Statistics 106: Analysis of Variance

Sample Final Examination Sketched Solutions

Fall Quarter, 2002

1. (a) *If subject effects are assumed to be fixed, specify the linear model that lies behind the computation of SSE1 above. Be sure to include all assumptions. The estimators for the parameters under this model turned out to be* $\hat{\mu} = 30.7, \hat{P}_1 = 5.4, \hat{P}_2 = -5.2, \hat{P}_3 = 3.4, \hat{P}_4 = -3.4, \hat{P}_5 = 0$. *Explain the origin of these numbers.*

   Write the linear model as

$$Y_{ij} = \mu + S_i + P_j + e_{ij}$$

   and specify the assumptions on the errors and constraints on the parameters. $\hat{\mu} = \bar{Y}_{..}, \hat{P}_j = \bar{Y}_{.j} - \bar{Y}_{..}$, substituting from the means that are given.

   (b) *If subject effects are assumed to be fixed, give the ANOVA table that would be used for testing the significance of subject and position effects.*

ANOVA

Source	df	SS	MS	F^*
Subjects	10	1990.8	199.08	8.49
Position	4	867.5	216.88	9.25
Error	40	938.1	23.452	
Total	54			

   (c) *Use Tukey's method to find the difference between two probe-word position means that would be required to say that they are significantly different. Use an overall level of 95%. Specify which probe-word positions have significantly different reaction times. What difference would have been required if one had used Bonferroni's method?*

$$T = \frac{1}{\sqrt{2}} q(.95, 5, 40) = 4.04/\sqrt{2} = 2.857$$

$$S(\hat{D}) = \sqrt{2(23.452)/11} = 2.0652$$

   and the Tukey value is $2.857(2.0652) \approx 5.9$. Ranking the means, we see that 1 is larger than 4,5,2, 3 is larger than 4,2. The Bonferroni value depends on $t(1 - .05/20, 40) = t(.9975, 40) \approx 2.97$, making the difference required to be $2.97(2.0652) = 6.14$. Note that $g = 10$ here.

   (d) *If subject effects are assumed to be random, how does the model given above change? Estimate the variance due to subject effects under this model. Estimate the correlation between the response times at two different probe-word positions.*

   $S_i, i = 1, \ldots, 5$ are now assumed to be independent normally distributed random variables with zero means and variances σ_S^2. From Table 24.1 with no interaction and $n = 1$, $\hat{\sigma}^2 = 23.452$ and $199.8 = \hat{\sigma}^2 + 5\hat{\sigma}_S^2$, so that $\hat{\sigma}_S^2 = 35.27$ and the correlation is $\hat{\rho} = 35.27/(23.45 + 35.27) = .6$

2. (a) *Give the linear models corresponding to the null and alternative hypothesis corresponding to testing the effect of concentration. Compute the error sum of squares that would have been obtained using a model that ignored concentration.*

   Write the usual three factor model equations with all terms in and again with α_i missing. $SSE(\text{Aout}) = 30,517 + 10,286 = 40,803$ with the error sum of squares computed by subtracting the total of the first sever numbers in the table from the last column.

   (b) *Exhibit the analysis of variance and test all main effects and interactions at an overall level of .07. Use Bonferroni's result (not the Kimball inequality).*

ANOVA

Source	df	SS	MS	F^*
Concentration	2	30,517	15,259	71.3
Batches	3	3,492	1164	5.58
Strain	3	7,777	2592	12.11
AB	6	309	52	.24
AC	6	1,044	174	.81
BC	9	2,120	236	1.10
ABC	18	1,487	83	.39
Error	48	10,286	214	
Total	95			

   Looking at the F^* values in the Table, we only need to check the first three. For an overall $\alpha = .07$, i.e. .01 for each of the seven tests, we would use $F(.99, 2, 48) \approx 5.14$ for Concentrations and $F(.99, 3, 48) \approx 4.26$ for Batches and Strains. We conclude that all main effects are significant and that no interactions are signficant.

   (c) *If one wishes to make all possible comparisons among concentrations using Bonferroni's method and an overall confidence level of 95%, determine how big a difference there needs to be between the response at two concentrations to say they are different.*

   There were 3 concentrations, hence, there will be $g = 3$ comparisons. and $t(.993, 48) \approx 2.52$. Also, $S(\hat{D}) = \sqrt{214/10} = 6.54$ for an overall required difference of 17.48.

   (d) *What conclusions can you draw from this experiment. How would you determine the strain-concentration combination that produces the most antibodies? You need not do this.*

   We conclude that the three main effects of concentration, batch and strain are statistically signficant and the interactions are not. To find intervals for the best Concentration strain concentration, look at $\bar{Y}_{i \cdot k \cdot}$, which is an average over batches and observations within each cell, i.e., 16 possible means, with 6 observations in each mean. In this case, we should compare the values derived from Scheffé with those obtained by the Bonferroni argument, with $g = 16$.