**Homework 3**

**Sergiu Buciumas**

1. Do problems 7.18, 7.19 and 7.20 in your text.

a. y = -1.77 +0.421†x1 +0.222†x2 -0.1280x3 -0.0193x1\*x1 -0.0074x2\*x2 +0.00082x3\*x3

-0.0199x1\*x2 +0.00915x1\*x3 +0.00258x2\*x3

**Regression Analysis: y versus x1, x2, x3**

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Regression 9 0.655671 0.072852 19.63 0.000

x1 1 0.007595 0.007595 2.05 0.172

x2 1 0.010745 0.010745 2.89 0.108

x3 1 0.012323 0.012323 3.32 0.087

x1\*x1 1 0.004913 0.004913 1.32 0.267

x2\*x2 1 0.001419 0.001419 0.38 0.545

x3\*x3 1 0.001213 0.001213 0.33 0.575

x1\*x2 1 0.010120 0.010120 2.73 0.118

x1\*x3 1 0.005352 0.005352 1.44 0.247

x2\*x3 1 0.000497 0.000497 0.13 0.719

Error 16 0.059386 0.003712

Total 25 0.715057

Model Summary

S R-sq R-sq(adj) R-sq(pred)

0.0609233 91.69% 87.02% 0.00%

Coefficients

Term Coef SE Coef T-Value P-Value VIF

Constant -1.77 1.29 -1.37 0.188

x1 0.421 0.294 1.43 0.172 521.01

x2 0.222 0.131 1.70 0.108 401.59

x3 -0.1280 0.0702 -1.82 0.087 688.02

x1\*x1 -0.0193 0.0168 -1.15 0.267 501.51

x2\*x2 -0.0074 0.0120 -0.62 0.545 173.60

x3\*x3 0.00082 0.00144 0.57 0.575 99.68

x1\*x2 -0.0199 0.0120 -1.65 0.118 204.43

x1\*x3 0.00915 0.00762 1.20 0.247 456.01

x2\*x3 0.00258 0.00704 0.37 0.719 349.97

Regression Equation

y = -1.77 +†0.421†x1 +†0.222†x2 -†0.1280†x3 -†0.0193†x1\*x1 -†0.0074†x2\*x2 +†0.00082†x3\*x3

-†0.0199†x1\*x2 +†0.00915†x1\*x3 +†0.00258†x2\*x3

Fits and Diagnostics for Unusual Observations

Obs y Fit Resid Std Resid

1 0.2220 0.2727 -0.0507 -2.02 R

10 0.4560 0.3485 0.1075 2.05 R

25 0.2510 0.1285 0.1225 2.23 R

26 0.0000 -0.0246 0.0246 2.80 R

R Large residual

b. For our model F = 19.63 with p=value = 0.000 being significant. All variable being non-significant.

|  |  |
| --- | --- |
| Coeficients | T-value P-value |
| β1 | 1.43 0.172 |
| β2 | 1.70 0.108 |
| β3 | -1.82 0.087 |
| β11 | -1.15 0.267 |
| β22 | -0.62 0.545 |
| β33 | 0.57 0.575 |
| β12 | -1.65 0.118 |
| β13 | 1.20 0.247 |
| β23 | 0.37 0.719 |

c. Per our Residual Plots we do observe several outliers which affect normality and the residual plot.



d. F= 1.6122 which is not significant.

7.19 VIF are very large, as we see in below table output, what indicating that there we have a serious problem with multicollinearity.

Coefficients

Term Coef SE Coef T-Value P-Value VIF

Constant -1.77 1.29 -1.37 0.188

x1 0.421 0.294 1.43 0.172 521.01

x2 0.222 0.131 1.70 0.108 401.59

x3 -0.1280 0.0702 -1.82 0.087 688.02

x1\*x1 -0.0193 0.0168 -1.15 0.267 501.51

x2\*x2 -0.0074 0.0120 -0.62 0.545 173.60

x3\*x3 0.00082 0.00144 0.57 0.575 99.68

x1\*x2 -0.0199 0.0120 -1.65 0.118 204.43

x1\*x3 0.00915 0.00762 1.20 0.247 456.01

x2\*x3 0.00258 0.00704 0.37 0.719 349.97

7.20 Points of interest: x1 = 8, x2 = 3 , x3 = 5

a. Regression Equation:

y = -1.77 + 0.421 x1 + 0.222 x2 - 0.1280 x3 - 0.0193 x1\*x1 - 0.0074 x2\*x2 + 0.00082 x3\*x3

- 0.0199 x1\*x2 + 0.00915 x1\*x3 + 0.00258 x2\*x3

Variable Setting

x1 8

x2 3

x3 5

Fit SE Fit 95% CI 95% PI

0.268864 0.0274987 (0.210569, 0.327158) (0.127165, 0.410562)

Predicted response at the point y^ = 0.26886 and at 95% CI on the mean response at the point is (0.210569, 0.327158).

b. Predicted response at the point is y^ = 0.25121 and a 95% CI on the mean response at point is (0.218464, 0.283960)

Prediction for y

Regression Equation

y = -0.259 + 0.0782 x1 + 0.1209 x2 - 0.1102 x3 - 0.0124 x1\*x2 + 0.00842 x1\*x3 + 0.00233 x2\*x

Variable Setting

x1 8

x2 3

x3 5

Fit SE Fit 95% CI 95% PI

0.251212 0.0156462 (0.218464, 0.283960) (0.118747, 0.383677)

c. Comparing the lengths of the confidence intervals in part a and b, appear that our model what include only main effect and two-factor interactions (b) is better based on confidence interval, but predicted value seems almost the same and are basically the same.

1. Use the wine data from set B11 in the Appendix..
2. Build a model using the wine quality as the response variable. Use flavor and region as predictors. Write out the specific model for each region.

• when x2 = 0 and x3 = 0, region = 1

• when x2 = 1 and x3 = 0, region = 2

• when x2 = 0 and x3 = 1, region = 3

Overall model regression is:

Quality = 7.094 + 1.116 Flavor – 1.533 x2 + 1.223 x3

Regression Equation

Region

1 Region 1 (x2 = 0 and x3 = 0):Quality = 7.094 + 1.116 Flavor

2 Region 2 (x2 = 1 and x3 = 0): Quality = 5.561 + 1.116 Flavor

3 Region 3 (x2 = 0 and x3= 1): Quality = 8.32 + 1.116 Flavor



1. Comment on the tests of significance of the model and of the variables in it.

The p-value for the regression is 0, so at least one of the predictors used in the model is significant. Each of the individual predictors (flavor, Region (x2 and x3) as was listed coded in option a) are also significant (p-value < .05).



1. Full Regression Equation

Quality = 6.73 + 1.199 Flavor + 0.0 Region\_1 - 2.89 Region\_2 + 3.38 Region\_3

+ 0.0 Flavor\*Region\_1 + 0.311 Flavor\*Region\_2 - 0.403 Flavor\*Region\_3

Regression Equation for each region:

Region

Region 1(x2 = 0 and x3 = 0): Quality = 6.73 + 1.199 Flavor

Region 2(x2 = 1 and x3 = 0): Quality = 3.84 + 1.509 Flavor

Region 3(x2 = 0 and x3= 1): Quality = 10.11 + 0.796 Flavor

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Regression 5 129.360 25.8720 32.56 0.000

Flavor 1 17.799 17.7990 22.40 0.000

Region 2 5.284 2.6418 3.32 0.049

Flavor\*Region 2 1.785 0.8923 1.12 0.338

Error 32 25.429 0.7946

Lack-of-Fit 23 20.164 0.8767 1.50 0.271

Pure Error 9 5.265 0.5850

Total 37 154.788

Model Summary

S R-sq R-sq(adj) R-sq(pred)

0.891427 83.57% 81.01% 77.68%

1. I would choose the model without the interaction terms because in the model with the interactions, both Flavor \* Region are not significant (p-value of 0.338). There is also not much of an improvement in the adjusted R2 as it increases from 80.87% to 81.01%. The standard deviation of the model is also almost the same, at 0.8914. It does not appear that adding the interaction terms improves the model by too much.
2. Use the “Wine Quality of Young Red Wines” data in table B19 to fit a linear model. Use predictors x2 through x10 to predict the quality rating y. Using so many predictors has likely led to some multicolinearity issues.
3. Regression Equation:

y = -5.2 + 6.22 x2 + 0.00494 x3 - 3.16 x4 + 6.52 x5 - 0.24 x6 - 15.27 x8 - 0.274 x9 + 22 x10

Coefficients

Term Coef SE Coef T-Value P-Value VIF

Constant -5.2 13.7 -0.38 0.708

x2 6.22 3.47 1.79 0.086 3.77

x3 0.00494 0.00773 0.64 0.529 3.47

x4 -3.16 2.17 -1.45 0.160 545.61

x5 6.52 4.09 1.60 0.124 778.64

x6 -0.24 3.21 -0.07 0.942 98.80

x8 -15.27 7.48 -2.04 0.053 6.75

x9 -0.274 0.178 -1.54 0.137 23.00

x10 22 128 0.17 0.867 112.97