are generally falling and overdue loan payments are piling up. This article presented data on

   y = Percentage change in average price from July 2006 to July 2007 (based on the S&P/Case-Shiller national housing index); and
   x = Percentage of mortgage loans 30 days or more overdue in latest quarter (based on data from Equifax and Moodys).
   The R output given below is was obtained from this data set and is intended to fit the Normal error SLR model $Y = \beta_0 + \beta_1 x + \varepsilon$. Assume that the data satisfied all the necessary assumptions.

```
indicators=read.table("C:/Users/Mahinda/Desktop/indicators.txt", header=1)
> fit = lm(y ~x , data = indicators)
> summary(fit)

Call:
lm(formula = y ~ x, data = indicators)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.5145     3.3240   1.358   0.1933
x            -2.2485     0.9033  -2.489   0.0242 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 3.954 on 16 degrees of freedom
Multiple R-squared:  0.2792,Adjusted R-squared:  0.2341
F-statistic: 6.196 on 1 and 16 DF,  p-value: 0.02419

> mean(indicators$x)
[1] 3.532222
```

   (a) (5 points) Calculate a 95% confidence interval for β_1.

   Solution:

```
> confint(fit)
                 2.5 %     97.5 %
(Intercept) -2.532112 11.5611000
x           -4.163454 -0.3335853
```

   (b) (3 points) Test the null hypothesis $H_0 : \beta_1 = -1$ against the alternative $H_1 : \beta_1 \neq -1$. Use $\alpha = 0.05$.

Question 3 continues on the next page. . .

Solution: -1 is in the 95 % confidence interval and so do not reject the null hypothesis $H_0 : \beta_1 = -1$ at $\alpha = 0.05$.

(c) (5 points) Calculate a 95% confidence interval for the mean of Y at $x = 2.5$.

Solution: $\hat{y} = 4.5145 - 2.2485 \times 2.5 = -1.10675$
$s_{\hat{y}} = s\sqrt{\frac{1}{n} + \frac{(x_h - \bar{x})^2}{SS_{XX}}} = \sqrt{\frac{s^2}{n} + \frac{s^2}{SS_{XX}} \times (x_h - \bar{x})^2}$
$= \sqrt{\frac{3.945^2}{n} + s_{b_1}^2 \times (2.5 - 3.532222)^2} = \sqrt{\frac{3.954^2}{18} + (0.9033)^2 \times (2.5 - 3.532222)^2}$.
Though not necessary you can also find SS_{XX} as follows: $s_{b_1} = \frac{s}{\sqrt{SS_{XX}}}$ and so $SS_{XX} = \frac{s^2}{s_{b_1}^2} = \frac{3.954^2}{0.9033^2} = 19.16060904$ Here is an R output just to check calculations:

```
> indicators=read.table("C:/Users/Mahinda/Desktop/indicators.txt",
header=1)
> sd(indicators$x)
[1] 1.061633
> sx = sd(indicators$x)
> n = 18
> ssxx = (n-1)*sx^2
> ssxx
[1] 19.16011
> fit = lm(y ~x , data = indicators)
> confint(fit, level=0.95) # CIs for model parameters
                2.5 %     97.5 %
(Intercept) -2.532112 11.5611000
x           -4.163454 -0.3335853
> x0 <- data.frame(x=2.5)
> predict(fit, x0)
        1
-1.106806
> #CI for the mean of Y at x=x0
> predict(fit, x0, interval="confidence", level=0.95)
        fit       lwr      upr
1 -1.106806 -3.901517 1.687905
```

(d) (2 points) What proportion of the variability in Y is explained by this linear regression model with x?

Solution: The proportion of the variability in Y is explained by the linear regression model with $x = R^2 = 0.2792$.

(e) (12 points) Calculate the ANOVA table. Just give the values of a, b, c, d, f, and g

in the table below.

Source of variation	Degrees of freedom	Sum of squares	Mean squares	F
Regression	a	b	c	d
Error	e	f	g	

Solution:

```
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value  Pr(>F)
x          1   96.87  96.870  6.1961 0.02419 *
Residuals 16  250.15  15.634
```

Question 3 continues on the next page. . .

You may continue your to answer to question 3 on this page

Question 3 continues on the next page...

You may continue your to answer to question 3 on this page

4. (8 points) (Ott, L et.al) A manufacturer of laundry detergent was interested in testing a new product prior to market release. One area of concern was the relationship between the height(y) of the detergent suds in a washing machine as a function of the amount(x) of detergent added in the wash cycle. For a standard size washing machine tub filled to the full level, the manufacturer made random assignments of amounts of detergent and tested them on the washing machine. The data and the ANOVA table for the linear regression of y on x is given below:

x	y
6	28.1
6	27.6
7	32.3
7	33.2
8	34.8
8	35.0
9	38.2
9	39.4
10	43.5
10	46.8

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	330.48	330.48	169.21	0.000
Residual Error	8	15.62	1.95		
Total	9	346.11			

Conduct a test for lack of fit of the linear regression model. Test at $\alpha = 0.05$

Solution: Note: This is question 11.53 9 582 Ott.

```
The regression equation is
y = 3.37 + 4.06 x
```

Predictor	Coef	SE Coef	T	P
Constant	3.370	2.539	1.33	0.221
x	4.0650	0.3125	13.01	0.000

```
S = 1.39752   R-Sq = 95.5%   R-Sq(adj) = 94.9%
```

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	330.48	330.48	169.21	0.000
Residual Error	8	15.62	1.95		
Lack of Fit	3	8.91	2.97	2.21	0.205
Pure Error	5	6.71	1.34		
Total	9	346.11			

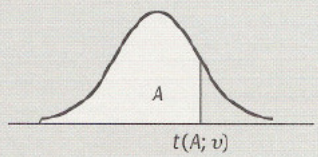
Question 4 continues on the next page...

You may continue your to answer to question 4 on this page

END OF TEST

TABLE B.2
Percentiles of the t Distribution.

Entry is $t(A;\nu)$ where $P\{t(\nu) \leq t(A;\nu)\} = A$



	A						
ν	.60	.70	.80	.85	.90	.95	.975
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.537	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960