Name:_____

STATISTICS 217, SPRING 2005

FINAL EXAM; May 4, 2005; 100 points

v. 1

Instructions: Write legibly. Closed book. Calculator allowed. Double-sided exam.

Multiple Choice Questions (3 pts each). Circle the SINGLE best answer for each question.

1. The records of 12 students were obtained in a SRS and each student's GPA (response variable y) and ACT score (explanatory variable x) were recorded. The estimated regression equation is

$$\hat{y} = -0.3994 + 0.1364x.$$

The proper interpretation of b_1 is:

A. for every unit increase in GPA, the estimated mean ACT score increases by 0.1364 units.

B. for every unit increase in ACT score, the estimated mean GPA decreases by 0.3994 units.

C. for every unit increase in ACT score, the estimated mean GPA increases by 0.1364 units.

D. for every decrease of 0.3994 units in GPA, the estimated mean ACT score increases by 0.1364 units.

2. A small p-value for the F-test used in simple linear regression indicates

A. $\beta_0 \neq 0$

B. $\beta_1 \neq 0$

C. there is a linear relationship between the predictor variable and the mean response.

D. both **A** and **B**

E. both **B** and **C**

3. Which of the following is **NOT** an assumption in simple linear regression?

A. The ϵ_i error terms are independent of one another.

B. The ϵ_i error terms are normally distributed.

C. The response variable y is normally distributed for each x.

D. The ϵ_i error terms are linearly related to x.

4. In simple linear regression, the distribution of the response variable y at any x value is

A. Normal with mean equal 0 and standard deviation equal 1.

B. Normal with mean equal 0 and standard deviation equal σ.

C. Normal with mean equal $\beta_0 + \beta_1 x$ and standard deviation equal σ.

D. F with p and $n - p - 1$ degrees of freedom.

5. The parameters of the simple linear regression model are:

A. $\beta_0,\ \beta_1,$ and σ

B. $\beta_0,\ \beta_1,$ and ϵ

C. $b_0,\ b_1,$ and e

D. $b_0,\ b_1,$ and s

6. What is the MLR model using 4 predictor variables?

A. $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i$

B. $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4$

C. $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$

D. $\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$

7. The proportion of variation in the response explained by all the predictor variables in a multiple linear regression is given by

   A. R^2
   B. $+\sqrt{R^2}$
   C. $-\sqrt{R^2}$
   D. R

8. A multiple linear regression model using 3 predictor variables results in a p-value= 0.002 for the F-test. What can we conclude?

   A. β_1, β_2, β_3 are *all* different from 0.
   B. x_1, x_2, and x_3 are *all* different from 0.
   C. at least one of β_1, β_2, β_3 differs from 0.
   D. none of β_1, β_2, β_3 are different from 0.

USE the following information to answer the next 4 questions.

The article "Multiple Linear Regression for Lake Ice and Lake Temperature Characteristics" presents data on maximum ice thickness in millimeters (y), average number of days per year of ice cover (x_1), average number of days the bottom temperature is lower than 8 degrees Celsius (x_2), and the average snow depth in millimeters (x_3) for 13 lakes in Minnesota.

```
Regression Analysis: y versus x1, x2, x3
Predictor        Coef     SE Coef          T        P
Constant       -286.2       238.7      -1.20    0.261
x1              2.813       1.177       2.39    0.041
x2              3.696       1.090              0.008
x3            -1.8370      0.8389      -2.19    0.056

Analysis of Variance
Source          DF        SS       MS       F       P
Regression       3     29839     9946    6.82   0.011
Residual Error   9     13131     1459
Total           12     42969
```

9. The R^2 value is

   A. 0.0110
   B. 0.1467
   C. 0.3056
   D. 0.6944

10. Find the estimate of σ^2.

    A. 38.1969
    B. 114.5906
    C. 1459
    D. 13131

11. Predict the response for $x_1 = 150$, $x_2 = 200$, and $x_3 = 85$. These x values are within the range of x values used to calculate the estimated multiple linear regression line.

    A. 674.6550
    B. 718.8050
    C. 1031.0950
    D. 1291.2050

12. The value of the t-statistic for testing $\mathrm{H}_0\colon \beta_2 = 0$ versus $\mathrm{H}_a\colon \beta_2 \neq 0$ given x_1 and x_3 are in the model is

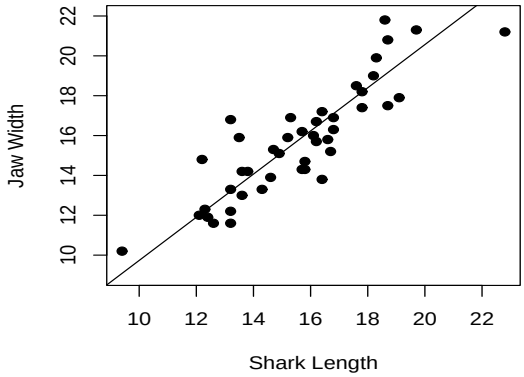
    A. 0.2949
    B. 2.3900
    C. 2.6100
    D. 3.3908

SHOW YOUR WORK: Problems and Discussion

Instructions: SHOW ALL WORK to receive full credit. Perform all calculations to 4 decimal places.

13. **Sharks!** Physical characteristics of sharks are of interest to surfers and scuba divers as well as to marine researchers. The following data on x = shark length (in feet) and y = jaw width (in inches) for 44 sharks was found in various articles appearing in the magazines *Skin Diver* and *Scuba News*. Because it is difficult to measure jaw width in living sharks, researchers would like to determine whether it is possible to estimate jaw width from shark length, which is more easily measured. Assume the necessary assumptions have been met.

```
Regression Analysis: Jaw Width versus Shark Length

Predictor        Coef     SE Coef           T        P
Constant        0.688       1.299        0.53    0.599
Shark Length  0.96345     0.08228       11.71    0.000

S = 1.376       R-Sq = 76.6%     R-Sq(adj) = 76.0%

Analysis of Variance

Source          DF       SS        MS         F        P
Regression       1   259.53    259.53    137.12    0.000
Residual Error  42    79.49      1.89
Total           43   339.02

Predicted Values for New Observation

New Obs     Fit   SE Fit       95.0% CI           95.0% PI
x*=17    17.067    0.238   (16.587,  17.546)  (14.249,  19.884)
```



(a) Give the equation of the least-squares regression line for this problem. (3 pts)

(b) Compute a 95% confidence interval for the slope parameter (β_1) of the simple linear regression model. (4 pts)

(c) Using the correct units of the data, give a practical interpretation of the confidence interval for β_1 given in the previous part. (5 pts)

(d) Give a practical interpretation of the 95% confidence interval for the mean response when $x^* = 17$. (5 pts)

(e) Interpret r^2 in terms of the problem. (4 pts)

14. **Old Faithful** Data was collected for three variables measured for 60 consecutive eruptions of the geyser Old Faithful in Yellowstone National Park. They are the duration of the eruption (ERUPTION DURATION, x_1), the duration of the dormant period immediately before the eruption (BEFORE DORMANT DURATION, x_2), and the duration of the dormant period immediately after the eruption (AFTER DORMANT DURATION, y). All the times are in minutes. Assume the necessary assumptions have been met. Find the Minitab output for this problem is attached.

(a) First consider conducting the multiple linear regression overall test.

i. State the appropriate hypotheses (H_0 and H_a). (3 pts)

ii. Give the value of the test statistic used to test the above hypotheses? (2 pts) _____

iii. Give the distribution of the test statistic assuming the H_0 is true? (3 pts) _____

iv. Give the p–value. (2 pts) _____

v. Give a practical conclusion *in terms of the problem.* (5 pts)

(b) Consider conducting a multiple linear regression follow-up test, assuming that the overall test is significant.

i. State the hypotheses (H_0 and H_a) which correspond to the test with t = 0.32. (3 pts)

ii. Give the distribution of the test statistic assuming the H_0 is true. (3 pts) _____

iii. Find the p-value. (3 pts) _____

iv. State your decision about H_0 at $\alpha = 0.05$? (2 pts) _____

v. Based on your decision, give a practical conclusion *in terms of the problem.* (5 pts)

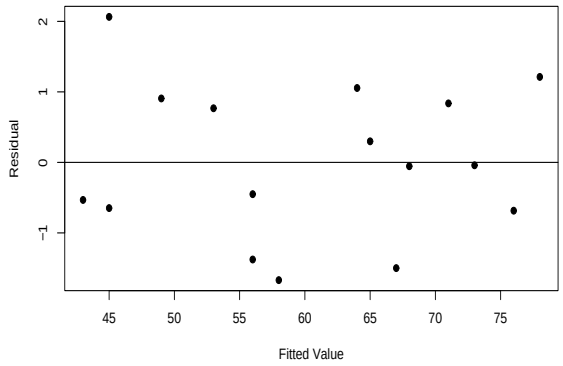
(c) Consider the usefulness of the predictor variables.

i. If you could find the value of either x_1 or x_2, but not both, which one would you want to know to predict the mean response? Circle one: $\boldsymbol{x_1}$ or $\boldsymbol{x_2}$ (3 pts)

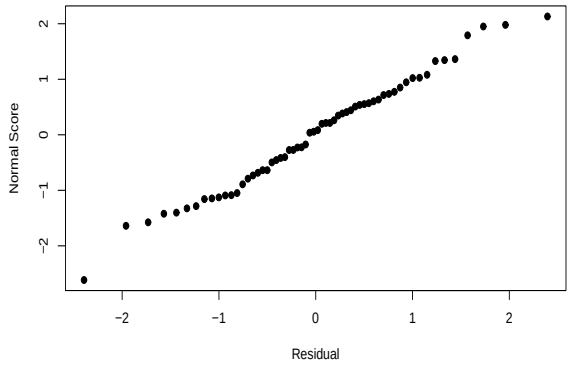
ii. Out of the three regression models given (Model 1 uses x_1 and x_2; Model 2 uses x_1 only; and Model 3 uses x_2 only), which one would you choose? *You must justify your answer to receive any credit.* (3 pts)

15. Consider Plots A and B to answer the following questions. (2 pts each)

PLOT A



PLOT B



(a) Which plot (A or B) is used to check the linear relationship assumption? _____

(b) Which plot (A or B) is used to check the constant variance assumption? _____

(c) Which plot (A or B) is used to check the normality of the residuals assumption? _____

Minitab Output for problem 14

MLR: y versus x1 and x2

```
Regression Analysis: y versus x1, x2

Predictor        Coef     SE Coef          T        P
Constant      119.904       7.427      16.14    0.000
x1              0.743       2.327       0.32
x2            -0.7292      0.1770      -4.12    0.000

S = 10.01       R-Sq = 46.6%     R-Sq(adj) = 44.8%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2      4985.9      2492.9     24.90    0.000
Residual Error    57      5706.9       100.1
Total             59     10692.7
```

SLR: y versus x1

```
Regression Analysis: y versus x1

Predictor        Coef     SE Coef          T        P
Constant       95.668       5.121      18.68    0.000
x1             -7.288       1.437      -5.07    0.000

S = 11.30       R-Sq = 30.7%     R-Sq(adj) = 29.5%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      3286.0      3286.0     25.73    0.000
Residual Error    58      7406.8       127.7
Total             59     10692.7
```

SLR: y versus x2

```
Regression Analysis: y versus x2

Predictor        Coef     SE Coef          T        P
Constant      119.090       6.921      17.21    0.000
x2          -0.68189     0.09598      -7.10    0.000

S = 9.928       R-Sq = 46.5%     R-Sq(adj) = 45.6%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      4975.7      4975.7     50.48    0.000
Residual Error    58      5717.1        98.6
Total             59     10692.7
```