Stat 704 Homework 4

Carry out all hypothesis tests at the 5% significance level.

1. Consider the **brand preference** data of Problem 6.5.

   (a) Obtain and report the scatterplot matrix; what does it tell you about the relationship between 'liking' Y and each of the predictors x_1 'moisture' and x_2 'sweetness'?

   (b) Fit the regression model $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$. Report the ANOVA table and the table of regression effects.

      i. Using the p-value from the F-statistic in the ANOVA table, test $H_0 : \beta_1 = \beta_2 = 0$. What does this imply about β_1 and β_2?

      ii. Report each of b_1 and b_2 along with tests of $H_0 : \beta_1 = 0$ and $\beta_2 = 0$. Can either predictor be dropped in the presence of the other?

      iii. Intepret both estimated coefficients.

   (c) Obtain residual plots of e_i vs. $\hat{Y}_i$, e_i vs. x_{i1}, and e_i vs. x_{i2}. Obtain the normal probability plot and a histogram of the residuals. What do these plots tell you?

   (d) Use SAS to conduct the Breusch-Pagan test of $H_0 : \alpha_1 = \alpha_2 = 0$ in the variance model $\sigma_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2}$.

   (e) Report R^2; how is it interpreted here?

   (f) Obtain *and interpret* an 95% interval estimate of $E\{Y_h\}$ when $x_{h1} = 5$ and $x_{h2} = 4$.

   (g) Obtain *and interpret* an 95% prediction interval for a new Y_h when $x_{h1} = 5$ and $x_{h2} = 4$.

   (h) Obtain *and interpret* $\text{SSR}(x_1|x_2)$ and $\text{SSR}(x_2|x_1)$.

   (i) Obtain $\text{SSR}(x_1)$, $\text{SSR}(x_2|x_1)$, and verify $\text{SSR}(x_1, x_2)=\text{SSR}(x_1)+\text{SSR}(x_2|x_1)$.

   (j) Obtain *and interpret* $R^2_{Y1|2}$ and $R^2_{Y2|1}$.

   (k) Obtain *and interpret* the variance inflation factors VIF_1 and VIF_2.

2. Consider the **commercial properties** data of Problem 6.18.

   (a) Obtain and report the scatterplot matrix; what does it tell you about the relationship between 'rental rate' Y and each of the predictors x_1 'age', x_2 'operating expense', x_3 'vacancy', and 'square footage'?

   (b) Fit the regression model $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + +\beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i$. Report the ANOVA table and the table of regression effects.

      i. Using the p-value from the F-statistic in the ANOVA table, test $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$. What does this imply about β_1, β_2, β_3, and β_4?

      ii. Report each of b_1, b_2, b_3, b_4 along with tests of $H_0 : \beta_j = 0$ for $j = 1, 2, 3, 4$. Can any predictor be dropped in the presence of the other three?

      iii. Intepret all four estimated coefficients.

   (c) Obtain residual plots of e_i vs. $\hat{Y}_i$, e_i vs. x_{i1}, e_i vs. x_{i2}, e_i vs. x_{i3}, and e_i vs. x_{i4}. Obtain the normal probability plot and a histogram of the residuals. What do these plots tell you?

   (d) Use SAS to conduct the Breusch-Pagan test of $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ in the variance model $\sigma_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \alpha_4 x_{i4}$.

   (e) Report R^2; how is it interpreted here?

   (f) Problem 6.20. You don't need to find a family of intervals, just compute the 95% interval for each of the four. You can compute the family for extra credit. The easiest way is Bonferroni (p. 159), where you would use `alpha=0.0125`.

   (g) Problem 6.21. Ignore the "family confidence interval" part; just find three 95% prediction intervals.

   (h) Obtain *and interpret* $\text{SSR}(x_1)$, $\text{SSR}(x_2|x_1)$, $\text{SSR}(x_3|x_1, x_2)$, $\text{SSR}(x_4|x_1, x_2, x_3)$.

   (i) Verify that the above extra sums of squares in (h) sum to $\text{SSR}(x_1, x_2, x_3, x_4)$.

   (j) Obtain *and interpret* $R^2_{Y1|234}$, $R^2_{Y2|134}$, $R^2_{Y3|124}$, $R^2_{Y4|123}$.

   (k) Obtain *and interpret* the variance inflation factors VIF_j for $j = 1, 2, 3, 4$.

3. The following textbook problems: 6.1, 6.22, 7.1 (df for the associated test), 7.2 (what is x_{i1} added on to?), 7.28, 7.29a (use the definition).