

Solution to GO Problem 13.8

The description of the study identifies three factors to be taken into account: temperature (1–3), slant range (1–3), and days (1–3). It is stated that the “procedure was repeated on 3 days”; this indicates the days to be a blocking factor. Due to different meteorological conditions on the 3 days, it is unreasonable a priori to assume that there are no day effects; however, the days would constitute a grouping into heterogeneous units (on the same day), and should therefore most naturally be considered as blocks. On the other hand, temperatures and slant ranges were the factors of interest controlled by experimenter, so they are treatment factors.

Each day, the 9 combinations of temperatures and slant ranges were randomized onto groups of 3 rockets that were fired simultaneously (in a “volley”). The experimental unit is therefore such a group of 3 rockets. Because the meteorological conditions may vary from one firing (of three rockets) to the next, we cannot consider the 3 rockets in a group as (true) replicates; additionally, there may be some interference between the rockets fired simultaneously (without a more detailed description of the experiment it’s difficult to say). The rockets are therefore subsamples of the same experimental unit; alternatively, we could say that the measurement unit (i.e., rocket) does not equal the experimental unit (i.e., volley). The potential dependence between (the outcomes of) the 3 rockets fired simultaneously invalidates the independence assumption of a standard linear model, and makes the analysis by such a model invalid. Thus standard linear modelling for the full dataset will not be considered further. This leaves us with two possibilities for modelling:

- i) average (or otherwise aggregate, see comments at the end of this solution) the outcomes of the three rockets in a volley and analyse the averages (same approach as in GO Example 13.1),
- ii) explicitly account for the potential dependence between the 3 rockets in each volley; approaches to do so may not have been discussed in the course by the time of the exercise, but the first idea would be to include effects of the volleys as so-called random effects.

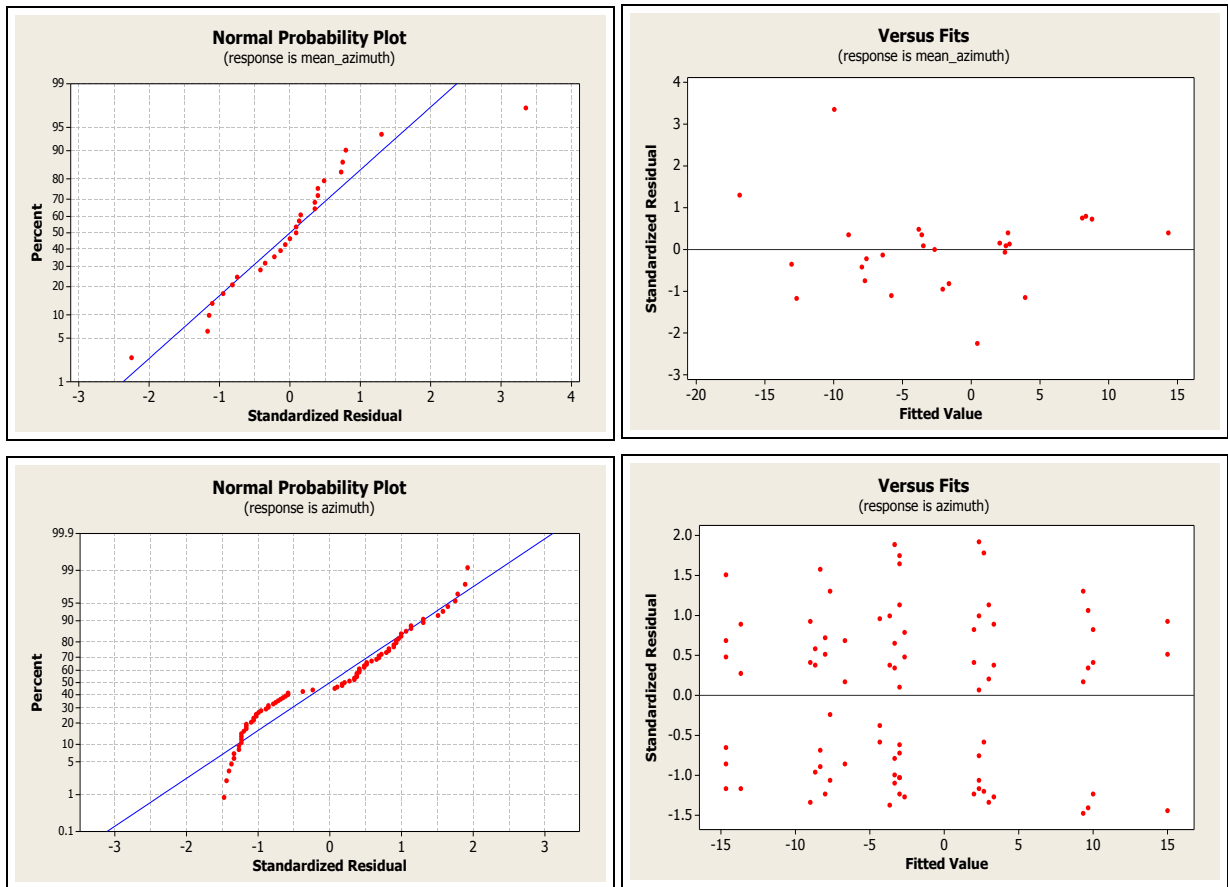
Denote by y_{ijkl} the azimuth error measured for rocket l at temperature level i and slant range level j on day k , where $i, j, k, l = 1, 2, 3$. The two models then take the form,

$$\begin{aligned} \text{i)} \quad \bar{y}_{ijk} &= \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + \varepsilon_{ijk}, \\ \text{ii)} \quad y_{ijkl} &= \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + ABC_{ijk} + \varepsilon_{ijkl}, \end{aligned}$$

where the α_i ’s and β_j ’s correspond to the main effect of temperature and slant range, respectively, the $(\alpha\beta)_{ij}$ ’s correspond to the interaction, and the γ_k correspond to day effects. The errors in models i) and ii), ε_{ijk} ’s and ε_{ijkl} ’s, respectively, are assumed i.i.d. and $\sim N(0, \sigma^2)$. Additionally in model ii), the volley random effects ABC_{ijk} ’s are assumed i.i.d. and $\sim N(0, \sigma_v^2)$.

The choice between the two models depends on the focus of the analysis. The inference for the effects of temperatures, slant ranges and days will be the same in both models, but model ii) additionally allows to separate the error variance into between-group and within-group variances. This may be of limited interest in practice.

The two sets of residuals plots on the next page show the error residuals in the two models, the first row for model i) and the second row for model ii).



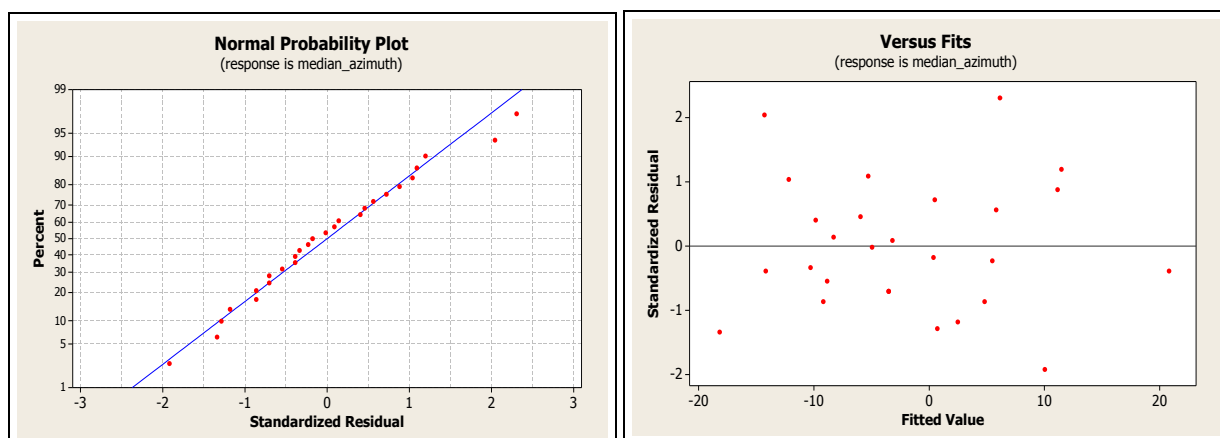
Both sets of residuals plots show problems. Among the residuals for model i), there seems to be one strong positive outlier. Inspection of the data shows this outlier to correspond to temperature 3 and slant range 2 on day 2. The deletion residual is 5.96, with a P -value of $2 \cdot 27 \cdot P(t(15) > 5.96) = 0.0007$ and therefore strong evidence of an outlying data point. For this observation, the three observations contributing to the average of -4.33 are $-10, -8, 5$. As the residual is positive, one may speculate that the observed azimuth value of 5 is an error. However, without any additional information about the data, we cannot make such an assumption.

The residuals for model ii) do not show a single extreme outlier, but the pattern of the residuals is not well described by a normal distribution. The normal plot looks strange, and a histogram shows a bimodal distribution with an apparent lack of residuals just below zero. As there seems little prospect of fixing this problem with the residuals by transformation, and the primary objective of our analysis is to make inference about the treatments, we decide to use model i) for the averaged azimuth errors.

Before turning to analysis of aggregated data, it is worth noting that the random effects model ii) is in fact *not* effective in accounting for dependence between the 3 rockets in a volley. That is because the model implicitly assumes the 3 rockets fired together to be non-negatively correlated, and the fitted model ends up showing a correlation of essentially zero. This more or less cancels the random effects out of the model, and we are left with a model assuming independence between rockets. There is a better, and more complex, way of accounting for the dependence between rockets in the same volley where the correlation can take both positive and negative values, and this model shows a much improved fit for a negative correlation (estimated at -0.43) between rockets in the same volley. That model also has nicer (and acceptable) residuals and must be considered the best model for the full data. As mentioned, this model is more complex and well beyond the course (and the GO textbook), so we will not pursue it further here.

When rerunning the analysis for model i) without the observation for the group with the extreme outlier, another extreme outlier appears, for temperature 3 and slant range 1 on day 2, and with a deletion residual of 4.92 ($P = 0.006$). As this also signals a clear outlier, we would need to remove both outliers before carrying out the analysis (and after that the residuals looks fine). This leads us to consider transformation to another scale. The data take both positive and negative values so the Box-Cox transformation analysis does not work because it requires strictly positive values of the outcome. It is always possible to add some number to all observations to make them all positive, but the transformation will then depend on the chosen number, and this approach is mainly used when adding a “small” number to make observations non-zero or when there is a biological lower bound for the observations. Another potential problem with transformation for these data is that values below and above zero will not be treated equally, contrasting that all the signs probably could be reversed without changing anything but the measurement direction for the data (and hence we should reach the same conclusions when analyzing the sign-reversed data).

As an alternative to removing the two observations, we consider here another approach: to replace the mean among the rockets in a group by the median within the group. This will no longer give the same inference as analysing the original data, but it leads to acceptable residuals (below) and may be viewed as a more robust procedure than analysing the means. Results for this analysis are shown below, including means with model-based standard errors.



Analysis of Variance for median_azimuth, using Adjusted SS for Tests

| Source | DF | Seq SS | Adj SS | Adj MS | F | P |
|--------------|----|---------|---------|--------|-------|-------|
| slant1 | 2 | 589.85 | 589.85 | 294.93 | 39.49 | 0.000 |
| temp1 | 2 | 585.41 | 585.41 | 292.70 | 39.20 | 0.000 |
| slant1*temp1 | 4 | 49.93 | 49.93 | 12.48 | 1.67 | 0.206 |
| day1 | 2 | 1039.19 | 1039.19 | 519.59 | 69.58 | 0.000 |
| Error | 16 | 119.48 | 119.48 | 7.47 | | |
| Total | 26 | 2383.85 | | | | |

S = 2.73269 R-Sq = 94.99% R-Sq(adj) = 91.86%

Least Squares Means for median_azimuth

| day1 | Mean | SE Mean |
|------|--------|---------|
| 1 | -8.111 | 0.9109 |
| 2 | -4.222 | 0.9109 |
| 3 | 6.556 | 0.9109 |

| | | |
|--------|--------|--------|
| slant1 | | |
| 1 | -7.556 | 0.9109 |
| 2 | -2.111 | 0.9109 |
| 3 | 3.889 | 0.9109 |
| temp1 | | |
| 1 | 4.111 | 0.9109 |
| 2 | -2.667 | 0.9109 |
| 3 | -7.222 | 0.9109 |

It is seen that the interaction between temperature and slant range is non-significant whereas all main effects are strongly significant. (Analyses of the mean azimuth error, with or without the extreme values, lead to the same significant effects, and to similar pairwise comparisons for the factors.) In order to compare factor levels, it is useful to compute an LSD value; this value can be used for comparisons for all three factors because the same number of observations (9) is involved in every factor level mean. For an unadjusted LSD value, we use $t(16, .025) = -2.12$, and for Bonferroni-adjustment for 3 comparisons, we use $t(16, .025/3) = -2.67$. Thus

$$\text{LSD} = 2.12 s / \sqrt{2/9} = 2.73, \quad \text{LSD}_{\text{Bon}} = 2.67 s / \sqrt{2/9} = 3.44.$$

It is seen that all differences between estimated factor levels exceed the LSD values. Therefore, it can be said that: all three slant ranges yield different azimuth errors, the numerically smallest for range 2; all three temperatures yield different azimuth errors, the numerically smallest for temperature 2; all three days yield different azimuth errors, the numerically smallest for day 2. It is not clear from the description of the study whether signed azimuth errors are preferably small or large, so for the interpretation above it has been assumed that errors numerically close to zero are preferable.

As a final note, one could think of other ways of achieving one measurement per experimental unit than computing means or medians. If the variation between the three rockets fired simultaneously was of interest, one could instead take the standard deviation between them as the outcome. If the deviations from zero (corresponding to no azimuth error) are important only by their magnitude but not by their direction, it would seem reasonable to analyze the numerical values of azimuth error; if this had been the intent of the exercise, one would perhaps assume the values to have been given as such. Generally speaking, our lack of information about the purpose of the experiment and of what constitutes a “good” outcome, limits our ability to choose the best analytical approach.