**CHE32000**

**Statistical Modeling and Quality Enhancement**

**Midterm Examination 3**

**Spring Semester 2017**

**Time Allowed: 60 minutes**

**Total Marks: 100 marks**

**Topics Covered Chapter 6 and Part of Chapter 7 (till page 386)**

**Students may bring 2 pages of crib sheet.**

**Graphing calculators allowed.**

**SHOW ALL WORKING ON THE EXAM PAPER. Numerically correct answers with insufficient justification may not receive full marks. Please print your name on the upper right corner of each page and on the last page where indicated. Your signature is required on the bottom of the last page where indicated. Read all questions carefully; answer them concisely and neatly. Show all work on these pages. Point values in the left-hand margins are based on 100 points total for the examination.**

**STUDENT NAME: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**PUID: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**RECITATION SECTION: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Question 1**

An electrical power consumed by a chemical process plant was thought to depend on the following variables: average ambient temperature (x1), number of days in a month (x2), average product purity (x3) tons of product produced (x4).

1. Find the mean power consumption given that x1= 85oF, x2=25 days, x3=90%, and x4=100 tons using the table of Parameter Estimates provided.  
    [4 marks]  
   Y=-102.7132+0.6054\*X1+8.9236\*X2+1.4375\*X3+0.0136\*X3

= 302.566 (3 pts for method, 1 pt for result)

1. Complete the ANOVA table by filling in the missing values. (1 pts each, no partial credits)  
    [6 marks]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | |  |  |  |  |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** | **Prob > F** |
| Model | 4 (1 pt) | 4957.2407 | 1239.31 (1 pt) | 5.106 (1 pt) | 0.0303 |
| Error | 7 | 1699.0093 (2 pts) | 242.72 |  |  |
| C. Total | 11 (1 pt) | 6656.25 |  |  |  |

2. Compute the 95% CI on this mean response (in a)? Compute the 95% PI on a future observation. What do you notice about the relative size of the two intervals? Which is wider and why?  
    [10 marks]

|  |  |  |  |
| --- | --- | --- | --- |
| **95% CI** |  | **95% PI** |  |
| \_\_278.991\_\_ | \_\_326.142\_ | \_\_258.829\_\_ | \_\_346.304\_ |
|  |  |  |  |

(1.5 pts each (0.5-1 pt was given for methods), 6 pts total)

The 95% PI is wider because it includes prediction error/additional variance/etc.

(4 pts (3 pts for partially reasonable answer))

1. Also attached is the residual plot. What can you infer from the Residual plot? How will you evaluate influential observations?

[10 marks]  
The residual plot shows that the normality assumptions are correct. (6 pts, partial credits were given to reasonable answer)

Influential observations can be evaluated using Cooks Distance/Studentized/Standardized residuals/etc. (4 pts, partial credits were given to reasonable answer)

JMP output from the ANOVA is attached.

**TOTAL FOR QUESTION 1 = 30 MARKS**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| RSquare | | | | | | | 0.74475 | | |
| RSquare Adj | | | | | | | 0.598893 | | |
| Root Mean Square Error | | | | | | | 15.57933 | | |
| Mean of Response | | | | | | | 278.75 | | |
| Observations (or Sum Wgts) | | | | | | | 12 | | |
| **Analysis of Variance** | | | | |  | | | | | |  | |  | | |  |
| **Source** | | | **DF** | | **Sum of Squares** | | | | | | **Mean Square** | | **F Ratio** | | | **Prob > F** |
| Model | | | \_\_4\_ | | 4957.2407 | | | | | | \_\_1239.31\_\_ | | 5.106 | | | 0.0303 |
| Error | | | 7 | | \_\_1699.0093\_\_ | | | | | | 242.72 | |  | | |  |
| C. Total | | | \_\_11 | | 6656.25 | | | | | |  | |  | | |  |
| **Parameter Estimates** | | | | | | |  | |  | | |  | |
| **Term** | | **Estimate** | | **Std Error** | | | | | **t Ratio** | | | **Prob>|t|** | |
| Intercept | | -102.7132 | | | 207.8589 | | | | -0.49 | | | 0.6363 | |
| X1 | | 0.6053705 | | | 0.368897 | | | | 1.64 | | | 0.1448 | |
| X2 | | 8.9236442 | | | 5.300522 | | | | 1.68 | | | 0.1361 | |
| X3 | | 1.4374567 | | | 2.391621 | | | | 0.6 | | | 0.5668 | |
| X4 | | 0.0136093 | | | 0.733821 | | | | 0.02 | | | 0.9857 | |
| **Predicted Values for New Observations** | | | | | | | | | | | |  | | |  | |  |
| **New Obs** | **Fit** | | | | | **SE Fit** | | **95% CI** | | | |  | | | **95% PI** | |  |
| 1 | \_\_302.566\_\_\_ | | | | | 9.9702 | | \_\_278.991\_\_ | | | | \_\_326.142\_ | | | \_\_258.829\_\_ | | \_\_346.304\_ |

****

**Figure 1. Residuals vs. Predicted Values**

**Show working for Q.1 here. Show all your working clearly to get full marks.**

**Working for Q. 1 (continued)**

**Question 2.**

a. Consider the “all possible regressions” approach in Table A, “forward selection” model in Table B, and “backward elimination” model in Table C. Which would be the most appropriate final model for each of these approaches and why? List your answers for each approach. [15 marks]

b. Table D provides the full model equation. Write the equation for the reduced model based on your conclusion in Part (a) [5 marks]

c. (i) If a strong multicollinearity is observed in a regression model, suggest two remedial measures you could implement to minimize its effect.

(ii) If the residual plot for regression indicates that variance increases with time, suggest one variance-stabilizing transformations for the response. [15 marks]

**TOTAL FOR QUESTION 2 = 35 MARKS**

**TABLE (A): All possible regressions.**

Response is Tumor Reduction.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Variables** | **R2** | **Adjusted R2** | **Mallows Cp** | **S** | **Age** | **Oxygen Saturation** | **Male/Female** | **Drugs** |
| A | 1 | 82.1 | 80.8 | 4.4 | 9.3577 | x |  |  |  |
| B | 1 | 57.0 | 54.0 | 27.3 | 14.487 |  | X |  |  |
| C | 2 | 87.9 | 86.1 | 1.1 | 7.9723 | x | X |  |  |
| D | 2 | 83.0 | 80.4 | 5.5 | 9.4476 | X |  |  | x |
| E | 3 | 88.0 | 85.0 | 3.0 | 8.2768 | X | x |  | X |
| F | 3 | 87.9 | 84.9 | 3.0 | 8.2942 | X | X | X |  |
| G | 4 | 88.0 | 83.6 | 5.0 | 8.6446 | x | X | x | X |

**Table (B) Forward selection.**

Alpha-to-Enter: 0.25

Response is Tumor reduction on 4 predictors, with N = 16

|  |  |  |
| --- | --- | --- |
| **Step** | **1** | **2** |
| **Constant** | 136.2 | 146.7 |
| **Age** | -1.43 | -1.12 |
| T-value | -8.01 | -5.76 |
| P-value | 0.000 | 0.000 |
| **Oxygen saturation** |  | -0.56 |
| T-value |  | -2.51 |
| P-value |  | 0.026 |
| **S** | 9.36 | 7.97 |
| **R2** | 82.07 | 87.92 |
| **R2 adjusted** | 80.79 | 86.06 |
| **Mallows Cp** | 4.4 | 1.1 |

**TABLE (C) Backward elimination.**

Alpha-to-Remove: 0.1

Response is Tumor reduction on 4 predictors, with N = 16

|  |  |  |  |
| --- | --- | --- | --- |
| **Step** | **1** | **2** | **3** |
| **Constant** | 146.2 | 146.2 | 146.7 |
| **Age** | -1.12 | -1.12 | -1.12 |
| T-value | -5.25 | -5.51 | -5.76 |
| P-value | 0.00 | 0.00 | 0.00 |
| **O2 saturation** | -0.59 | -0.59 | -0.56 |
| T-value | -2.11 | -2.22 | -2.51 |
| P-value | 0.058 | 0.046 | 0.026 |
| **Male/Female** | 0.1 |  |  |
| T-value | 0.03 |  |  |
| P-value | 0.979 |  |  |
| **Drugs** | 0.5 | 0.6 |  |
| T-value | 0.22 | 0.25 |  |
| P-value | 0.832 | 0.809 |  |
| **S** | 8.64 | 8.28 | 7.97 |
| **R-Sq** | 87.98 | 87.98 | 87.92 |
| **R-Sq(adj)** | 83.61 | 84.97 | 86.06 |
| **Mallows C-p** | 5.0 | 3.0 | 1.1 |

**Table D: Full Model Equation**

Tumor Reduction = 146.21 -1.12\*Age -0.59\*Oxygen saturation (%) + **A**\*(Male/Female) + 0.54\*Drugs (mg)

Where “**A**” for corresponding Male/Female is:

0 🡪 -0.06

1 🡪 0.06

Else 🡪 0

**Show working for Q.2 here:**

* 1. For All Possible Regressions: Model C with Age and Oxygen Saturation because smallest Cp, largest Adjusted R^2

Y = 146.2101.12\*Age-0.59\*OxygenSaturation

(5 pts if mentioned the reason as low Cp/S and high adjusted R^2. 5 pts even if only low Cp/S mentioned as a reason. 4 pts if only mentioned adjusted R^2 as a reason. 3 pts if answer is correct but reason doesn’t make sense).

* 1. Forward Selection: Step 2 with Age and Oxygen Saturation because P-values of significance are below 0.25 alpha to enter

Y = 146.2101.12\*Age-0.59\*OxygenSaturation

(5 pts if discussed significance/p-values/t-values. 4 pts if only mentioned low Cp/high adjusted R^2 as a reason. 3 pts if meaning of t- and p-values have been interchanged)

* 1. Backward Elimination: Step 3 with Age and Oxygen Saturation because P-values of significance are above 0.1 alpha to leave

Y = 146.2101.12\*Age-0.59\*OxygenSaturation

(Same scoring system as that for part (b))

1. Y = 146.2101.12\*Age-0.59\*OxygenSaturation

(5 pts if model is correct. 0 is wrong. If multiple models shown, out of which only one is correct, 0 pts.)

1. Parts:
   1. Strong multicollinearity can be remedied with: dropping variables, obtain more data, different regression methods, etc. Answers may vary. (5 pts per remedy metric)

(5 pts for each correct answer = 10 pts. 5 pts + 4 pts = 9 pts if one answer is removing one variable and the other answer is combining variables because both of them are almost the same.)

* 1. Transformations: 1/y, ln y, sqrt(y), etc. (5 pts)

(5 pts for correct answer)

**Working for Q.2 continued:**

**Question 3**.

In a manufacturing process, an engineer is interested in evaluating the effect of cutting speed (A), metal hardness (B), and cutting angle (C) on the life of a cutting tool. Two levels of each factor are chosen and three replicates of a 23 factorial design are run. The tool life data (in hrs) are shown in the table below:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Treatment Combination | Replicate | | | Total | Average | Variance |
|  | I | II | II |  |  |  |
| (1) | 221 | 255 | 311 | 787 | 262.33 | 2065.33 |
| a | 325 | 375 | 435 | 1135 | 378.33 | 3033.33 |
| b | 354 | 352 | 348 | 1054 | 351.33 | 9.33 |
| ab | 552 | 520 | 472 | 1544 | 514.67 | 1621.33 |
| c | 440 | 447 | 453 | 1340 | 446.67 | 42.33 |
| ac | 406 | 397 | 377 | 1180 | 393.33 | 220.33 |
| bc | 605 | 555 | 505 | 1665 | 555.00 | 2500.00 |
| abc | 392 | 405 | 419 | 1216 | 405.33 | 182.33 |
|  |  |  |  | 9921 | 413.38 | 1209.29 |

1. Calculate the Standard error for each of the coefficients in the parameter estimates table. [5 marks]
2. In the ANOVA table, below fill in the Degrees of freedom information for the Model, error, and Total, as well as the MSE. [10 marks]
3. What is the ‘F” value for the model. Test the significance of the model at α = 0.05. Clearly state the hypothesis it is testing. [10 marks]
4. From the table below with information from the regression coefficients, pick out the variables that will have the most significant effect. Justify your answer. [10 marks]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | |  |  |  |  |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** | **Prob > F** |
| Model | \_\_\_\_\_\_\_\_\_\_\_\_ | 179348.96 | 25621.3 | \_\_\_\_\_\_\_\_\_ | <.0001 |
| Error | \_\_\_\_\_\_\_\_\_\_\_\_ | 19348.67 | \_\_\_\_\_\_\_\_\_\_ |  |  |
| C. Total | \_\_\_\_\_\_\_\_\_\_\_\_ | 198697.63 |  |  |  |
|  |  |  |  |  |  |
| **Parameter Estimates** | |  |  |  |  |
| **Term** | **Estimate** | **Std Error** | **t Ratio** | **Prob > |t|** |  |
| Intercept | 413.375 | \_\_\_\_\_\_\_\_\_\_\_\_ | 58.24 | <.0001 |  |
| A | 9.5416667 | \_\_\_\_\_\_\_\_\_\_\_\_ | 1.34 | 0.1976 |  |
| B | 43.208333 | \_\_\_\_\_\_\_\_\_\_\_\_ | 6.09 | <.0001 |  |
| C | 36.708333 | \_\_\_\_\_\_\_\_\_\_\_\_ | 5.17 | <.0001 |  |
| A\*B | -6.125 | \_\_\_\_\_\_\_\_\_\_\_\_ | -0.86 | 0.401 |  |
| A\*C | -60.29167 | \_\_\_\_\_\_\_\_\_\_\_\_ | -8.49 | <.0001 |  |
| B\*C | -13.125 | \_\_\_\_\_\_\_\_\_\_\_\_ | -1.85 | 0.083 |  |
| A\*B\*C | -17.95833 | \_\_\_\_\_\_\_\_\_\_\_\_ | -2.53 | 0.0223 |  |

**Working for Question 3**.

**Working for Question 3 (contd)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | |  |  |  | |  |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | | **F Ratio** | **Prob > F** |
| Model | 7 | 179348.96 | 25621.3 | | 21.187 | <.0001\* |
| Error | 16 | 19348.67 | 1209.3 | |  |  |
| C. Total | 23 | 198697.63 |  | |  |  |
|  | 3X3 = 9 pts. 3 pts for each correct DF |  | 1 pt for correct answer. 0.5 pt if incorrect but following previous dof steps | |  |  |
| **Parameter Estimates** | |  |  |  | |  |
| **Term** | **Estimate** | **Std Error** | **t Ratio** | **Prob > |t|** | |  |
| Intercept | 413.375 | 7.098391 | 58.24 | <.0001\* | |  |
| A | 9.5416667 | 7.098391 | 1.34 | 0.1976 | |  |
| B | 43.208333 | 7.098391 | 6.09 | <.0001\* | |  |
| C | 36.708333 | 7.098391 | 5.17 | <.0001\* | |  |
| A\*B | -6.125 | 7.098391 | -0.86 | 0.401 | |  |
| A\*C | -60.29167 | 7.098391 | -8.49 | <.0001\* | |  |
| B\*C | -13.125 | 7.098391 | -1.85 | 0.083 | |  |
| A\*B\*C | -17.95833 | 7.098391 | -2.53 | 0.0223 | |  |

(5 pts if all standard errors are correct. 3 pts if some are correct)

1. Hypothesis is testing whether any effects are significant (Beta\_i != 0)

(6 pts for correct F ratio. 4 pts for model significance)

1. Most significant variables are: (both answers work, as long as correct justification is stated)

B,C: Main effects from P-value

A, B, C: main effects B, C but A since AC interaction is significant from P-value

(10 pts if all correct variables chosen. 7 pts if not all are correct)

Please provide the following information.

**Print your name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Sign your name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Please do not write below this line.

Total: