

Durham University
MATH1541 Statistics
Exercise Sheet 9

Kamil Hepak
Tutorial Group 4

Dec 2018

1 a) Decomposition of first value

A datapoint can be decomposed into the sum of the group effect for its group, its residual, and the mean for all datapoints in all groups: $y_{ij} = g_i + e_{ij} + \bar{y}$.

The group effect for Radius 1: $\bar{y}_i - \bar{y} = 41.292 - 22.856 = 18.436$

The residual for $y_{1,1}$: $y_{1,1} - \bar{y}_i = 100.7 - 41.292 = 59.408$

Total mean of all datapoints: $\bar{y} = 22.856$

Thus, the datapoint $y_{1,1}$, 100.7, can be decomposed into $18.436 + 59.408 + 22.856$

2 b) Effects and residuals plot

Attached

3 c) ANOVA Table

Total Sum of Squares = Group Sum of Squares + Residual Sum of Squares

Group SS: $\sum_{i=1}^3 n_i \cdot g_i^2 = 12(18.436^2 + (-7.497)^2 + (10.939)^2) = 6189.034$

Residual SS: $\sum_{i=1}^3 (n_i - 1) \cdot s_i^2 = 11(26.578^2 + 5.025^2 + 3.676^2) = 8196.691$

	Sum of Squares	Contribution to Total SS
Radius Group	6189.034	43.02%
Residual	8196.691	56.97%
Total	14385.725	100%

4 d) Interpretation of analysis

As visible on the effects and residuals plot, the group effects and residuals both have a noticeable effect on the variation of a recorded response. The residuals seem to have a fairly normal distribution based on their boxplot, with a few considerable outliers skewing the data slightly. It is not immediately clear if the group effects or residuals have a greater influence on the response variation - though it does appear that the group effects are slightly more powerful in this regard, sitting above and below the third and first quartiles of the residual boxplot, the ANOVA table disagrees. Here, we see that the residuals contribute about 14% more to variation than group effects - a small, but still noteworthy,

differential. Overall, it may be sensible to conclude that residuals and group effects seem to have a similar effect on variation, or that residuals seem only slightly more powerful - further testing would be warranted to reach a stronger conclusion.

However, our analysis has to be taken with some caution, as the datapoint $y_{1,1}$ is very large compared to the vast majority of datapoints - though it does not qualify as an outlier (when plotting a boxplot using R), it still positively skews the group effect and standard deviation for Radius 1, the residual plot, and total mean. A more accurate analysis might possibly be obtained if the measurement of 100.7 were to be classified as an outlier and excluded from calculations.

5 e) Assessment of homogeneity assumption

Homogeneity plot attached

The plot shows that Radii 4 and 7 have a significant overlap, but Radius 1 is very far from the other values, probably due to the extreme $y_{1,1}$ datapoint. We might conclude that homogeneity is limited only to the larger two Radii, or that there is no homogeneity in general across all three groups.