

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Civil and Environmental Engineering

1.017/1.0101.017/1.010 Computing and Data Analysis for Environmental Applications / Uncertainty in Engineering

Problem Set 8: Analysis of Variance
Due: Tuesday, Nov. 20, 2003

Please turn in a hard copy of your MATLAB program as well as all printed outputs (tables, plots, etc.) required to solve the problem.

Problem 1

Consider the following set of dissolved oxygen concentration data (x_{ij}) obtained in 4 different seasons (rows), 6 samples per season (columns):

5.62	6.12	6.62	6.21	7.08	5.36
7.70	8.31	8.80	8.24	7.87	7.44
2.52	5.44	4.94	2.99	4.39	4.44
6.77	6.65	6.01	6.26	7.09	6.05

Write a MATLAB function to carry out a single factor ANOVA to test the influence of season on this data set. Your function should check the **normality assumption** by plotting the raw data with `normplot`. It should derive the **treatment** and **error sums-of-squares** and mean-squared values, the corresponding F statistic, and the p value for the test. Check your program by comparing its outputs to the ANOVA table given below (values obtained from the internal MATLAB `anova1` function). **When using `anova1` be sure to transpose the data array** (MATLAB requires treatments in columns and replicates in rows).

Source	SS	df	MS=SS/df	F	P
Treatments	47.1642	3	15.7214	29.7986	1.4094e-007
Error	10.5518	20	0.52759		
Total	57.716	23			

Problem 2

This problem is a continuation of the analysis of Massachusetts Water Resources Authority (MWRA) Boston Harbor data started in Problem Set 7. In the earlier problem set you used a large sample hypothesis test to examine the difference between population means for two different parts of the harbor (the outer and inner harbors). In the present problem set you consider rainfall to be the factor in a single factor ANOVA, with the treatments being three

different rainfall intensity categories. The rainfall categories are quantified in terms of the root mean squared sum (RMS) of the current and preceding two days precipitation values, measured in inches. This RMS value is tabulated, together with station ID, date, coliform count C (reported in organisms per 100 ml), and a few other variables in the text file `lowercharles.txt`.

You should group the coliform counts for all times and stations in the data file into three treatments according to the following RMS rainfall classifications (treatments):

- Treatment 1: $\text{RMS} = 0$ (or blank)
- Treatment 2: $0 < \text{RMS rain} < 0.25$ in.
- Treatment 3: $0.25 < \text{RMS rain}$

Make sure that you have an equal number of replicates for each treatment. Also, convert the coliform count C to the transformed count $C_T = \ln(C+1)$. Check for normality as in Problem 1 above. Carry out a single factor ANOVA, using either your own ANOVA program (from Problem 1) or the MATLAB `anova1` function. Present the ANOVA table, report the p value, and discuss the significance of your results.