Statistics Project—Parameter Estimation

The Question: Standard statistical procedures like Least-Squares Regression are intended to divine the underlying values for certain parameters in the presence of noise. This project addresses the question of how well those procedures do in the presence of varying levels of noise.

Background: Many statistical procedures are based on the premise that there is an underlying behavior described by $y = g(x)$ where $g(x)$ is some relationship described by a few values called "parameters". In the case of linear regression $g(x) = a+bx$ and the parameters are a and b. In the usual situation a and b are unknown. What is known are the measurements (x_i,y_i) which are assumed to be related by $y_i = g(x_i) + \varepsilon_i$ where ε_i is some unknown random value we can call noise. The statistical procedure then attempts to determine $g(x)$ by estimating the parameters that describe $g(x)$. In the linear case this means to estimate the a and b in $g(x) = a+bx$.

Basic Test: This project tests the ability of the statistical procedure to estimate the parameters in the presence of different noise levels. Each group of trails begins with a specification of the parameter values and the noise level. For each individual trial we use randomly generated ε_i values to generate a set of measurements (x_i,y_i). Now apply the standard statistical procedure and measure how closely it estimates the parameters we started with.

Study Design: To setup your study you will need to determine the following—

1. Type of procedure being analyzed, e.g. linear regression, quadratic regression, etc.
2. Values of the parameters being considered
3. How to describe the noise levels
4. What noise levels to test
5. How many trails at each setting to use
6. What statistics to use to describe the results
7. What explanatory values to use

Getting Started: To understand the Basic Test make an Excel sheet that takes a and b as input, generates (x_i,y_i) using $g(x) = a+bx$ and the rand() function for ε_i, the Excel sheet then estimates a and b using the intercept() and slope() functions. Run several trials (F9 will recalculate all the ε_i values) and observe the quality of the parameter estimates.

First Meeting: Prior to the meeting do the "Getting Started" exercise and print out three runs. Explain the output to your instructor. Determine the following for your project--which parameters, how to measure noise levels, how to use Excel for your trials, how to choose the parameter test values, how to choose the explanatory variable values, and what could go wrong. Write a summary of this meeting and turn it in to the instructor.

Construct the Instrument: Based on the decisions from the First Meeting make an Excel sheet that can be used to carry out the trials. Experiment with this Excel sheet until you have some intuition as to how this experiment is going to go.

Second Meeting: Present your instrument to the instructor and describe your expectations as to how these trials are going to go. Determine what parameter values to test, what noise levels to test, how many trials at each setting, what measure(s) of accuracy to use, and how to tabulate the test data. Write a summary of this meeting and turn it in to the instructor.

Collect Data and Compute Statistics: Do the stuff specified at the Second Meeting. You should end up with a table of test results and a table of statistics based on the test results.

Third Meeting: Present your results. If there are problems with the testing procedures resolve them and run the tests again. Discuss what conclusions are justified. Discuss every section of the project report and what should be there.

Write Project Report Draft: Do the stuff specified at the Third Meeting.

Fourth Meeting: Present your draft report to the instructor. Use your instructor's critique to write the final report.

Write the Project Report.

Additional Project Guidelines:
Due Dates
Report Format
Report Writing Cautions.