Exercise 9 – ANOVA

An alloy is being prepared with three different catalysts: A, B, and C. The concentration of the major element is determined for each catalyst in several samples. The data for each catalyst can be generated by:

A=transpose([42.7 45.6 43.1 41.6])
B=transpose([44.9 48.3 46.2])
C=transpose([41.9 44.2 40.5 43.7 41.0])

Depict the data graphically to compare the effects of the different catalysts and to visualize the dispersion:

plot(A,1,'o',B,2,'o',C,3,'o')

This plots the concentrations as dots at heights 1, 2, and 3 for the catalysts A,B,C, respectively.

The hypotheses that we want to test are:

H_0: all catalysts give statistically the same result
H_1: at least one catalyst gives a different result.

Calculate the means and variances:

mu1=mean(A)
mu2=mean(B)
mu3=mean(C)
s1=std(A)
s2=std(B)
s3=std(C)

Create variables corresponding to the number of data points for each catalyst:
        n1=4; n2=3; n3=5;
Calculate the global mean:
        mu=(n1*mu1+n2*mu2+n3*mu3)/(n1+n2+n3)

Fill out the table:

	Catalyst		
	A	B	C
Sample mean, μ_I			
Sample s.d., s_i			
Number of values, n_i			

Calculate the estimate of the variance as the total sum of squares divided by the number of degrees of freedom:

s = sqrt((($n1-1$)*s1^2 + ($n2-1$)*s2^2+($n3-1$)*s3^2)/($n1+n2+n3-3$))

We now want to know if the sum of squares for each catalyst is significant compared to that from random variations.

In following steps below, you will fill out this table, which is a standard way of presenting the ANOVA results:

Source of variation	Degrees of freedom	Sum of squares	Mean square	Mean square ratio, F
Between catalysts	2			
Residual	9			-
Total	11		-	-

Residual SS = ($n1-1$)*s1^2 + ($n2-1$)*s2^2+($n3-1$)*s3^2

Treatment SS = n1*(mu1-mu)^2+n2*(mu2-mu)^2+n3*(mu3-mu)^2

Mean square = sum of squares divided by the respective number of degrees of freedom.

The significance of the ratio F can be evaluated using a so-called F distribution (will discuss later).

The next step is to examine the residuals but we will do this later.

2. Now do the same much quicker with the Matlab's 'anova' function.

   p=anova1(X) performs the ANOVA analysis and returns the probability that the null hypothesis is true.
The data corresponding to each treatment (i.e. each catalyst) should be the columns in the matrix X. To construct the matrix, you first need to make sure that your data columns are of equal length. The simplest way is to go to the Array Editor and enter NaN for the 5th element of the array A and the 4th and 5th elements of the array B.
Now, enter
       X=[A B C]
       p=anova1(X)

Inspect the table and the figure that are produced.