Practice Exam 4
(Answers are found on the last page.)

Name:_____

Instructions:

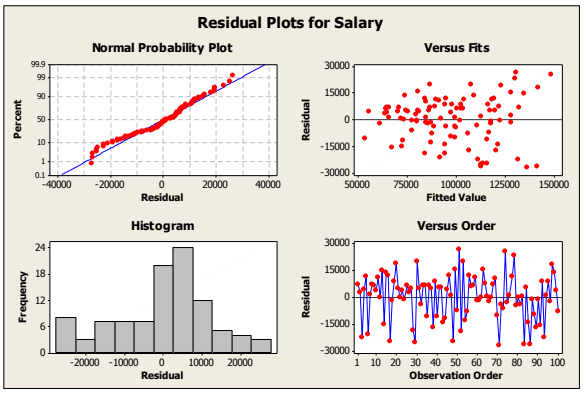
1. Do not start until instructed to do so.
2. You may use a scientific calculator (no graphing calculators allowed).
3. No other aids are allowed.
4. The work you turn in must be your own.
5. Use $\alpha = .05$, unless otherwise specified.
6. SHOW ALL WORK to receive full credit.

Questions 1 – 7: An international management consulting company develops a multiple regression model for executive salaries of its client firms. Shown below is regression output for the following model fit to data collected from a random sample of company executives:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

y = annual salary (in dollars)
x_1 = experience (years)
x_2 = education (years)
x_3 = number of employees supervised
x_4 = corporate assets (millions of dollars)

Regression Analysis: Salary versus X1, X2, X3, X4




```
The regression equation is
Salary = - 37086 + 2696 X1 + 2656 X2 + 41.1 X3 + 245 X4


Predictor     Coef  SE Coef      T      P
Constant   -37086    17053  -2.17  0.032
X1          2696.4    173.7  15.53  0.000
X2          2656.0    563.5   4.71  0.000
X3          41.093    7.807   5.26  0.000
X4          244.59    83.42   2.93  0.004


S = 12685.9   R-Sq = 75.7%   R-Sq(adj) = 74.7%

Analysis of Variance

Source           DF           SS           MS      F      P
Regression        4  47663114838  11915778709  74.04  0.000
Residual Error   95  15288485309    160931424
Total            99  62951600146


Predicted Values for New Observations

New Obs    Fit  SE Fit      95% CI           95% PI
      1  83396    1763  (79896, 86895)  (57969, 108822)


Values of Predictors for New Observations

New Obs    X1    X2   X3   X4
      1  10.0  16.0  200  175
```

1. **4 points** Give the value of $\hat{\beta}_1$ and interpret this value in the context of this problem.

2. **1 point** How many executives were in the sample?

3. **3 points** Note that s = 12,685.9. Interpret this number in the context of this problem.

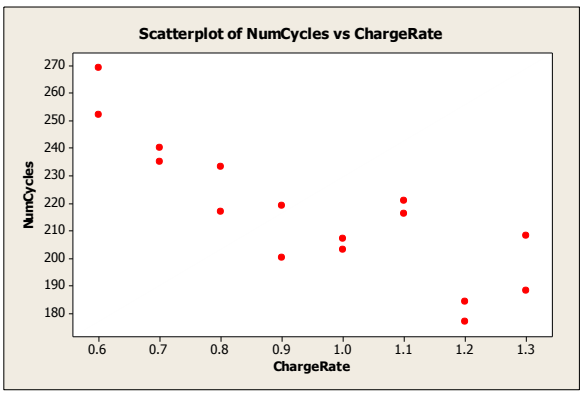
4. **5 points** Perform a test of overall model adequacy. Give the hypotheses, the value of the test statistic, either the rejection region or p-value, and write a conclusion in terms of the problem.

5. **3 points** List the predictor variables that are significantly related to salary.

6. **3 points** Interpret the numbers under the heading 95% CI in the output above.

7. **2 points** Which of the following regression assumptions seem to be least satisfied?
   a. residuals are independent
   b. residuals have a normal distribution
   c. residuals have a mean of 0 at each fitted value
   d. residuals have constant variance

Questions 8 – 10: A researcher studied the effects of the charge rate (amperes) on the performance of batteries. Two batteries were randomly assigned to each setting of charge rate (from .6 to 1.3) and the number of cycles (discharge-charge) that the battery underwent until it failed was observed. The data and some analyses are shown below.



NumCycles	ChargeRate
269	0.6
252	0.6
235	0.7
240	0.7
233	0.8
217	0.8
200	0.9
219	0.9
203	1.0
207	1.0
221	1.1
216	1.1
177	1.2
184	1.2
208	1.3
188	1.3

Regression Analysis: NumCycles versus ChargeRate, ChargeRate^2

```
The regression equation is
NumCycles = 415 - 343 ChargeRate + 134 ChargeRate^2


Predictor       Coef  SE Coef      T      P
Constant      414.82    59.32   6.99  0.000
ChargeRate    -342.8    129.9  -2.64  0.020
ChargeRate^2  133.63    68.01   1.96  0.071


S = 12.4664   R-Sq = 78.2%   R-Sq(adj) = 74.8%


Analysis of Variance

Source          DF        SS        MS        F      P

Regression     ___  ________  ________  _______  0.000

Residual Error ___    2020.4  ________

Total          ___    9254.4
```

8. **3 points** Write down the quadratic <u>model</u> (not the least-squares regression equation) that relates the number of cycles to the charge rate.

9. **4 points** Complete the ANOVA table above by filling in the blanks.

10. **5 points** Conduct a test that compares this model to the straight-line model. Is there enough evidence that the quadratic model fits the data better?

Questions 11 – 12: Data on sales last year in 26 sales districts are given below for a maker of asphalt roofing shingles. A regression model is used to try to predict sales by using data on promotional expenditures and the number of active accounts in each district.

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

$y =$ sales (thousands of squares)
$x_1 =$ promotional expenses (thousands of dollars)
$x_2 =$ number of active accounts

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.822806
R Square	0.677009
Adjusted R Square	0.632965
Standard Error	51.22281
Observations	26

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	120991.1	40330.37	15.37112	1.3E-05
Residual	22	57723.08	2623.777		
Total	25	178714.2			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	199.4904	140.0301	1.424625	0.168298	-90.9142	489.895
x1	-49.6612	25.05911	-1.98176	0.060137	-101.631	2.308178
x2	-0.53737	2.589201	-0.20754	0.837496	-5.90704	4.832307
x1x2	0.938325	0.450894	2.081034	0.04928	0.003229	1.873422

11. **3 points** Explain what it means to say that promotional expenses and number of accounts interact.

12. **3 points** Is there significant evidence of interaction in the data? Explain.

13. **3 points** Define a set of dummy variables that would be used to account for the information in the categorical variable "tool model" at 5 levels (M1, M2, M3, M4, M5).

An economist wishes to relate the speed with which a particular insurance innovation is adopted to the type of firm (stock company or mutual company). Consider the following model:

$E(y) = \beta_0 + \beta_1 x$

$y =$ number of months elapsed from when the first firm in the industry adopted the innovation to when the given firm adopted it

$x = 1$ if stock firm; 0 otherwise

Regression Analysis: Months versus x

```
The regression equation is
Months = 16.7 + 5.40 x


Predictor    Coef  SE Coef     T      P
Constant   16.700    2.920  5.72  0.000
x           5.400    4.130  1.31  0.207


S = 9.23460   R-Sq = 8.7%   R-Sq(adj) = 3.6%


Analysis of Variance

Source          DF       SS      MS     F      P
Regression       1   145.80  145.80  1.71  0.207
Residual Error  18  1535.00   85.28
Total           19  1680.80
```

14. **8 points** Estimate the average time until adoption for all stock firms. Repeat for all mutual firms. Is the observed difference statistically significant? Explain.

Answers

1. 2696.4; We estimate that a 1 year increase in experience is associated with a $2,696.40 increase in annual salary, holding years of education, number of employees supervised, and corporate assets constant.

2. 100

3. The average difference between an executive's actual annual salary and the predicted salary from this multiple regression model is estimated to be $12,685.90.

4. H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$
   H_a: $not\ H_0;\ \ at\ least\ one\ \beta\ is\ nonzero; at\ least\ one\ predictor\ term\ is\ useful$

   $F = 74.04$ RR: $F >$ a value between 2.45 and 2.53 $p - value = .000$

   We have strong evidence that at least one predictor term is useful in explaining variation in annual salaries.

5. Years experience, years of education, number of employees supervised, value of corporate assets

6. We are 95% confidence that the mean annual salary for all executives having 10 years experience, a bachelor's degree, supervise 200 employees, and whose company has assets of $175 million is between $79,896 and $86,895.

7. residuals have constant variance

8. $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ or $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$

   where $x =$ charge rate and $y =$ number of cycles until failure

9.

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	7234.0	3617.0	23.27	0.000
Residual Error	13	2020.4	155.4		
Total	15	9254.4			

10. H_0: $\beta_2 = 0$ H_a: $\beta_2 \neq 0$ $t = 1.96$ RR: $|t| > 2.160$ $p - value = .071$
    We do not have enough evidence at the 5% significance level that the quadratic model fits the data better than the straight-line model.

11. The effect of promotional expenses on average sales depends on the number of active accounts.
    or
    The effect of the number of active accounts on average sales depends on promotional expenses.

12. Yes, just barely at the 5% significance level. The test for H_0: $\beta_3 = 0$ gives a p-value of .04928.

13. $x_1 = 1\ if\ M1; 0\ otherwise$
    $x_2 = 1\ if\ M2; 0\ otherwise$
    $x_3 = 1\ if\ M3; 0\ otherwise$
    $x_4 = 1\ if\ M4; 0\ otherwise$

14. 22.1 months; 16.7 months
    No. The test for H_0: $\beta_1 = 0$ gives a p-value of .207 which is not significant at the usual levels of significance.