Statistical analysis

Full Factorial Experimental Design and Interpretation

This experiment has a full factorial design; that is the dependency of any variable on any other variable, as well as interactions between variables can be estimated. This particular design is complicated by the presence of both categorical variables (i.e. variables which cannot be ranked in any meaningful order; in this case Treatments 1-4) and quantitative variables (numerical measures). It is also complicated by the presence of four values for the variable Treatment.

Such an experiment is analysed using analysis of variance; that is, the observed variance (a measure of how far values lie from the mean) is attributed quantitatively to the different measured variables, thus allowing the relative significance of the different variables in controlling the outcome of the experiment to be quantified.

While the results of analysis can be presented in many ways, the box-and-whisker plot is a useful way of summarizing the data graphically. For each variable, the minimum and maximum values in the range are shown (the ‘whiskers’), together with the lower and upper quartiles, which define the top and bottom of the ‘box’, and the median value.

Model fitting

A linear model was fitted using predictors in both categorical and numeric terms; the data for which are shown in Table 1. The categorical equation the data were fit to was:

$$\%Pu = \mathrm{constant}_{Term1} + \mathrm{coeff}_{T1} \times (T=1) + \mathrm{coeff}_{T2} \times (T=2) + \mathrm{coeff}_{T3} \times (T=3) + \mathrm{coeff}_{T4} \times (T=4) + \mathrm{coeff}_{C} \times (C=H) - \mathrm{coeff}_{C} \times (C=L) + \mathrm{coeff}_{P} \times (P=H) - \mathrm{coeff}_{P} \times (P=L) + \mathrm{coeff}_{T*C(1)} \times (C=H \,\&\, T=1) - \mathrm{coeff}_{T*C(1)} \times (C=L \,\&\, T=1) + \mathrm{coeff}_{T*C(2)} \times (C=H \,\&\, T=2) - \mathrm{coeff}_{T*C(2)} \times (C=L \,\&\, T=2) + \mathrm{coeff}_{T*C(3)} \times (C=H \,\&\, T=3) - \mathrm{coeff}_{T*C(3)} \times (C=L \,\&\, T=3) + \mathrm{coeff}_{T*C(4)} \times (C=H \,\&\, T=4) - \mathrm{coeff}_{T*C(4)} \times (C=L \,\&\, T=4) + \mathrm{coeff}_{T*P(1)} \times (P=H \,\&\, T=1) - \mathrm{coeff}_{T*P(1)} \times (P=L \,\&\, T=1) + \mathrm{coeff}_{T*P(2)} \times (C=H \,\&\, T=2) - \mathrm{coeff}_{T*P(2)} \times (P=L \,\&\, T=2) + \mathrm{coeff}_{T*P(3)} \times (P=H \,\&\, T=3) - \mathrm{coeff}_{T*P(3)} \times (P=L \,\&\, T=3) + \mathrm{coeff}_{T*P(4)} \times (P=H \,\&\, T=4) - \mathrm{coeff}_{T*P(4)} \times (P=L \,\&\, T=4) + \mathrm{coeff}_{C*P} \times (C=H \,\&\, P=L \mid C=L \,\&\, P=H) - \mathrm{coeff}_{C*P} \times (C=H \,\&\, P=H \mid C=L \,\&\, P=L) + \mathrm{coeff}_{T*C*P(1)} \times (C=H \,\&\, P=L \,\&\, T=1 \mid C=L \,\&\, P=H \,\&\, T=1) - \mathrm{coeff}_{T*C*P(1)} \times (C=H \,\&\, P=H \,\&\, T=1 \mid C=L \,\&\, P=L \,\&\, T=1) + \mathrm{coeff}_{T*C*P(2)} \times (C=H \,\&\, P=L \,\&\, T=2 \mid C=L \,\&\, P=H \,\&\, T=2) - \mathrm{coeff}_{T*C*P(2)} \times (C=H \,\&\, P=H \,\&\, T=2 \mid C=L \,\&\, P=L \,\&\, T=2) + \mathrm{coeff}_{T*C*P(3)} \times (C=H \,\&\, P=L \,\&\, T=3 \mid C=L \,\&\, P=H \,\&\, T=3) - \mathrm{coeff}_{T*C*P(3)} \times (C=H \,\&\, P=H \,\&\, T=3 \mid C=L \,\&\, P=L \,\&\, T=3) + \mathrm{coeff}_{T*C*P(4)} \times (C=H \,\&\, P=L \,\&\, T=4 \mid C=L \,\&\, P=H \,\&\, T=4) - \mathrm{coeff}_{T*C*P(4)} \times (C=H \,\&\, P=H \,\&\, T=4 \mid C=L \,\&\, P=L \,\&\, T=4)$$

where terms are H (high concentration) or L (low concentration), C is the CMS term, P is the polyelectrolyte term, and T is the Treatment term (1-4). Constant = coefficient$_{Term\ 1}$ (refer to Table 1 for coefficients).

The complexity of the above equations is due to the use of categorical variables, (the Treatment variable is categorical). However, each of the terms in brackets is a Boolean expression equating to either zero or one; and in use only 7 terms were non-zero (the constant, the three main predictors and three interaction terms) for analysis of each individual experimental result.

Table 1 ANOVA table from fit of data to multi-linear regression model.

Least Squares Coefficients, Response O, Model MREG_FF_B

Term	Coeff.	Std. Error	T-value	Signif.
1	21.5625	3.444359		
T<3df>				
1	30.4375	5.965804		
2	4.6875	5.965804		
3	-17.8125	5.965804		
4	-17.3125	5.965804		
C_N	2.4125	0.487106		
P_N	139.821429	13.917311		
T*C_N<3df>				
1	-2.0625	0.843692		
2	1.0375	0.843692		
3	3.2875	0.843692		
4	-2.2625	0.843692		
T*P_N<3df>				
1	-43.392857	24.105489		
2	37.321429	24.105489		
3	118.75	24.105489		
4	-112.678571	24.105489		
C_N*P_N	-7.303571	1.968205		
T*C_N*P_N<3df>				0.01505
1	5.660714	3.409031	1.660505	0.103332
2	-2.839286	3.409031	-0.832872	0.409042
3	-9.625	3.409031	-2.823383	0.006898
4	6.803571	3.409031	1.995749	0.051653

No. cases = 64 R-sq. = 0.8909 RMS Error = 13.78
Resid. df = 48 R-sq-adj. = 0.8568 Cond. No. = 11.66