Instructions: Read and then answer all questions carefully and completely for full/partial credit. Each part should have a conclusion or final statement. For any testing problems be sure to explicitly list the null and alternative hypotheses, test statistic, rejection region, decision and conclusion of the test. Good Luck! and Have a Great Thanksgiving!

1. The technique of multiple linear regression is one of the most frequently used statistical tools in practice, with many varying fields of application. The following is an application of multiple regression in geography. P. J. Taylor (Geography, 1980) chose to investigate the variation in average annual precipitation in California.

Data on the average annual precipitation (Y), altitude (X1), latitude (X2), and distance from the Pacific coast (X3) were obtained for 30 meteorological stations scattered throughout the state. In addition, whether the station was a "westward-facing" or "leeward-facing" station was also recorded. This variable, X4, was defined to be equal to 1 (one) if the station was 'leeward' and 0 (zero) if the station was "westward".

It was thought that the interrelation between annual rainfall and the independent variables may also include curvature and interaction. Thus the following four models were fit:

M1:

\[ E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \]

First Order

M2:

\[ E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{14} X_1 X_4 + \beta_{23} X_2 X_3 + \beta_{24} X_2 X_4 + \beta_{34} X_3 X_4 \]

Interaction

M3:

\[ E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{14} X_1 X_4 + \beta_{24} X_2 X_4 + \beta_{33} X_3^2 + \beta_{33} X_3^2 \]

Subject Matter

M4:

\[ E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{14} X_1 X_4 + \beta_{23} X_2 X_3 + \beta_{24} X_2 X_4 + \beta_{34} X_3 X_4 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2 \]

Complete Second Order
a) Concerning the models M1, M2, and M4, give a concise description of what each model is implicitly assuming about the relationship between Y and the X's. (Be brief but as complete as possible. You should give summary sketches of the relationship).

Answer the following questions concerning M1: First Order

b) Report the least squares equation and test its overall usefulness. Use $\alpha = 0.01$.

c) Of particular interest in this case is the effect of the variable X4. Investigate this effect by building a 95% confidence interval for $\beta_4$ and by giving a careful interpretation of the interval. The researchers are particularly interested in specifying and quantifying the role that X4 plays in this model.
d) LA (Los Angeles) has an altitude of 312 ft., a latitude of 34.1 degrees, is 16 miles from the Pacific coast (the location of the station) and is obviously "westward-facing". Use the Minitab output to find a 99% interval for the mean annual rainfall for LA.

e) The residual plots for a full residual analysis were obtained for the fit of Model 1: First Order Model. These should help, in part, in developing a Subject Matter Model. Make an assessment of these residual plots. What model assumption(s) may be violated? For each of the possible assumptions listed below mention what an ideal plot would look like, what a plot that indicates a problem, and mention a specific remedy that might alleviate the corresponding problem.

i) Normality

ii) Model Lack of Fit

iii) Nonconstant Variance
The remaining questions might concern any of the mentioned models.

f) Interactions play important roles in many models. It is obvious (why?) that in M3: Subject Matter Model that X1 and X2 interact. Define what it means for X1 and X2 to interact in their effect on Y. In the process, analyze the interaction profile plots given below.
g) As mentioned in class the sequence of model is useful in testing certain aspects of our process, i.e., whether there is interaction or curvature present. Please assess these issues by completing the following tests: (Use $\alpha = 0.01$)

i) Formally test whether Model 2: Interaction is better than Model 1: First Order.

ii) Formally test whether Model 4: Complete Second Order is better than the model chosen as best in part (i) above.
h) Please note that within the output of the Model 4: Complete Second Order there is a warning associated with the results of the predict statement: **X denotes a row with X values away from the center.** Explain what this message means.

i) Since the Subject Matter Model doesn't naturally fit between the (First order and the Interaction) or (Interaction and the Complete Second Order) models, we can't formally test whether the Model 2: Interaction is better than the Model 3: Subject Matter Model or whether the Model 3: Subject Matter Model is better than the Model 2: Interaction. However, we can use our Ad Hoc method of comparison. To do so fill in the following chart. For each of the statistical criterion mentioned, choose the model that is superior and then use the results to "see" if our Subject Matter model is superior.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s$</td>
<td>_______</td>
<td>_______</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>$R_a^2$</td>
<td>_______</td>
<td>_______</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>$F$</td>
<td>_______</td>
<td>_______</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>$S_y$</td>
<td>_______</td>
<td>_______</td>
<td>_______</td>
<td>_______</td>
</tr>
</tbody>
</table>

j) Based on your conclusion to part (h), re-evaluate the mean annual rainfall LA can expect to receive. (Note that all other models except for the first order model already has 99% intervals listed, so you may quote those as is, along with a careful explanation.)