Matlab Homework #1
ChE 231
Spring 2019

You are an engineer in the materials science division of a rocket propulsion company. The lead scientist of a contracting company has recently been testing multiple grinding parameters to determine their effects on the performance of sintered silicon nitride, a material commonly used in the thrusters of various small satellites and space probes due to its ability to survive high thermal gradients and thermal shock. However, the scientist requires you to analyze the data he's collected so far in lab. The response variable measured is the mean of the ceramic strength, where the number of observations is 32 which comprises a complete 2^5 factorial design with 2 levels. The 5 variables manipulated are as follows,

- Factor 1 = Table Speed - (slow (.025 m/s) and fast (.125 m/s))
- Factor 2 = Feed Rate - (slow (.05 mm) and fast (.125 mm))
- Factor 3 = Wheel Grit - (140/170 and 80/100)
- Factor 4 = Direction - (longitudinal and transverse)
- Factor 5 = Batch - (1 and 2)

No centerpoints were included as the effects of factors 1-3 were expected to produce monotone effects, in addition, direction and batch are variables qualitative in nature. You are tasked with performing statistical analyses of these data including the development of a regression model, as well as identifying and quantifying the contributions of each variable on ceramic strength. Text in different font and encapsulated within parentheses, such as (`mean`) for example, are functions within MATLAB that have supporting documentation which can be accessed by entering `help mean` within the command window and by clicking on `Reference page for mean`.

1. Consider the dataset collected for high performance ceramics in randomized order within `data.p`, which has the following usage,

```
[X, Y, Z] = data()
```

   where X is an $n \times 5$ experimental design matrix, Y is an $n \times 1$ matrix of the output variable ceramic strength, with each row corresponding to a single run, and Z is an $n \times 1$ matrix corresponding to the order in which the experiments were run.

   (a) Generate a run order plot of the data versus the sample number in which the data was collected (`plot`). By visual inspection, is time a variable consideration? In other words, is the ceramic strength invariant with respect to time?

   (b) Generate a normal (Gaussian) probability plot (`normplot`). Looking at the graph, is the plot linear? Assess whether the data could have come from a normal distribution and describe in detail how a normal probability plot can be used to determine the type of distribution a data comes from.

   (c) Calculate the sample mean of the ceramic strength $\hat{\mu}$.

   (d) Calculate the sample standard deviation of the ceramic strength $\hat{\sigma}$.

   (e) Generate a histogram with 7 bins (`histogram`). Describe what information this type of plot generates and what you observe.

(f) Calculate the sample variance of the ceramic strength $\hat{\sigma}^2$.

(g) Perform a mean test (t-test) to test if the data comes from a normal distribution with mean $\hat{\mu}$ (`ttest`). Use $\alpha = 0.05$. Can the hypothesis that the data comes from a normal distribution with mean $\hat{\mu}$ be rejected at the 3% significance level? Report a 95% confidence interval on the mean.

(h) Perform a variance test to test if the data comes from a normal distribution with variance $\hat{\sigma}^2$ (`vartest`). Use $\alpha = 0.05$. Can the hypothesis that the data comes from a normal distribution with variance $\hat{\sigma}^2$ be rejected at the 3% significance level? Report a 95% confidence interval on the variance.

(i) Perform maximum likelihood estimation to estimate the mean, $\hat{\mu}$, and standard deviation, $\hat{\sigma}$, of a normal distribution data (`normfit`). Report the estimated mean and variance along with 95% confidence intervals for both. Compare these values to parts d & h.

2. For this experimental design, the factors affecting ceramic strength have certain disturbances on one another. This can be written in the following form of a regression model with interactions, which describes factors and their interactions with one another, in addition to quantifying the relative contribution of each variable and their interactions to the response variable, namely ceramic strength.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{12} x_1 x_2 \dots \tag{1}$$

(a) Typically, four-factor and and five-factor interaction terms are not significant and can be eliminated. Thus, we will perform an ANOVA analysis with up to three-factor interactions and type 3 sum of squares using the following commands,

```
varnames = 'Table Speed', 'Feed Rate', 'Wheel Grit', ...  'Direction', 'Batch';
[anova,tbl] =anovan(strength,factors,3,3,varnames);
```

(b) Assess the significance of each factor and their interactions by first extracting the p-values from the `anovan` analysis (`cell2mat`).

(c) Using a p-value < 0.05 as statistically significant, which interactions can be neglected? How many terms can the model be reduced to now? Report the new reduced design matrix, which should have dimensions $m \times 5$. Report also the corresponding ceramic strength values, which should also have dimensions $m \times 1$.

(d) Perform ANOVA analysis one more time on your new design matrix and

3. Given the new design matrix found previously, we can input a guess matrix of beta coefficients, and then perform regression to identify each beta coefficient.

(a) Create a separate function called `modelFun` with input parameters `beta` and `x` and output `y`. In effect, `modelFun` should be setup in such a way that it can be called to solve for equation 1.

(b) Perform a regression fit using `beta0 = [1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1]` as a guess matrix to identify bias terms using your `modelFun` function (`nlinfit`). For `nlinfit`, use the new design matrix and corresponding ceramic strength values identified from part 2d as input parameters, and use the `@modelFun` handle to reference the model equation function from part 3a. This command should have the following form,

```
betaValues = nlinfit(newX,newY,@modelFun,beta0)
```

(c) Export the fitted parameters, β, and report their values. Each column of the matrix `beta` corresponds to one of the four outputs. The first row of `beta` contains the bias terms, β_0, and the remaining rows contain the nonlinear regression coefficients, $\beta_1, \dots, \beta_{12}$ for the ceramic strength output. Write the nonlinear model equation that relates the ceramic strength output to the 4 inputs $x_1, \dots, x_5$ using the fitted parameter values.

(d) Comment on the effect of each predictor $x_1, \dots, x_5$ (inputs) on the response y (output). Recall what each of the inputs and outputs physically represents as you discuss the results.

(e) The response surfaces can be plotted to analyze the fit and response of the model developed and beta values derived using the following command,

```
nlintool(X,Y,@modelFun,betaValues,0.01)
```

(f) Adjust the nominal predictor values by either clicking on the graphs or entering new numbers in the boxes. Suggest a new set of inputs based on your model that increases the ceramic strength as high as possible. Take a screenshot and post this into your solution.

4. Draw some conclusions from your statistical analysis and model fitting. Comment on the factors on ceramic strength. Importantly, this experimental design has only 2 levels. What does this mean in terms of the predictor values you can actually choose? Are the inputs that you selected for 3f realistic? With this in mind, what values can your inputs really be and what is the maximum ceramic strength you can achieve by adjusting the inputs? What does this physically represent and what recommendations would you make to the scientist?