TENTAMEN: Experimental design (April 12, 2012)

Short solutions

1) a) Idea: minimize the sum of squares of the differences between the expected (modelled) and observed values. For the estimation, no assumptions needed.

   b) See the book (Chapter 11).

2) a) Code first the values $T = (\text{Temperature} - 55)/5$, $C = (\text{Concentration} - 8)/2$, and $S = (\text{Stirring rate} - 80)/20$, so that each of the factors have values -1 and 1.

   Main effects: $L_T = 1$, $L_C = 20$ and $L_S = 5$; Interaction effects: $L_{TC} = 0$, $L_{TS} = -1$, $L_{CS} = 6$, $L_{TCS} = 1$

   b) Variance of each observation is 4, i.e. $\sigma^2 = 4$. Therefore, Var(Effect) is $\sigma^2/2 = 2$. An effect is significant if the interval

   $$\text{Effect} \pm 2 \cdot \sqrt{\text{Var(Effect)}} = \text{Effect} \pm 2.8$$

   does not include 0. Therefore, the effects of concentration and stirring rate as well as their interaction are significant (active). Yield increases with increasing concentration and stirring rate. The effect of concentration increases with increasing stirring rate. Therefore, large values of concentration and stirring rate are preferred.

   c) Model for yield becomes $\hat{y} = \bar{y} + 10 \cdot C + 2.5 \cdot S + 3 \cdot CS$. If we move to the direction of steepest ascent (ignoring the interaction term), we could choose the two new experimental points such that (Temperature, Concentration, Stirring) is either (55°C,13%,92.5rpm) and (55°C,18%,105rpm). Alternatively, the stirring rate could be kept at 100rpm and concentration is given two values, e.g. 13% and 18%.

3) a) We want to test $H_0 : \mu_1 = \mu_2$ against $\mu_1 \neq \mu_2$.
   40 items total. Take 20 at random to be the times X_1 (no one waiting) and the remaining 20 to be the times X_2 (someone waiting). Compute $\bar{X}_1 - \bar{X}_2 = T_{\text{mean}}$. Repeat this N times.
   $\begin{pmatrix} 40 \\ 20 \end{pmatrix}$ randomizations possible.

   b) p value is $(17 + 13)/5000 = 0.006$ (which is much less than 0.05).

   c) $T = (\bar{X}_1 - \bar{X}_2)/(S_p\sqrt{1/n_1 + 1/n_2})$, where
   $S_p^2 = ((n_1 - 1)S_1^2 + (n_2 - 1)S_2^2)/(n_1 + n_2 - 2))$ and $T \sim T_{38} \approx T_{40}$. T gets the value -2.15 (which is not in the critical region since $T_{0.025}^{40} = 2.021$).

d) In b), H_0 would be rejected at the 5% level but not in c). Both b) and c) test whether the difference between the two means is significantly different from 0.
Randomization: no normal assumptions, no random sample, and $\sigma_1^2 = \sigma_2^2 = \sigma^2$
T-test: $X_1 \sim N(\mu_1, \sigma)$ and $X_2 \sim N(\mu_2, \sigma)$, two independent random samples
Since the distributions of X_1 and X_2 are very skewed (and we do not know whether the samples are random samples) randomization test would be recommended.

4) a) $Y_{ti} = \mu + \tau_t + \beta_i + \epsilon_{ti}$, where μ is the overall mean, τ_t the treatment effect, β_i the block effect and $\epsilon_{ti} \sim N(0, \sigma)$ independent random errors.

b) The randomization procedure would assign a number 1-16 to each of the rabbits. Put four slips of marked either 1, 2, 3, or 4 in a bowl. Select a number at random 1-16 to select a rabbit for Diet A, and pull a number out of the bowl to select a position on the top row. Repeat three times without replacement for Diets B, C, and D to complete the assignment to the top row. Follow the same procedure to assign the other three rows.

c) The ANOVA table becomes

Source	df	SS	MS	F ratio
Diet	$4-1=3$	191.5	63.833	12.6264
Shelf	$4-1=3$	38.0	12.667	2.5055
Residuals	9	45.5	5.0556	
Total	$16-1=15$			

d) Since $F(3,9) = 3.86$ at the 5% level, diet has a significant effect but not shelf.

e) If we had ignored the possible effect of the height of the shelf, regular ANOVA could have been used. If we wanted to take into account both the height of the cage (shelf) and the position on the shelf, we could use a Latin squares design.

5) a) 2_{III}^{7-4} fractional factorial design . Resolution is 3 and the defining relation is $I = ABD = ACE = BCF = ABCG = BCDE = ACDF = CDG = ABEF = BEG = AFG = DEF = BDFG = ADEG = CEFG = ABCDEFG$.

b) All three factor and higher order interactions are 0.

c) It seems that l_A, l_C and l_E are large indicating that either A, C and AC, or A, E and AE or C, E and CE are active.

d) Total flip-over, i.e. change all the signs in order to get the main effects free from confoundings.