STATISTICS 480: Project 2, Spring 2003
Statistics Software Reliability

The goal of this project is to assess the reliability of Minitab, JMP, and Excel with respect to linear regression calculations using a subset of the NIST Statistical Reference Data sets (StRD sets). Your final report will take the form of a web page (an HTML file with accompanying images) as well as a printed document.

1. Download the Linear Regression data sets: Norris, NoInt1, Longley, Wampler3. These are found at the NIST StRD web site:

   `http://www.nist.gov/itl/div898/strd/`

2. For the Norris, NoInt1, and Wampler3 data sets, obtain a scatterplot of y vs. x. For the Longley data set, obtain a matrix of scatterplots for all pairs of variables. You may use any one of the three software packages to do this part.

3. For each data set and for each program, obtain the software values for

   (a) the parameter estimates (Norris has 2 parameters, NoInt1 has 1, Longley has 7, and Wampler3 has 6),

   (b) the standard deviations of the parameter estimates

   (c) the Residual standard deviation, sometimes called the Root Mean Square Error

   (d) R^2

   For each software value, attempt to obtain as many significant digits as NIST reports for the certified values (15 significant digits). Note that this may require more than 15 decimal places in some cases since leading zeros (even after the decimal place) are not counted as significant digits. (Note: This is fairly easy to do in Excel and JMP but somewhat of a challenge in Minitab.)

4. Create a table which contains all the certified values, software values, and LRE values. Norris will require 6 LREs, NoInt1 will require 4, Longley will require 16, and Wampler3 will require 14 *for each program*—this project will require a total of 120 LRE calculations.

5. Numerically summarize your findings by creating a table with 3 entries for each data set and program. Let λ_β be the minimum of the parameter estimate LREs, λ_σ be the minimum of the standard deviation LREs, and λ_{R^2} is the LRE for R^2.

6. Write your report using Microsoft Word. Write the report as if it were to be read by someone who is only vaguely familiar with the NIST data sets. Be sure to briefly describe the purpose of your study, the data sets used (include the plots from step 2) and the models fit. Also, describe, in as much detail as needed, how the LRE-values are calculated and how they can be interpreted. Then present your tables from steps 4 and 5; indicate how you obtained the values in the table from step 5 from the values in the table from step 4. In your conclusion, summarize your findings *in words* and indicate your recommendations as far as software reliability for linear regression in these three packages goes. Also, indicate any adverse conditions that may be affecting the accuracy of your LRE-values themselves.

7. Once you are satisfied with your report save it as a Word file. Then ***first***—print it to turn in later (NO dot matrix printouts will be accepted); ***second***—use Word's Save As HTML menu command to save your report as a web page. When you save your report as an HTML file, you should name the file as follows:

   *lastname*proj2.html

   (replace *lastname* with your last name; for example, duckworthproj2.html)

   This will result in a file being created called *lastname*proj2.html as well as several image files (one for each plot in your report) usually named something like Image1.gif, Image2.gif, etc. All of these files together constitute your web page. Be sure to save your report as a Word file too—use the same name as your HTML file but use .doc rather than .html (for example, duckworthproj2.doc).

8. If you haven't already, you will need to register to have a personal home page on VINCENT. Instructions can be found at

   `http://www.public.iastate.edu/info/setup.html`

   Basically, you have to login to your VINCENT/Acropolis account (say, in 322 Snedecor for example) and issue to commands at the command line:

   ```
   add www
   setup_www
   ```

   but you can check the details at the above web page.

9. After completing step 8, FTP your .doc file, your .html file, and your image files to your WWW subdirectory.

10. Using a web browser, double check that your web page is viewable. The URL for your page should be

    `http://www.public.iastate.edu/~yourusername/lastnameproj2.html`

    (be sure to replace `yourusername` with your VINCENT user name and replace `lastname` with your last name)

11. Once you are satisfied with your web page, email me the URL for it.

Special considerations as you work on this project:

- Be sure you are actually fitting the model indicated by the NIST web pages. Norris is a simple linear model, NoInt1 is a model without an intercept/constant term, Longley is a multiple regression model, and Wampler3 is a high degree polynomial model.
- You must email me the URL for your web page by 5:00 PM, Friday, May, 2, 2003. You must also turn in a printed copy of the Word version of your report by this same deadline.
- At the time of the final exam, Wednesday, May 8, 9:45 to 11:45, you can pick up a critique of your project and find out your course grade.