STA 5207 Assignment 11

Due Friday December 3

You do not need to turn in any SAS output.

1. (25 points total) Sixteen batches of plastic were made, and from each batch one test item was molded. Each test item was randomly assigned to one of the four predetermined time levels, and hardness was measured after the assigned elapsed time. Use the data file “plastic.txt.” Notice that it gives *y* (hardness in Brinelle units) first for each observation, followed by *x* (elapsed time in hours).
2. (10 points) Perform the Lack of Fit test, giving the hypotheses, test statistic, P-value, and conclusion.
   1. 𝐻0: 𝐸(𝑦) = 𝛽0 + 𝛽1𝑥 vs 𝐻1: 𝐸(𝑦) ≠ 𝛽0 + 𝛽1x
   2. *FLF =* 0.82369 P-value = 0.4622,
   3. Fail to reject H0­, so cannot conclude that the model doesn’t fit the data well.
3. (10 points) Based on the results of the Lack of Fit test, would it be appropriate to perform simple linear regression on this data set? Why or why not?
   1. It would be appropriate as we failed to reject the hypothesis that the hardness changes linearly over time. Additionally, when we do a scatter plot for the data, it does appear that the relationship is linearly increasing.
4. (5 points) Which would be a more appropriate estimate for σ2 in this situation: MSE or MSPE?
   1. MSE as we failed to reject H0
   2. MSE = 10.45893
5. (35 points total) A chemist is interested in the concentration of a solution (*y*) over time (*x*). Fifteen identical solutions were prepared. They were randomly divided into five sets of three and were measured after 1, 3, 5, 7, and 9 hours respectively. The data is given in *solution.txt*. We saw earlier that the relationship between solution and time is not linear.
6. (5 points) Give the quadratic model using the centered predictor.
   1. E[y] = *β*0 + *β*1X + *β*11X2, where X = x - x̄
7. (10 points) Obtain the fitted equation for the quadratic model using the centered predictor. No output needed.
8. (20 points) Test for the quadratic term, giving the hypotheses, test statistic, P-value, and conclusion. Test for the linear term if necessary. No output needed.
   1. 𝐻0: 𝛽11 = 0 vs 𝐻1: 𝛽11 ≠ 0
   2. F = 78.248 with P-value < 0.0001
   3. Reject H0, keep both the quadratic and linear terms in the model.
9. (40 points total) A commercial real estate agency evaluates vacancy rates, square footage, rental rates, and operating expenses for commercial properties in order to provide information to clients. The data is contained in *properties.txt* and contains headings. The data set in SAS studio does not contain headings. The variables, given in order, are:

y: Rental rate

x1: Age

x2: Operating expenses

x3: Vacancy rate

x4: Square footage

**Only use X1, X2, and X4 in the model**. If using SAS, use property as the filename since properties is too long.

1. (10 points) Give the full model with all possible interactions.
   1. *E*[*y*]=𝛽0 + 𝛽1X1 + 𝛽2X2 + 𝛽3X3 + 𝛽12X1X2 + 𝛽13X1X3 + 𝛽23X2X3 + 𝛽123 X1X2X3
2. (30 points) Test for significant interactions, following the hierarchy of testing. Give the hypotheses, test statistics, and conclusion. Use centered predictors.
   1. Three-way interactions:
      1. H0: β123 = 0 vs H1: β123 ≠ 0
      2. F-test = 16.046 P-value < 0.001
      3. Reject H0. Stop and use the full model