1 Review problem set 6 solutions

Here are solutions to review problem set 6:

For the Friedman test with the data from problem 15.6: The rank means $\overline{R}_{i.}$ for the treatments are $1.79, 1.5$, and 2.71, for the No Music, Hard Rock, and Classical groups, respectively. Calculations lead to a Friedman statistic value of $FR = 5.59$. Since there is a tie of ranks in the data, we can adjust the statistic via the following formula:

$$FR' = \frac{FR}{K}, \text{ for } K = 1 - \sum \frac{T_i^3 - T_i}{bt(t^2-1)},$$

where the sum is over all groups of tied data that occur within the same block, T_i is the number of tied values in the i^{th} tied group (all occurring in the same block), and b and t are as usual the number of blocks and treatment levels, respectively. For these data, $K = 1 - (2^3 - 2)/[(7)(3)(8)] = 27/28$, so $FR' = 5.8$, which can be compared to a chi-squared distribution with 2 degrees of freedom. Here the P value is .054, which is larger than the RCB Anova P value of .012.

For the data from the second problem: The programs for defining dummy variables and reading in the data are on the website.

The regression model is $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \varepsilon_i$, where the x variables are group indicators for groups 1 through 4. For this model the ANOVA H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ is equivalent to the overall regression test H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$. The overall regression F test yields $F = .81, P = .54$, so we do not reject H_0.

For the group means, we have $\widehat{\mu}_1 = \widehat{\beta}_0 + \widehat{\beta}_1 = 53.2 + (-13.8) = 39.4$, $\widehat{\mu}_2 = \widehat{\beta}_0 + \widehat{\beta}_2 = 53.2 + (-9.0) = 44.2$, $\widehat{\mu}_3 = \widehat{\beta}_0 + \widehat{\beta}_3 = 53.2 + (-1.2) = 52.0$, $\widehat{\mu}_4 = \widehat{\beta}_0 + \widehat{\beta}_4 = 53.2 + (-12.4) = 40.8$, and $\widehat{\mu}_5 = \widehat{\beta}_0 = 53.2$.

For the data from the third problem: a) From the plot it appears that a linear relationship between x and y looks appropriate for all three groups, the lines appear to be parallel, and the intercepts appear to differ.

b) For the parallelism test, let x_1 be the x variable listed in the problem, x_2 and x_3 are indicator variables (dummy variables) for processes 1 and 2, and $x_4 = x_1 x_2$ and $x_5 = x_1 x_3$ are crossproduct variables. Then for the parallelism test the complete model is: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \varepsilon_i$ and the reduced model is: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$. The F statistic equals $F = \frac{(10369-10212)/2}{80.7} = .97$, the P value is much greater than .05 so we do not reject H_0: $\beta_4 = \beta_5 = 0$ and we conclude that parallelism is satisfied.

c) The test of equality of adjusted group means compares the complete model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$ to the reduced model $y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$. The F statistic equals $F = [\,(10212-4999)/2]/80.4 = 32.42$, which is greater than $F_{2,14,.05} = 3.74$. Thus we reject H_0: $\beta_2 = \beta_3 = 0$, and conclude that the adjusted group means differ.

d) For the adjusted group means we use the fact that $\overline{x} = 30.5$ to obtain $\widehat{\mu}_1 = \widehat{\beta}_0 + \widehat{\beta}_1\overline{x}_1 + \widehat{\beta}_2 = -25.5 + (4.16)(30.5) + 16.64 = 118.02$, $\widehat{\mu}_2 = \widehat{\beta}_0 + \widehat{\beta}_1\overline{x}_1 + \widehat{\beta}_3 = -25.5 + (4.16)(30.5) + 41,45 = 142.83$, and $\widehat{\mu}_3 = \widehat{\beta}_0 + \widehat{\beta}_1\overline{x}_1 = -25.5 + (4.16)(30.5) = 101.38$.

e) There is one outlier from process 1 that is visible on the residual plot as well as the normal plot. For an actual analysis of data this point can be further examined, and we can consider reanalyzing the data without it for comparison.