STA 302 / 1001 H - Fall 2003
Test 2
November 10, 2003

LAST NAME:________________________FIRST NAME:________________________

STUDENT NUMBER:____

ENROLLED IN: (circle one) STA 302 STA 1001

INSTRUCTIONS:
- Time: 50 minutes
- Aids allowed: calculator.
- A table of values from the t distribution is on the last page.
- A table of formulae is on the second to last page.
- Total points: 30

1 (a) (b)	2 (a)	2 (b)	3 (a)	3 (b) (c)	3(d)

1. (a) (4 points) State the simple linear regression model in matrix terms, defining all matrices and vectors. Include the Gauss-Markov assumptions.

   (b) (4 points) Use matrix properties to prove $\mathbf{e}'\hat{\mathbf{Y}} = \mathbf{0}$ where $\mathbf{e}$ is the vector of residuals and $\hat{\mathbf{Y}}$ is the vector of predicted values. (Do not work with the individual elements of these matrices.)

2. The times to failure (Y) for 20 light bulbs were measured at 20 temperatures (X). A simple linear regression was carried out to explore how the failure time is related to the temperature.

   (a) (5 points) Here are some calculated values from the 20 data points.

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 0.105 & -0.0136 \\ -0.0136 & 0.00335 \end{pmatrix}$$

$$\mathbf{X}'\mathbf{Y} = \begin{pmatrix} 367.9 \\ 965.6 \end{pmatrix}$$

$$\mathbf{e}'\mathbf{e} = 789.7$$

   What are the estimated slope and intercept of the regression line and their estimated standard errors?

(b) (2 points) A scatterplot and a plot of the residuals versus the predicted values for the light bulb problem are given below. What additional information does this provide?

3. An experiment was performed to determine which temperature in the manufacturing process results in the strongest product. In analysing the data, it was determined that a straight line model was an appropriate fit for the regression with square root of strength (`sqrtstrn`) as the dependent variable and temperature (`temp`) as the independent variable. Six temperatures (80^o to 180^o in increments of 20^o) were tested five times, resulting in 30 data points. Some SAS output is given below.

```
The REG Procedure
Descriptive Statistics
```

Variable	Sum	Mean	Uncorrected SS	Variance	Standard Deviation
Intercept	30.00000	1.00000	30.00000	0	0
temp	3900.00000	130.00000	542000	1206.89655	34.74042
sqrtstrn	20660	688.67158	15089253	29696	172.32658

```
The REG Procedure
Dependent Variable: sqrtstrn

Analysis of Variance
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	858611	858611	9295.97	<.0001
Error	28	2586.18683	92.36382		
Corrected Total	29	861197			

```
Parameter Estimates
```

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	44.78775	6.90488	6.49	<.0001
x	1	4.95295	0.05137	96.42	<.0001

(a) (6 points) Construct simultaneous 90% confidence intervals for the expected strength for temperatures of 80^o and 180^o. Use the Bonferroni method.

(b) (2 points) The Bonferroni method is "conservative". Explain what this means in relation to your answer to (a).

(c) (3 points) Before carrying out the regression using the square root of strength, a regression was first carried out with the raw data (strength regressed on temperature). What would be evident in the plot of residuals versus predicted values that would cause the people analysing the data to try the square root transform?

(d) (4 points) The normal probability plot of the residuals for the regression of square root of strength on temperature is given below. What additional information does this give you? Quote one number from the SAS output whose value is relevant to this additional information and explain how it is relevant.

Some formulae:

$$b_1 = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2} \qquad b_0 = \overline{Y} - b_1\overline{X}$$

$$\text{Var}(b_1) = \frac{\sigma^2}{\sum(X_i - \overline{X})^2} \qquad \text{Var}(b_0) = \sigma^2\left(\frac{1}{n} + \frac{\overline{X}^2}{\sum(X_i - \overline{X})^2}\right)$$

$$\text{Cov}(b_0, b_1) = -\frac{\sigma^2\overline{X}}{\sum(X_i - \overline{X})^2} \qquad \text{SSTO} = \sum(Y_i - \overline{Y})^2$$

$$\text{SSE} = \sum(Y_i - \hat{Y}_i)^2 \qquad \text{SSR} = b_1^2\sum(X_i - \overline{X})^2 = \sum(\hat{Y}_i - \overline{Y})^2$$

$$\sigma^2\{\hat{Y}^*\} = \text{Var}(\hat{Y}^*) = \sigma^2\left(\frac{1}{n} + \frac{(X^* - \overline{X})^2}{\sum(X_i - \overline{X})^2}\right)$$

$$\sigma^2\{\text{pred}\} = \text{Var}(Y^* - \hat{Y}^*) = \sigma^2\left(1 + \frac{1}{n} + \frac{(X^* - \overline{X})^2}{\sum(X_i - \overline{X})^2}\right)$$

$\hat{X}_h \pm \frac{t_{n-2,1-\alpha/2}}{|b_1|} *$ appropriate s.e.
(valid approximation if $\frac{t^2 s^2}{b_1^2 S_{XX}}$ is small)

$$\text{Cov}(\mathbf{X}) = E[(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})'] \qquad \text{Cov}(\mathbf{AX}) = \mathbf{A}\text{Cov}(\mathbf{X})\mathbf{A}'$$

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \qquad \text{Cov}(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

$$\hat{\mathbf{Y}} = \mathbf{Xb} = \mathbf{HY} \qquad \mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \qquad \text{SSR} = \mathbf{Y}'(\mathbf{H} - \tfrac{1}{n}\mathbf{J})\mathbf{Y}$$

$$\text{SSE} = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y} \qquad \text{SSTO} = \mathbf{Y}'(\mathbf{I} - \tfrac{1}{n}\mathbf{J})\mathbf{Y}$$