Tutorial 4: Fitting of Experimental Data - Single and Multiple Regression

Overview

We often do experiments to determine the relationships among variables. These relationships may follow a known form, for example the voltage and current in a resistor, or an unknown form, for example the strength and stiffness of wood. Invariably, measurement errors will be present in practical measurements, even despite our good efforts. Since our measurements can never be "exact", we must develop means of extracting the best results we can. A good strategy is to make multiple measurements over a wide range of data values so as to benefit from the effect of averaging and to see the character of the data response. Is the response linear, is it noisy, does it depend on one or more sources ?

These questions can be addressed using regression techniques, where we seek to find models of the data that give a "best-fit" in some sense. The most reliable of such models are those that are based on physics. The resulting numerical analysis is then a matter of finding the associated "calibration constants". In other cases, no direct model is available, so an empirical model based on the character of the observed data needs to be created. This should be done with care; the model should be kept simple and practical, and it should incorporate whatever physical intuition that may be available. This tutorial describes some examples of the creation and use of single and multiple regressions using the least-squares technique.

Objectives:

In this tutorial you will explore the statistical character of some real data that contain significant noise, in this case mostly due to the effects of other quantities not part of the measurements. You will get some experience in handling statistical data and in making appropriate decisions.

Pre-tutorial questions

The assigned questions will not be marked, but the tutorials are intended as a place to discuss the questions that you found challenging. These questions are intended to exercise skills needed for the tutorials and exams.

Textbook questions: **8-7, 8-15, 8-25**.

Course question:

When doing their Capstone experiment on the natural frequencies of a ski, some students made the following vibration amplitude vs. forcing frequency measurements near a particular natural frequency:

Frequency, Hz	33.2	33.4	33.6	33.8
Amplitude, mm	1.3	1.5	1.4	1.2

The students wish to identify the ski natural frequency by determining the forcing frequency at which the maximum vibration amplitude occurs. To get better accuracy, they propose to fit a least-squares parabola among the data and then find the peak of the parabola. You are asked to do the required calculation.

You are first asked to proceed algebraically and to insert the numbers only at the end. Use normalized variables $x_i = (f_i - f_{mean}) / (0.5\ \Delta f)$, where $\Delta f = 0.2$Hz. y_i = amplitudes. In this normalized format, $x_i = [-3, -1, 1, 3]$).

When finished, use Matlab to compare with your answer.

Tutorial Assignment

You should also start the tutorial assignment before the tutorial period so that you can identify the difficult points before class and therefore have more focused questions during the tutorial.

Data file DK.csv contains strength, stiffness and density data for 131 2x4 southern yellow pine boards. The objective of the measurements was to develop and test a method for predicting the strength of lumber based on non-destructive measurements, here the elastic modulus and the density. The following quantities were measured:

1. Modulus of Rupture (MOR) = wood strength measured in a bending test with the wood loaded in the "joist" direction, $x10^6$ psi.
2. Ultimate Compressive Strength (UCS), $x10^6$ psi.
3. Modulus of Elasticity (MOEj) = wood bending stiffness measured by a laboratory machine with the wood loaded in the "joist" direction, $x10^3$ psi.
4. Continuous Lumber Tester (CLTa) = average wood bending stiffness measured by a sawmill machine with the wood loaded in the "plank" direction, $x10^3$ psi.
5. Continuous Lumber Tester (CLTm) = minimum wood bending stiffness measured by a sawmill machine with the wood loaded in the "plank" direction, $x10^3$ psi.
6. Density (dens) = wood density, $x1000kg/m^3$.

You are asked to explore the measurements 3-6 as possible non-destructive predictors of wood strength MOR. Prepare scatter plots to visualize the data. Do both single and multiple regressions using various regressors and suggest suitable wood strength predictors. Make scatter plots of measured vs. predicted strength. Describe your observations, identify relevant statistical quantities and explain your reasoning. Note that MOEj is a laboratory measurement, so is not suitable for industrial use.

Some starter Matlab code:

```
clear all
% DK wood data regression

% read all data
indata = xlsread('DK.csv');
num  = indata(:,1);
MOR  = indata(:,2);
UCS  = indata(:,3);
MOEj = indata(:,4);
CLTa = indata(:,5);
CLTm = indata(:,6);
dens = indata(:,7);

% regress MOR with MOEj
X = [ones(size(MOEj)) MOEj];
Y = MOR;
[b,bint,r,rint,stats] = regress(Y,X);
r2 = stats(1);
sd = sqrt(stats(4));
bstd = (bint(:,2)-bint(:,1)) / 3.92;
```