Key: dbede dedec bddab caabc daeca

MATH202

Spring 2007

Exam 2a

Name:____

Section:_________

Instructions:

1. Do not start until instructed to do so.
2. If you brought a cell phone by mistake, turn it off and place it under your seat. You may NOT use it as a calculator.
3. You may use a calculator (NOT a cell phone calculator), but nothing else.
4. Code your **UDelNet ID** in the **Last Name** space on your scansheet and fill in the bubbles.
5. Write your name in the white space below the name box on your scansheet.
6. DO NOT put any part of your Social Security Number on your scansheet.
7. Choose the **best** answer to each question.
8. Use α = .05 unless otherwise specified.

Questions 1 – 4: In the distribution center of McCormick and Co., Inc., data were collected on a random sample of 9 orders. The time to fill the order (in minutes) and the size of the order (number of cases) were recorded. A simple linear regression analysis was performed to see if time is related to the size of the order, on average. If so, McCormick and Co. can use the results to predict fill time for future orders. Some regression results are shown below.

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3017.122	3017.122	12.21436	0.010062
Residual	7	1729.1	247.0143		
Total	8	4746.222			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	12.59432	6.58817	1.911656	0.09752
Size	0.109357	0.03129	3.494905	0.010062

1. Which of the following is the value of s, the standard deviation of the regression line?
   a. 0.03129 b. 1729.1 c. 247.0143 d. 15.7167 e. 6.58817

2. Interpret the slope in the above problem.
   a. For every additional case ordered, average fill time increases by 12.59432 minutes.
   b. For every additional case ordered, average fill time increases by 0.109357 minutes.
   c. For every additional 0.109357 cases ordered, average fill time increases by 1 minute.
   d. The average fill time for an order with no cases is 12.59432 minutes.
   e. For every additional 12.59432 cases ordered, average fill time increases by .109357 minutes.

3. Which of the following is a ***correct*** conclusion? Choose all that are correct.
   a. This model would probably not be useful for predicting time to fill orders.
   b. The estimated slope is significantly different from 0.
   c. There is enough evidence of a linear relationship between order size and time to fill the order in the population.
   d. There is not enough evidence of a linear relationship between order size and time to fill the order in the population.
   e. Both b and c.

4. Interpret the value of R^2 (.636).
   a. About 63.6% of the points fall on the estimated regression line.
   b. The estimated model would give us exact predictions of the time to fill orders about 63.6% of the time.
   c. About 63.6% of the variation in the sample of order sizes is explained by the estimated model.
   d. About 63.6% of the variation in the sample of times to fill orders is explained by the estimated model.
   e. There is a 63.6% chance that this model would be useful to predict fill time in the population of orders.

Questions 5 – 6: The data summarized below are from *Climatology Report* No. 77-3 (by J. F. Benci and T. B. McKee, Department of Atmospheric Science, Colorado State University). Let x be the elevation (thousands of feet) and let y be the average number of frost-free days in a year. Data on these two variables for the cities of Denver, Gunnison, Aspen, Crested Butte, and Dillon, Colorado were obtained.

$\bar{x} = 7.92$ $\bar{y} = 73.6$ $SS_{xx} = 11.408$ $SS_{xy} = -352.26$ $r^2 = .963$

5. Find the estimated regression equation.
   a. $\hat{y} = 2280.57 - 30.878x$
   b. $\hat{y} = 73.6 + 7.92x$
   c. $\hat{y} = -170.956 - 30.878x$
   d. $\hat{y} = -30.878 + 318.156x$
   e. $\hat{y} = 318.156 - 30.878x$

6. What is the value of the correlation coefficient?
   a. .963
   b. –.963
   c. .981
   d. –.981

Questions 7 – 8: The relationship between the rate at which crickets chirp (chirps per second) and the outside temperature (°F) was studied in detail by George W. Pierce, a physics professor at Harvard (*The Songs of Insects*, Harvard UP, 1948). The following is a summary of his data on 15 crickets using the regression model

$$chirps = \beta_0 + \beta_1 temp + \varepsilon .$$

Regression Analysis: chirps versus temp

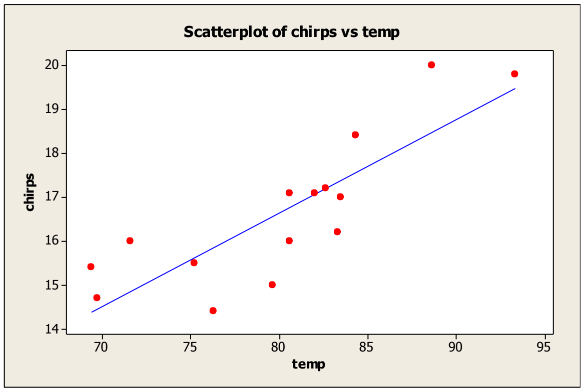
```
The regression equation is
chirps = - 0.31 + 0.212 temp


Predictor      Coef  SE Coef      T      P
Constant     -0.309    3.109  -0.10  0.922
temp        0.21193  0.03871   5.47  0.000


S = 0.971518   R-Sq =        R-Sq(adj) = 67.4%


Analysis of Variance

Source          DF      SS      MS      F      P
Regression                  28.287  29.97  0.000
Residual Error               0.944
Total
```



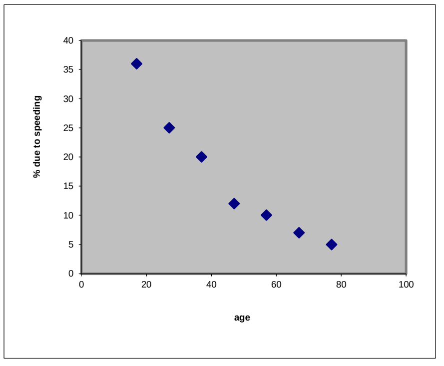
7. What is the value of R^2?
   a. .303
   b. .833
   c. .682
   d. .674
   e. .697

8. Interpret the value of s in the context of this problem.
   a. About 97% of the variation in chirp rates is explained by the estimated regression model.
   b. We estimate that an increase of 1° corresponds to an increase of .97 chirps per second on average.
   c. The temperature values differ from their mean by about .97°, on average.
   d. About 95% of chirp rates are within 1.94 chirps per second of their predicted values.
   e. Using the estimated model to predict temperature would result in correct predictions about 97% of the time.

Question 9: Data for this problem are based on information taken from *The Wall Street Journal*. Let x be the age (years) of a driver and let y be the percentage of all fatal accidents for that age that are due to speeding. A quadratic regression analysis is shown below based on the data.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.995831
R Square	0.991679
Adjusted R Square	0.987519
Standard Error	1.248809
Observations	7



ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	743.4762	371.7381	238.3664	6.92E-05
Residual	4	6.238095	1.559524		
Total	6	749.7143			

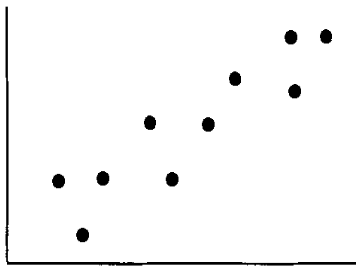
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	54.18964	2.743857	19.74944	3.88E-05	46.57146	61.80783
x	-1.24619	0.130237	-9.56864	0.000666	-1.60779	-0.88459
x^2	0.007976	0.001363	5.853817	0.004249	0.004193	0.011759

9. Which of the following is the ***correct*** statement?
   a. On average, as age increases by 1 year, fatal accidents due to speeding decreases by an estimated 1.25%.
   b. On average, as age increases by 1 year, fatal accidents due to speeding decreases by an estimated .008%.
   c. The curvature we see in the data is not statistically significant.
   d. The p-value of .0042 comes from the test of $Ho: x^2 = 0$ vs. $Ha: x^2 \neq 0$.
   e. The squared term in the model is statistically useful for predicting the percentage of fatal accidents due to speeding.

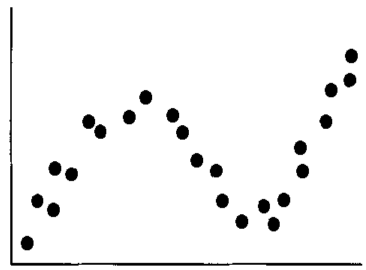
10. Which of the following plots would give the highest value of $|r|$, the absolute value of the correlation coefficient?

a. A b. B c. C d. D

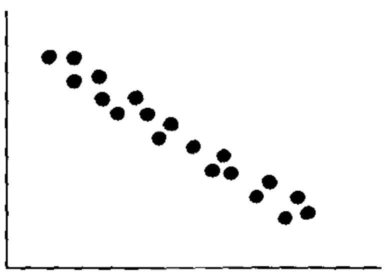
A



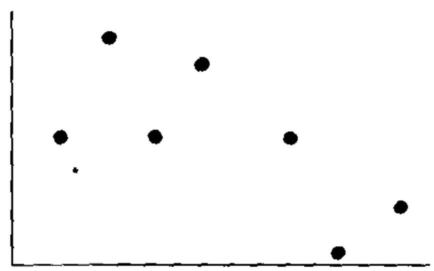
B



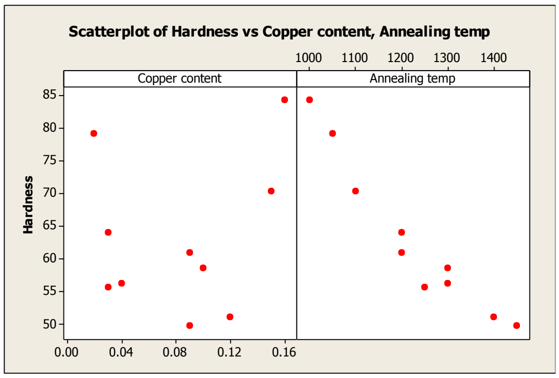
C

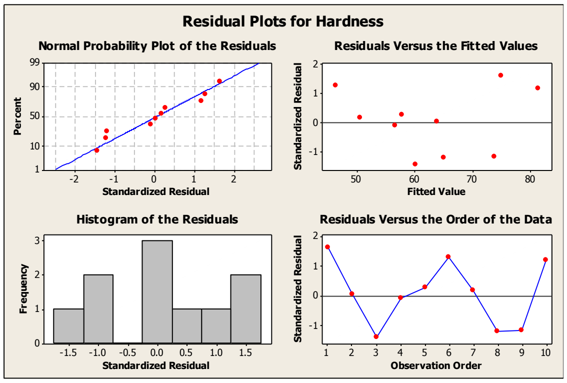


D



Questions 11 – 19: A steel company will be producing cold-reduced sheet steel consisting of 0.15 percent copper at an annealing temperature of 1150 degrees Fahrenheit. The company is interested in estimating the mean hardness of this steel. It collected data on 10 different specimens of sheet steel that had been produced at different copper contents and annealing temperatures. Data and Minitab output are given below.





Model1

Regression Analysis: Hardness versus Copper Content, Annealing Temp

The regression equation is
Hardness = 154 + 19.3 Copper Content - 0.0754 Annealing Temp

Predictor	Coef	SE Coef	T	P
Constant	153.84	10.43	14.74	0.000
Copper Content	19.30	23.12	0.84	0.431
Annealing Temp	-0.075431	0.008122	-9.29	0.000

S = 3.52168   R-Sq = 92.8%   R-Sq(adj) = 90.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	1122.45	561.22	45.25	0.000
Residual Error	7	86.82	12.40		
Total	9	1209.26			

Source	DF	Seq SS
Copper Content	1	52.73
Annealing Temp	1	1069.72

Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	69.99	1.94	(65.40, 74.58)	(60.48, 79.50)

Values of Predictors for New Observations

New Obs	Copper content	Annealing temp
1	0.150	1150

Model 2

Regression Analysis: Hardness versus Copper Content, Annealing Temp, ...

The regression equation is
Hardness = 172 - 159 Copper Content - 0.0913 Annealing Temp + 0.156 Content*Temp

Predictor	Coef	SE Coef	T	P
Constant	172.33	28.04	6.14	0.001
Copper Content	-159.0	250.6	-0.63	0.549
Annealing Temp	-0.09132	0.02377	-3.84	0.009
Content*Temp	0.1564	0.2189		0.502

S = 3.65167   R-Sq = 93.4%   R-Sq(adj) = 90.1%

```
Analysis of Variance

Source          DF       SS      MS      F      P
Regression       3  1129.26  376.42  28.23  0.001
Residual Error   6    80.01   13.33
Total            9  1209.26


Source          DF   Seq SS
Copper Content   1    52.73
Annealing Temp   1  1069.72
Content*Temp     1     6.81
```

Model 3

Regression Analysis: Hardness versus Copper Content, Annealing Temp, ...

```
The regression equation is
Hardness = 389 + 61 Copper Content - 0.459 Annealing Temp - 411 Content^2
           + 0.000155 Temp^2


Predictor             Coef     SE Coef      T      P
Constant            389.39       71.92   5.41  0.003
Copper Content        60.8       121.8   0.50  0.639
Annealing Temp     -0.4592      0.1161  -3.96  0.011
Content^2           -410.8       730.6  -0.56  0.598
Temp^2          0.00015462  0.00004585   3.37  0.020


S = 2.25371   R-Sq = 97.9%   R-Sq(adj) =


Analysis of Variance

Source          DF       SS      MS      F      P
Regression       4  1183.87  295.97  58.27  0.000
Residual Error   5    25.40    5.08
Total            9  1209.26

Source          DF   Seq SS
Copper Content   1    52.73
Annealing Temp   1  1069.72
Content^2        1     3.66
Temp^2           1    57.76
```

Model 4

Regression Analysis: Hardness versus Copper Content, Annealing Temp, ...

```
The regression equation is
Hardness = 386 - 26 Copper Content - 0.449 Annealing Temp + 0.059 Content*Temp
           - 288 Content^2 + 0.000148 Temp^2

Predictor             Coef     SE Coef      T      P
Constant               386       79.74   4.84  0.008
Copper Content         -26       277.2  -0.09  0.930
Annealing Temp      -0.449      0.1312  -3.42  0.027
Content*Temp         0.059      0.1642   0.36  0.739
Content^2             -288       875.1  -0.33  0.759
Temp^2            0.000148  0.00005335   2.78  0.050
```

```
S = 2.48054   R-Sq = 98.0%   R-Sq(adj) = 95.4%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       5  1184.65  236.93  38.51  0.002
Residual Error   4    24.61    6.15
Total            9  1209.26

Source          DF   Seq SS
Copper Content   1    52.73
Annealing Temp   1  1069.72
Content*Temp     1     6.81
Content^2        1     7.76
Temp^2           1    47.63
```

11. Use Model 1 to test if there is a linear relationship between the hardness of the steel, and copper content and/or annealing temperature. State the null hypothesis.
    a. $H_0: \beta_0 = \beta_1 = \beta_2 = 0$
    b. $H_0: \beta_1 = \beta_2 = 0$
    c. $H_0: \hat{\beta}_0 = \hat{\beta}_1 = \hat{\beta}_2 = 0$
    d. $H_0: \textit{all beta parameters are equal}$
    e. $H_0: \hat{\beta}_1 = \hat{\beta}_2 = 0$

12. Perform a hypothesis test on β_1 from Model 1. Does copper content contribute significantly to the prediction of the hardness of the steel?
    a. Yes, copper content contributes significantly to the prediction of the hardness of the steel, since the test statistic for β_1 has a p-value of 0.431.
    b. Yes, copper content contributes significantly to the prediction of the hardness of the steel, since the test statistic for β_1 has a p-value of 0.00.
    c. Copper content does not contribute significantly to the prediction of the hardness of the steel, since the test statistic (t= 0.84) for β_1, falls in the rejection region $|t| > 2.365$.
    d. Copper content does not contribute significantly to the prediction of the hardness of the steel, since the test statistic (t= 0.84) for β_1, falls in the non-rejection region $|t| \leq 2.365$.
    e. Since the F-value of 45.25 falls in the rejection region $F > F_{(0.05,\, 2,7)}$, we must conclude that copper content contributes significantly to the prediction of the hardness of the steel.

13. Interpret the estimated coefficient $\hat{\beta}_2$ from Model 1.
    a. The hardness of the steel is expected to increase by approximately 0.075 units for every $°F$ increase in annealing temperature.
    b. The hardness of the steel is expected to decrease by approximately 0.075 units for every $°F$ increase in annealing temperature.
    c. The hardness of the steel is expected to increase by approximately 0.075 units for every $°F$ increase in annealing temperature, for a given copper content.
    d. The hardness of the steel is expected to decrease by approximately 0.075 units for every $°F$ increase in annealing temperature, for a given copper content.
    e. The annealing temperature of the steel is expected to decrease by approximately 0.075 $°F$ for every unit increase in the hardness of the steel

14. Use Model 1 to find a 95 % interval estimate of the mean hardness of the steel to be produced for all specimens with .150 copper content and annealing temperature of 1150.
    a. (65.40, 74.58)
    b. (60.48, 79.50)
    c. (-35.38, 73.98)
    d. (-0.094, -0.056)
    e. (68.05, 71.93)

15. Use Model 2 to determine the effect of temperature on hardness.
    a. The hardness of the steel is expected to increase by 0.1564 units if the temperature increases by 1 $^\circ F$ for a given copper content.
    b. The hardness of the steel is expected to change by (-0.0913 + 0.156*content) units if the temperature increases by 1 $^\circ F$ for a given copper content.
    c. The hardness of the steel is expected to decrease by 0.0913 units if the temperature increases by 1 $^\circ F$ for a given copper content.
    d. The temperature has no effect of the hardness of the steel.
    e. It is virtually impossible to statistically determine the effect of temperature on the hardness of the steel.

16. Use Model 2 to test the inclusion of the interaction term. Calculate the test statistic and state the critical region, respectively.
    a. T = 0.71, $|z|$ > 1.96, the interaction term does not contribute significantly to the model.
    b. T = 0.71, $|t|$ > 2.447, the interaction term contributes significantly to the model.
    c. T = 0.71, $|t|$ > 2.447, the interaction term does not contribute significantly to the model.
    d. T = 0.23, $|z|$ > 1.96, the interaction term does not contribute significantly to the model.
    e. T = 0.71, $|t|$ > 1.943, the interaction term does not contribute significantly to the model.

17. Use Model 4 to determine the mean hardness of the steel with a copper content of 0.15 % at a temperature of 1150 $^\circ F$.
    a. approximately 65.18 units of hardness.
    b. approximately 72.16 units of hardness.
    c. approximately 95.40 units of hardness.
    d. approximately 69.99 units of hardness.

18. Compute R^2_{adj} for Model 3.
    a. Approximately .962
    b. Approximately .038
    c. Approximately .979
    d. Approximately .988
    e. Approximately .012

19. Interpret $\hat{\beta}_0$ = 153.84 for Model 1.

a. The mean hardness of the steel is estimated to increase by 153.84 units, if the steel is at 0 $°F$ and contains no copper.
b. The mean hardness of the steel that contains no copper, and its temperature is at 0 $°F$ is estimated to 153.84 units.
c. The mean hardness of the steel increases linearly at a rate of 153.84 units for every $°F$ increase in temperature, for a given percent of copper.
d. We estimate that the mean hardness of the steel will increase linearly at a rate of 153.84 units for every 1 % increase in copper content at a given temperature level.

Questions 20 – 22: Interviews were conducted with over 1,000 vendors in a certain city on the East Coast in order to study the factors influencing vendor's incomes. The researchers collected data on age, hours worked per day, and annual earnings. Minitab was used to obtain some output shown below.

Regression Analysis: Annual Earnings,y versus Age, X1, Hours Worked perDay, X2

```
Predictor            Coef  SE Coef      T      P
Constant            -20.4    652.7  -0.03  0.976
Age, X1            13.350    7.672   1.74  0.107
HrsWorked/Day, X2  243.71    63.51   3.84  0.002


S = 547.737   R-Sq = 58.2%   R-Sq(adj) = 51.3%


Analysis of Variance

Source          DF       SS       MS     F      P
Regression       2  5018232  2509116         0.005
Residual Error  12  3600196   300016
Total           14  8618428


Source             DF   Seq SS
Age, X1             1   600498
HrsWorked/Day, X2   1  4417734


Unusual Observations

Obs  Age, X1  AnnEarnings,y   Fit  SE Fit  Residual  St Resid
  4     18.0           1552  2657     205     -1105    -2.18R

R denotes an observation with a large standardized residual.
```

20. Write a hypothesized first-order model for annual earnings as a function of age and hours worked per day.

a. $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$

b. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

c. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

d. $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

21. Is hours worked per day a statistically useful predictor of annual earnings? The p-value used for this test is

a. 0.005 b. 0.976 c. 0.107 d. 0.002 e. none of these

22. Determine the global F-test statistic for testing the overall utility of the model.

a. 8.36 b. 2.81 c. 3.89 d. 5.10 e. 6.93

Questions 23 – 25: Data on company profit (in \$ millions), CEO's annual income (in \$ thousands), and percentage of the company's stock owned by the CEO are collected on a random sample of companies.

Y = company profit
X_1 = CEO's annual income (in thousands of dollars)
X_2 = CEO's stock in a company (in %)

The interaction model was fit to the data with the following results:

Regression Analysis: Profit, Y versus Income, X1, Stock, X2 (%), X1X2

```
Predictor          Coef   SE Coef      T      P
Constant         1170.9     993.1   1.18  0.272
Income, X1      0.12144   0.04251   2.86  0.021
Stock, X2 (%)      5.89     61.24         0.926
X1X2           -0.03531   0.01166  -3.03  0.016


S = 2311.30   R-Sq = 57.0%   R-Sq(adj) = 40.9%


Analysis of Variance

Source          DF        SS        MS     F      P
Regression       3  56742740  18914247  3.54  0.068
Residual Error   8  42736721   5342090
Total           11  99479461


Source         DF    Seq SS
Income, X1      1   1683416
Stock, X2 (%)   1   6056233
X1X2            1  49003090


Unusual Observations

Obs  Income, X1  Profit, Y  Fit  SE Fit  Residual  St Resid
  3         855        346  207    2304       139      0.75 X

X denotes an observation whose X value gives it large influence.




Predicted Values for New Observations
```

```
New
Obs   Fit  SE Fit      95% CI          95% PI
  1  8707    2275  (3461, 13953)  (1228, 16185)


Values of Predictors for New Observations

New              Stock,
Obs  Income, X1  X2 (%)   X1X2
  1       70000    4.00  28000
```

23. Which of these is the statistical model?
    a. $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

    b. $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$

    c. $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$

    d. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \varepsilon$

    e. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$

24. Give the p-value for testing whether CEO's annual income and CEO's stock in the company interact.

    a. 0.272 b. 0.068 c. 0.016 d. 0.021 e. 0.926

25. Determine the t-test statistic for the test concerning β_2.

    a. 0.10 b. 0.003 c. 3.54 d. 1.18 e. 3.03