TENTAMEN: Experimental design (April 28, 2011)

Short solutions

1) a) A distribution that represents a characteristic set of outcomes which could occur if there was no difference between the two groups.

   b) Do not assume normality or that there is no autocorrelation

   c) To increase precision

   d) May obtain an upper/lower limit for the difference; may help to decide how many observations would be needed in order to obtain a certain precision (width of the interval).

2) $L_1 = 11.25$, $L_2 = 4.75$, $L_3 = -0.25$, $L_{12} = 9.25$, $L_{13} = 0.25$, $L_{23} = 0.75$, and $L_{123} = -0.75$. Plot the effects and a normal plot. Based on the plots, factors 1 and 2 as well as the interaction between them seem active. The best yield we obtain at high temperatures and high pH, but the agitation rate seems not to have an effect on the yield.

3) a) D=ABC, i.e. I=ABCD, and AB=CD, AC=BD, AD=BC, A=BCD, B=ACD, C=ABD, and D=ABC.

   b) IV

   c) All the main effects are confounded with a third order effect which are assumed to be 0. Therefore, you can estimate the main effects alone.

   d) $L_A = 5.67$, $L_B = -0.155$, $L_C = -2.905$, $L_D = -7.19$, $L_{AB} = L_{CD} = 0.645$, $L_{AC} = L_{BD} = 0.065$, and $L_{AD} = L_{BC} = -0.14$. The variance of an effect can be estimated by $s^2/4 = 0.112$, where $s^2 = (\sum(x_{1i} - x_{2i})^2/2))/8$. Therefore, s.d.(effect)=0.33 and $t^{(8)}_{0.025} = 2.306$ so that the effects should be compared with $0.33 \cdot 2.306$. A, C and D seem active.

4) a) $y_i = a + b_1 n_i + b_2 m_i + b_{12} n_i \times m_i + \epsilon_i$, where $\epsilon_i \sim N(0, \sigma^2)$ and ϵ_i's are independent.

   b) Residual sum of squares is 124, and df's 4, 8 and 12.

   c) $s_E = 16$, $s_L = 124 - 16 = 108$, and $(s_L/df_L)/(s_E/df_E) = (108/2)/(16/6) = 20.25$. Since $F(2, 6) = 5.14$ at the 5% level, it seems that the fit is not good.

   d) $\hat{\sigma}^2 = s_E/df_E = 2.67$

5) a) A randomized block design, groups are blocks.

b) Df's are 2,4,8, and 14, MS's 9.87, 376.93 and 11.53, F's 0.8555 and 32.6821, and p-values 0.4606 and $5.3 \cdot 10^{-5}$.

c) There methods do not seem to differ but the groups do. It seems that group 1 has lost least weight, and group 4 and group 5 most.

d) Errors independent and $N(0, \sigma^2)$-distributed.

e) One could look at the relative weight loss instead. It is obvious that the men in group 5 should loose most wieght in pounds since they are the heaviest.