Regression Assessment

The SPSS file for this assessment is catalog.sav. This data file contains hypothetical monthly sales figures for three products sold by a catalog company: women’s clothing, men’s clothing, and jewelry. Data for five possible predictor variables are the number of catalogs mailed, the number of pages in the catalog, the number of phone lines open to receive orders, the amount spent on print advertising, and the number of customer service representatives available to take calls.

1. Looking at the histograms for the three products, what do you conclude about the nature of the three distributions?
2. Using the sales of women’s clothing as the dependent variable, create a regression equation with all five predictors included in the equation. What is the total proportion of variance in the sale of women’s clothing that is explained by all five predictors?
3. What is the best estimate of the proportion of variance in the sales of women’s clothing that these five predictors explain in the population?
4. Which of the five predictor variables appears irrelevant in this equation?
5. Now, build an equation with the four best predictors. What is the proportion of variance explained in the sales of women’s clothing by these four independent variables?
6. What is the best estimate of r-squared in the population for the equation in question 5?
7. What is the coefficient (slope) for number of pages in the catalog?
8. What is the predicted value for the sales of women’s clothing using the following values for the independent variables (round to two decimal places):  
   Number of Catalogs Mailed: 11000  
   Number of Pages in the Catalog: 100  
   Amount Spent on Print Advertising: 30000  
   Number of Customer Service Representatives: 50
9. What is the approximate 95% confidence interval for the predicted value in question 8 rounded to the nearest whole number? The predicted value plus or minus \_\_\_\_\_\_\_\_\_\_\_\_.
10. Does there appear to be a first-order auto correlation?
11. Looking at the plots of the standardized residuals and the Normal P-P Plot, what do you conclude about the assumption of normal residuals?
12. Do a stepwise regression analysis using all five independent variables and the natural log of the sales of women’s clothing at the dependent variable. What is the r-squared for the equation this stepwise analysis produces?
13. What impact did the log transformation have on the nature of the residual plot?
14. Which of the five independent variables is a non-significant predictor when the log of women’s sales is used as the dependent variable?
15. Turning to the analysis of the sale of men’s clothing, which variable is non-significant in the stepwise prediction of the sale of men’s clothing using the sale of men’s clothing as the dependent variable?
16. In the analysis of the sale of men’s clothing using a stepwise regression analysis using all five independent variables as possible predictors, what is the proportion of variance explained in the sale of men’s clothing by the number of phone lines open for ordering over and above the proportion of variance explained by the number of catalogs mailed?
17. In the equation generated in question 15, what do you conclude about the residuals from this model?
18. Perform both a forward and backward equation building technique. Do these procedures result in the same prediction model? (Yes/No Answer)
19. Continuing with the prediction equation for the sale of men’s clothing, what is the largest deviation of a predicted value from the actual value of the sale of men’s clothing (round to two decimal places)?
20. Comparing the prediction models for men and women, what is a possible explanation for the difference in the two models?