Solution to additional exercise 7.3

The data stem from an experiment to compare the survival of virus exposed to different solutions (thiosemicarbazon preparations). The outcome measured is proportional to the averaged inverse survival time of the virus in 6 eggs. As the purpose is to kill the virus, "good" results correspond to larger values of the outcome. There were 3 preparations, each applied in 3 concentrations, and the entire experiment was repeated once.

y_{ijk} = averaged inverse survival time (multiplied by a constant),
x_j = concentrations (logarithmic, the powers of 10: -4.6, -4.3, -4.0),

for the ith ($i = 1, 2, 3$) preparation measured at concentration x_j in run/replication k ($k = 1, 2$) of the experiment.

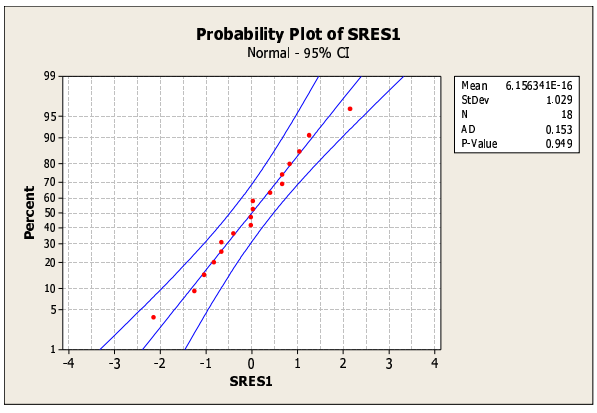
The design would most naturally be considered a two-factor (3×3) design with the 2 runs/replications as blocks. Apriori, there might be some difference between the first and second run of the experiment, and it seems therefore more natural to take them as blocks. As a general rule blocks are not entered into interactions with treatments, even though exceptions occur. The rationale behind the rule is that we do not expect a division of the experimental units into homogeneous groups (blocks) to influence treatment effects. The term *replication* unfortunately has two common usages: as a run (or repetition) of the entire experiments, and as an indication of identical experimental units (replicates) in the design. These two situations correspond to a block design and a completely randomized design, respectively. If one did not want to consider the two runs as a blocking factor, the design would be a three-factor design without replication (replicates). Then one would most naturally omit the three-factor interaction from the model in order to use it to estimate the error variance.

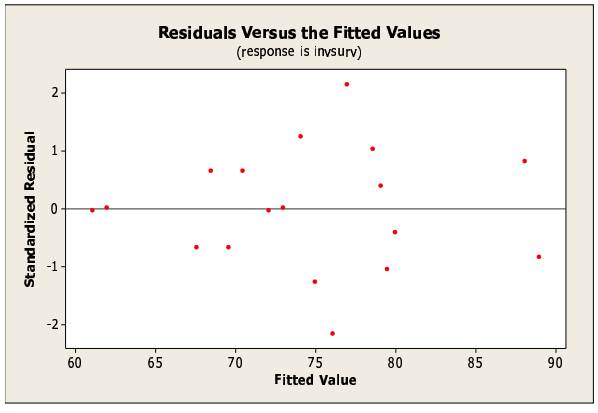
The statistical model for the block design is,

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + \varepsilon_{ijk}, \quad i = 1, 2, 3;\ j = 1, 2, 3;\ k = 1, 2, \tag{1}$$

where the errors ε_{ijk} are as usual assumed to be i.i.d. and $\sim$ N(0,σ^2). This model takes the 3 concentrations as levels and does not assume any particular relationship with the actual concentrations.

Model (1) may be analysed as a General Linear Model in Minitab; different portions of the output are shown below: normal and residual plots (standardized residuals versus fitted values), the ANOVA table, main effects and interaction plots for the two treatment factors.





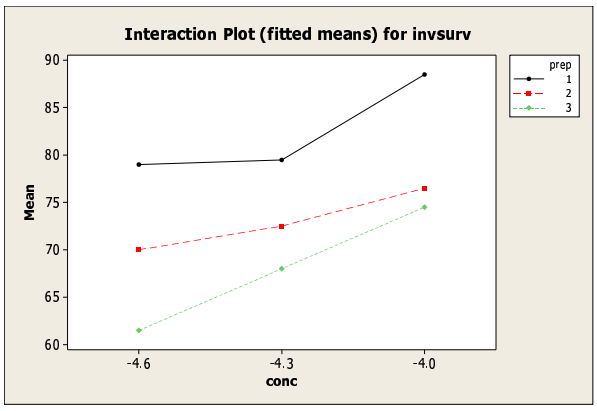
The residual plot shows some indication of increasing errors with increasing fitted values, although the last pair of points somewhat contradicts the pattern. Note that the plot has pairs of positive and negative residuals at almost the same fitted values; these are for the first and second replication (whose fitted values differ only little, by the lacking effect of replication). A Box-Cox analysis indicates, quite unusually, an optimal λ-value between -3 and -4 ($\hat{\lambda} = -3.63$ by the `boxcox` command in Stata), but the confidence interval even includes 1 (no transformation). The message hereof is that there is little gain in any transformation.

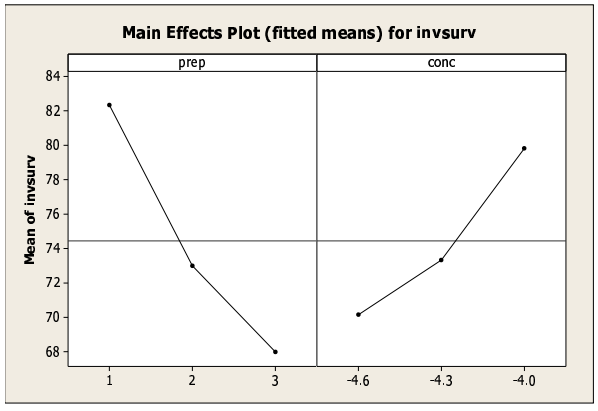
The normal plot of the residuals looks very good. There is, however, one pair of large standardized residuals, for observations 4 and 13. The deletion residuals are 3.10 and -3.10, which is far from statistically significance ($P = 2 \cdot 18 \cdot \mathrm{P}(t_7 > 3.10) = 0.31$, using the Bonferroni method). This would lead us to conclude that there is no urgent need to remove any outliers. It is noteworthy that after model reductions (below) the pair of outliers is tranformed into a single outlier (observation 13) with a weakly significant P-value for the outlier test in some models. The message here is that it is only one of the values in the pair that seems outlying, and the fact that the residuals come in pairs for the full model generates an artificial pair of outliers. This difference between residual patterns for the full and reduced models shows that it is sometimes useful to redo the model checks for the chosen, final model. In general, it is not considered necessary to redo model checks for every model reduction. As to whether observation 13 should be considered an outlier, we will take the conservative approach of keeping it in the model, at the cost of a higher residual error. The inference with and without observation 13 is not dramatically different.

ANOVA table:

Source	DF	Seq SS	Adj SS	Adj MS	F	P
prep	2	635.11	635.11	317.56	25.55	0.000
conc	2	291.44	291.44	145.72	11.72	0.004
prep*conc	4	34.89	34.89	8.72	0.70	0.612
replication	1	3.56	3.56	3.56	0.29	0.607
Error	8	99.44	99.44	12.43		
Total	17	1064.44				

The ANOVA table shows the interaction between preparations and concentrations as well as the replications to be non-significant. That is, the concentrations have similar impact for the 3 preparations, and no difference can be seen between the 2 runs of the experiment. Due to the balancedness of the design, removing non-significant terms serves only to pool the error variance; with a fairly residual DF one could argue both in favour of and against it, but we won't do it here. Further, we see strong effects of both preparations and concentrations: we can be pretty sure that differences exist both between preparations and concentrations. We proceed with the main effects and interaction plots.





The deviations from parallel curves may be considered as minor because the ordering of preparations is preserved across all concentrations, and vice versa. There does not seem to be any single non-parallel feature in the plot that could further be explored by contrasts. By the clearly non-significant interaction, we therefore proceed to look at preparations and concentrations separately.

The impact of concentrations may be seen as roughly linear with the actual power of the concentration. This can be explored further (below) but it is also perfectly valid to leave both factors as categorical. We then present the results as (least squares) means with standard errors and letter coding for pairwise statistical significance (using the Bonferroni method).

factor	level	least squares mean	standard error	letter coding
preparation	1	82.33	1.44	b
	2	73.00	1.44	a
	3	68.00	1.44	a
concentration	-4.6	70.17	1.44	a
	-4.3	73.33	1.44	a
	-4.0	79.83	1.44	b

It is seen that for both factors there is a complete separation of the factor levels (quite unusually). The P-value for the Bonferroni pairwise comparisons of preparations 2 and 3 is 0.067, falling just short of significance at the simultaneous 5% level to declare them different. Summarised in plain words, preparation 1 and concentration -4.0 work best (have shortest survival times).

To investigate a linear modelling of concentration, we may fit a reduced model of (1) with a regression term for the concentration (polynomial contrast analysis is another option). Before doing so, we should consider the possibility of eliminating non-significant terms from model (1). With a clearly non-significant interaction, there is really no need to fit separate regression lines for the 3 preparations. Also, the replications could be removed without affecting the model fit; however, as the blocks are truly part of the statistical design (to repeat the experiment is something else than having two replications at each treatment combination), we decide to retain the blocks in the model. Thus, our reduced model is

$$y_{ijk} = \mu + \alpha_i + \beta x_j + \gamma_k + \varepsilon_{ijk}, \quad i = 1, 2, 3;\ j = 1, 2, 3;\ k = 1, 2, \tag{2}$$

with the corresponding ANOVA table:

Source	DF	Seq SS	Adj SS	Adj MS	F	P
prep	2	635.11	635.11	317.56	28.38	0.000
conc	1	280.33	280.33	280.33	25.06	0.000
replication	1	3.56	3.56	3.56	0.32	0.583
Error	13	145.44	145.44	11.19		
Total	17	1064.44				

The F-test of linearity is computed as follows,

$$F = \frac{[145.44 - (99.44 + 34.89)]/[13 - (8 + 4)]}{(99.44 + 34.89)/(8 + 4)} = 0.99,$$

which is far from significant (in $F(1, 12)$). The estimated regression coefficient for (log) concentration is $\hat{\beta} = 16.1$ (with $\mathrm{SE}(\hat{\beta}) = 3.2$). That is, increasing the dose by a logarithmic unit of 0.1 implies an estimated increase in the outcome by 1.61. The least squares means for preparations are the same as above, and correspond here to a concentration of -4.3 (the mean value among the concentrations in the dataset). The comparisons among preparations also lead to the same conclusions.