Stat 501 Nov. 10 LAB Turn in by Nov. 12

For parts 1-5, use the dataset marketshare.mtw at www.stat.psu.edu/~rho/501data/. The dataset gives information about the market share (of sales) for a food product over n = 36 months. Y = market share, x1 = price of the product, x2 = Nielsen rating of advertising exposure for product, x3 = 1 if discount price promotion was in effect and 0 otherwise, x4 = 1 if package promotion was in effect and 0 otherwise. Also the dataset includes the six possible multiplicative interactions between pairs of x variables.

1. Use Stat>Regression>Best Subsets to identify potentially good models for predicting Y = market share. All x-variables and interactions are possible predictors. Based on the results, what variables are in the model you think might be the best? Why do you think this might be the best model?

2. A forward stepwise procedure identifies a model by first picking the strongest predictor, then adding the next strongest predictor given the first x-variable is in the model, and so on until variables can't be added due to lack of statistical significance. There's no guarantee the procedure stops at the best model, so these days stepwise procedures aren't used as often as the best subsets procedures.
Use Stat>Regression>Stepwise to carry out a stepwise procedure. What variables are in the final model (the last column of output)? Does this agree with what you found in part 1?

3. Do a multiple regression using the predictors you think are in the best model. Store the residuals and fits. Write the estimated model.

4. Using Graph>Scatterplot, With Groups, plot Fits versus x1 = Price using x3 and x4 as grouping variables. What is indicated about how x1= price, x3 = discount pricing, and x4 = package promotion affect estimated market share (Y)?

5. Using Graph>Scatterplot, With Groups, plot Residuals versus Fits and use x3 and x4 as grouping variables. Briefly interpret the result.

For parts 6-11 , use the dataset bodyfat.mtw at www.stat.psu.edu/~rho/501data/. Y = measure of body fat, x1 = triceps skinfold measurement, x2 = thigh circumference, x3 = midarm circumference

6. Do a simple regression using x2 = thigh to predict y. (a) What is the estimated slope? (b) What is the standard error of this estimated slope? (c) Is the linear relationship statistically significant?

7. Do a multiple regression using all three x-variables to predict y. (a) What is the estimated coefficient multiplying x2=thigh? (b) What is the standard error of this estimated coefficient? (c) Compare the standard error here to the standard error found in part 6. (d) Is x2= thigh statistically significant within this multiple regression?

8. Do a multiple regression in which you predict x2 = thigh (response variable for this) using the other two x-variables as predictors. **Store the residuals.** What is the value of R^2 for this regression? What is indicated about the x-variables?

9. Do a multiple regression in which you predict y = bodyfat using x1 = triceps and x3= midarm as predictors. **Store the residuals**. Then, do a simple regression using the residuals from this part as the response variable and the residuals from the previous part (part 8) as the predictor variable. (a) What is the slope? (b) Compare this value to the estimated coefficient found in part 7a.

10. Again use all three x-variables to predict y in a multiple regression. Use Options and select Display: Variance Inflation Factors. What is the VIF (variance inflation factor) reported on the output for x2 = thigh?

11. Refer back to part 8 in which we found the R^2 for the relationship between x2 = thigh and the other two x-variables. Using the R^2 found there, calculate $\frac{1}{1-R^2}$ and compare this to the value found in part 10.

Interpretation: A variance inflation factor measures how correlation among the x-variables affects the standard error (or variance = squared standard error) of an estimated coefficient in a multiple regression. In the presence of a high VIF (book suggests high is > 10) we have imprecise estimates of the coefficients.