Assignment 6

Gilbert Watson

Monday, November 4th, 2013

Contents

1 6.5

In a small-scale experimental study of the relation between degree of brand liking (Y) and moisture content (X_1) and sweetness (X_2) of the product, the following results were obtained from the experiment based on a completely randomized design (data are coded):

a) Obtain the scatter plot matrix and the correlation matrix. What information do these diagnostic aids provide here?

b) Fit regession model (6.1) to the data. State the estimated regression function. How is b_1 interpreted here?

c) Obtain the residuals and prepare a box plot of the residuals. What information does this plot provide?

d) Plot the residuals against Y, X_1, X_2 , and X_1 X_2 on separate graphs. Also prepare a normal probability plot. Interpret the plots and summarize your findings.

e) Conduct the Breusch-Pagan test for constancy of the error variance, assuming $log(\sigma^2) = \gamma_0 + \gamma_1 X_{i1} + \gamma_2 X_{i2}$: use $\alpha = .01$. State the alternaltives, decision rule, and conclusion.

1.1 Answer:

2 7.3

Refer to Brand Preference Problem 6.5:

a) Obtain the analysis of variance table that decomposes the regression sum of squares into extra sums of squares associated with X, and with X_2 , given X_1.

b) Test whether X_2 can be dropped from the regression model given that X_1 is retained. Use the F^* test statistic and level of significance .01. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

2.1 Answer:

3 7.12

Refer to Brand Preference Problem 6.5: Calculate R^2_{Y1}, R^2_{Y2}, $R^2_{|2}$, $R^2_{Y1|2}$, $R^2_{Y2|1}$, and R^2. Explain what each coefficient measures and interpret your results.

3.1 Answer:

4 7.24

Refer to Brand Preference Problem 6.5:

a) Fit first-order simple linear regression model (2.1) for relating brand liking (Y) to moisture content (X_1). State the fitted regression function.

b) Compare the estimated regression coefficient for moisture content obtained in part (a) with the corresponding coefficient 0btained in Problem 6.5b. What do you find?

c) Does $SSR(X_1)$ equal $SSR(X_1|X_2)$ here? If not, is the difference substantial?

d) Refer to the correlation matrix obtained in Problem 6.5a. What bearing does this have on your findings in parts (b) and (c)?

4.1 Answer:

5 7.28

a Define each of the following exta sums of squares: (I) $SSR(X_5X_1)$, (2) $SSR(X_3, X_4|X_1)$, (3) $SSR(X4|X_1, X_2, X_3)$.

b For a multiple regression model with five X variables, what is the relevant exta sum of squares for testing whether or not $\beta_5 = O$? whether or not $\beta_2 = \beta_4 = O$?

5.1 Answer:

6 7.29

Show that:

a) $SSR(X_1, X_2, X_3, X_4) = SSR(X_1) + SSR(X_2, X_3|X_1) + SSR(X4|X_1, X_2, X_3)$.

b) $SSR(X_1, X_2, X_3, X_4) = SSR(X_2, X_3) + SSR(X_1|X_2, X_3) + SSR(X4|X_1, X_2, X_3)$.

6.1 Answer:

7 7.30

Refer to Brand Preference Problem 6.5:

a) Regress Y on X_1 using simple linear regression model (2.1) and obtain the residuals.

b) Regress X_1 on X_2 using simple linear regression model (2.1) and obtain the residuals.

c) Calculate the coefficient of simple correlation between the two sets of residuals and show that it equals $r_{Y1|2}$.

7.1 Answer:

8 8.11

Refer to Brand Preference Problem 6.5:

a) Fit regression model (8.22).

b) Test whether or not the interaction term can be dropped from the model; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

8.1 Answer:

9 8.16

Refer to Grade point average Problem 1.19. An assistant to the director of admissions conjectured that the predictive power of the model could be improved by adding information on whether the student had chosen a major field of concentration at the time the application was submitted. Assume that regression model (8.33) is appropriate, where X_1 is entrance test score and $X_2 = 1$ if student had indicated a major field of concentration at the time of application and 0 if the major field was undecided.

a) Explain how each regression coefficient in model (8.33) is interpreted here.

b) Fit the regression model and state the estimated regression function.

c) Test whether the X_2 variable can be dropped from the regression model; use $\alpha = .01$. State the alternatives. decision rule. and conclusion.

d) Obtain the residuals for regression model (8.33) and plot them against $X1X_2$.Is there any evidence in your plot that it would be helpful to include an interaction term in the model?

9.1 Answer:

10 8.20

Refer to Grade point average Problems 1.19 and 8.16:

a) Fit regression model (8.49) and state the estimated regression function.

b) Test whether the interaction term can be dropped from the model; use $\alpha = .05$. State the alternatives. decision rule, and conclusion. If the interaction term cannot be dropped from the model, describe the nature of the interaction effect.

10.1 Answer:

11 8.42

Refer to Market share data set in Appendix C.3. Company executives want to be able to predict market share of their product (Y) based on merchandise price (X_1), the gross Nielsen rating pOints (X_1, an index of the amount of advertising exposure that the product received); the presence or absence of a wholesale pricing discount ($X_3 = 1$ if discount present: otherwise $X_3 = 0$); the presence or absence of a package promotion during the period ($X_4 = 1$ if promotion present: otherwise $X_4 = 0$): and year (X_5). Code year as a nominal level variable and use 2000 as the reference year.

a) Fit a first-order regression model. Plot the residuals against the fitted values. How well does the first-order model appear to fit the data?

b) Re-fit the model in part (a). After adding all second-order terms involving only the quantitative predictors. Test whether or not all quadratic and interaction terms can be dropped from the regression model: use $\alpha = .05$. State the alternatives. decision rule, and conclusion.

c) In part (a), test whether advertising index (X_2) and year (X_1) can be dropped from the model; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

11.1 Answer:

12 System Information

```
> sessionInfo();

R version 3.0.1 (2013-05-16)
Platform: x86_64-apple-darwin10.8.0 (64-bit)

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

loaded via a namespace (and not attached):
[1] tools_3.0.1
```