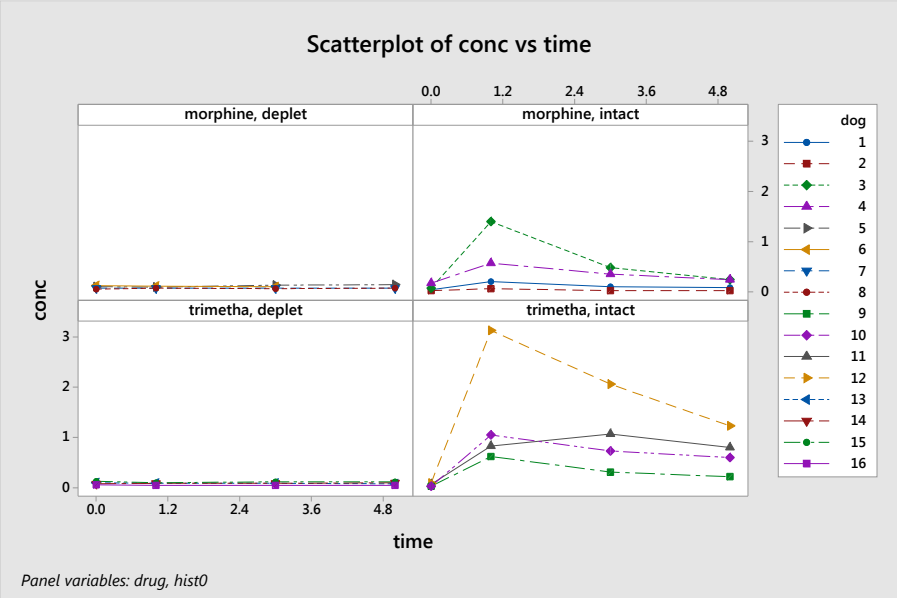
## Solution to additional exercise 10.8

A study investigated the effect of the drugs morphine and trimethaphan on histamine release in 16 mangrel dogs. Among the 8 eight dogs in each drug treatment group, half of the dogs had been treated so that their supplies of available histamine were depleted at the time of innoculation with the treatment drug; the remaining dogs had intact histamine levels. Measurements (blood histamine concentrations) were taken at minutes 1, 3 and 5 minutes after drug innoculation, as well as at time 0 immediately prior to innoculation.

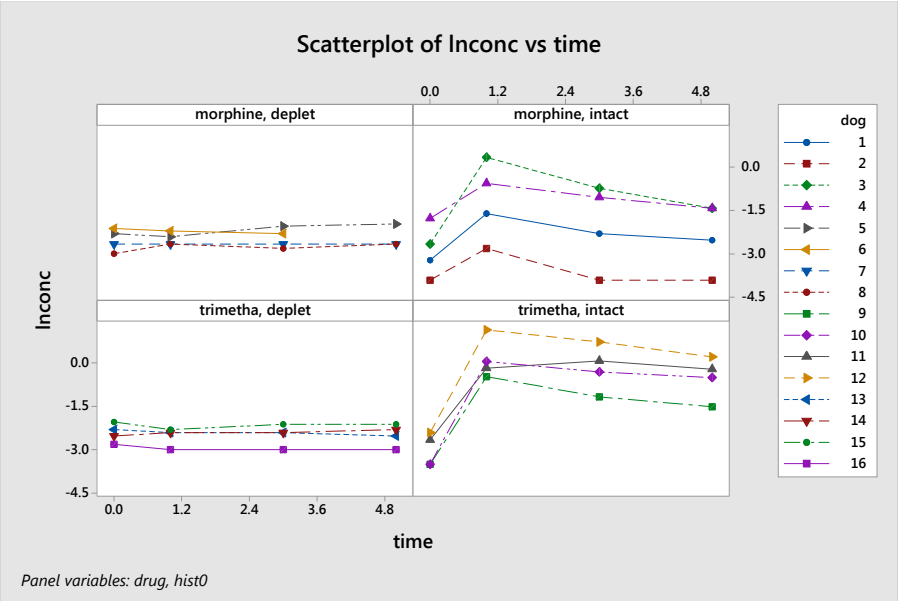
Notation:

$y_{ij}$ = blood histamine concentration of $i$th dog measured at $j$th time,
$i$ = $1, \ldots, 16 \sim$ dogs,
$j$ = $1, 2, 3, 4 \sim$ time (minutes 0,1,3,5).

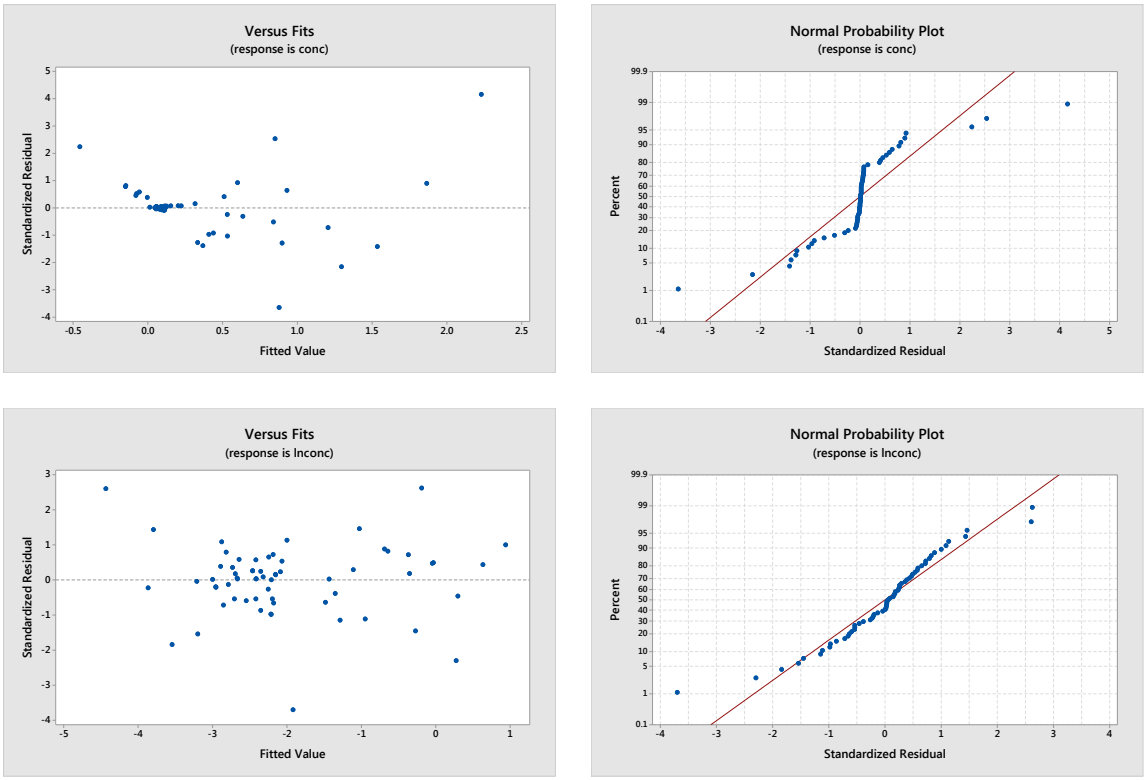
The data are longitudinal with 4 measurements over time on each dog. Two treatment factors (drug and initial histamine levels) were applied to the dogs, and there are 4 replicates (dogs) for each of the 2×2=4 treatment levels. Therefore, any analysis of a single outcome per dog would be a 2-way factorial with replication. One measurement is missing, but we will use the remaining measurements of that dog whenever possible in the analysis. The task is to analyse the data using different univariate methods for longitudinal data. We start by plotting the data against time, with connecting lines for measurements on the same dogs and panels for the four treatment groups.




The values of some dogs show large variation over time, whereas others have more or less constant values. It is seen that the presence or absence of variability is clearly associated with treatment groups. We try to plot (next page) the logarithmic concentrations also because the raw data plot seems to indicate much less variation at low histamine levels (this is also evident from the data, and in part a consequence of the concentrations being bounded from below at zero). Our first conclusion is that the spread seems more homogeneous among groups for log-transformed data.




The first question to clarify is the need of transformation. It is natural to base the decision on the residuals from the hierarchical (or split-plot) model for longitudinal data with random dog effects. We show here only the residual plots (standardised residuals versus predicted values, and the normal probability plot) for the untransformed and log-transformed data, and return to the results of the analysis later on.




It is very obvious that the residual plots for the raw data are unsatisfactory: there is a clear "edge" to the left caused by the lower bound of zero, clearly increasing variation with increasing predicted values, and the distribution of residuals is skewed. The plots for the log-transformed concentrations

are much better, even if not "perfect" (for example, there seems to be one large negative residual). The log-transformed values have the advantage of simple interpretation, and would seem sufficiently well-behaved for these data to not go further into finding an optimal transformation. The Box-Cox transformation analysis works only for fixed effects models (in its standard implementations in Stata and Minitab), so as an approximation one might try a model with fixed effects of dogs in addition to the other fixed effects; this gives an "optimal" power of $-0.15$ with a fairly narrow confidence interval that does not include 0 (e.g., in Minitab: $(-0.245, -0.055)$). A correct Box-Cox analysis (beyond the scope of the course) yields an optimal power around $-0.14$ with 0 just included in the 95% CI. With the estimate so close to zero it is natural to consider the log transformation as a reasonable choice of transformation. Somewhat surprisingly, the residual plots look decent also with the inverse square-root transformation, although this is not supported by the Box-Cox analysis. For this solution, we will stick to the log transformation.

**Separate analyses for times of measurements**

As our first exploratory analysis of the data, we analyse at each time point. The statistical model for fixed time $j$ is

$$\ln(y_{ij}) = \mu + \alpha_{\text{drug}(i)} + \beta_{\text{onset}(i)} + (\alpha\beta)_{\text{drug}\times\text{onset}(i)} + \varepsilon_{ij}, \quad i = 1, \ldots, 16,$$

where the two treatment factors of dog $i$ are determined by drug($i$) and onset($i$). The following summary (partial) ANOVA tables gives $F$-statistics and $P$-values for each of the 4 time points. The residual plots for the separate analyses do not look great (there is even after log-transformation very little variation in some of the treatment groups), but nor do they look catastrophic considering the low number of observations (16) for each analysis.

| | | 0 mins | | 1 min | | 3 mins | | 5 mins | |
|---|---|---|---|---|---|---|---|---|---|
| Source | df | $F$ | $P$ | $F$ | $P$ | $F$ | $P$ | $F$ | $P$ |
| Drug | 1 | 0.00 | 0.96 | 2.50 | 0.14 | 4.35 | 0.059 | 4.79 | 0.051 |
| Histamine onset | 1 | 2.68 | 0.13 | 25.5 | $<$0.0005 | 10.3 | 0.008 | 6.74 | 0.025 |
| D×H | 1 | 0.15 | 0.71 | 2.88 | 0.12 | 4.64 | 0.052 | 5.45 | 0.040 |
| Error | 12 | | | | | | | | |

In interpreting the test statistics, one should apply a Bonferroni correction to account for multiple testing. Tests of treatment effects are only of interest for the last 3 times, so a Bonferroni correction should multiply the above $P$-values by 3. After correction, the only significant results are differences between histamine onset levels at 1 and 3 minutes. The previous graphs show this quite clearly: the dogs with depleted levels show no reaction whereas the dogs with intact levels respond to the treatment by increased histamine levels. The measurements prior to treatment show no significant differences between the 4 groups of dogs, which is exactly what one would hope for. Overall, the analysis at separate time points seems too week here, where there is much variation between dogs and strong patterns within dogs. It is certainly of interest to explore more powerful methods to compare the impact of the drugs.

**Analysis of summary statistic (response feature): range over time**

The graphs of the data show the response (whenever present) to innoculation as an abrupt increase in blood histamine, typically followed by a slower decrease. One might expect the values to eventually decline to the initial level but the series over time may be too short to investigate that hypothesis. To capture the "peak" feature of the response profiles, the most common response features such as mean, slope, curvature, and "gain" do not seem natural. Two obvious alternative statistics are the

area under the curve (AUC) and the range (maximal value minus minimal value). For the latter, one might also calculate maximal value minus initial value, but the range seems simpler to work with. In Minitab, the range can be calculated using the descriptive statistics menu. We continue to use the logarithmic values, corresponding to "multiplicative ranges" (ratio of largest to smallest value). The statistical model is the same as above, only using as the outcome instead of $\ln(y_{ij})$ (for fixed time $j$) the summary statistic $\tilde{y}_i = \max_j \ln(y_{ij}) - \min_j \ln(y_{ij})$.

Not unexpectedly, the analysis shows the two groups of dogs with depleted histamine levels to have very small ranges, contrasting the two groups with intact levels. To facilitate the comparison between the groups we present the results in the layout of a 1-way ANOVA with 4 treatment groups (the interaction between drug and onset level is clearly significant).

```
One-way ANOVA: Range versus Tx

Source  DF      SS     MS      F      P
Tx       3  24.495  8.165  33.78  0.000
Error   12   2.900  0.242
Total   15  27.395
                                    Individual 95% CIs For Mean Based on
                                    Pooled StDev
Level           N    Mean   StDev  ---+---------+---------+---------+------
morphinedeplet  4  0.2402  0.1923   (---*---)
morphineintact  4  1.7284  0.8729               (---*----)
trimethadeplet  4  0.2227  0.0327  (----*---)
trimethaintact  4  3.2149  0.4084                          (----*---)
                                    ---+---------+---------+---------+------
                                     0.0       1.2       2.4       3.6
Pooled StDev = 0.4916
```

For this analysis, there is one quite outlying observation: dog no. 3, which has considerably higher range than the other 3 dogs in its group (`morphineintact`). We choose however to retain this observation in the data because there is nothing in the data to indicate an error. Its impact is to give its treatment group a high estimated standard deviation, and it is seen that the assumption of equal variation in the 4 groups is not met by the data. $P$-values for group comparisons would therefore preferably be based on two-sample $t$-tests (where no assumption of equal variance is needed). As seen from the estimates and the standard deviations, however, the conclusion is pretty clear even without carrying out the analysis: the trimethaphan drug gives a significantly higher relative rise in blood histamine levels than does morphine. As already noted, the dogs with depleted levels have only very little variation over time in their values.

**Analysis of hierarchical ("split-plot") model**

The hierarchical model has random effects at the additional level in the data: the dogs. Formulated in the split-plot terminology, we consider the dogs as "whole plots" and the measurements over time on each dog as "split plots". The model assumes the same correlation between two measurements on the same dog no matter their distance in time. According to the rule of thumb of short series, this assumption may be ok for a series of 4 measurements. However, the data plots show strong patterns for each dog, which would indicate that close observations have the strongest correlations.[1] Therefore,

[1] A repeated measures ANOVA (not shown here, see the SAS solution file) gives $\epsilon$-values for "the sphericity" parameter of 0.57 and 0.84, and a clearly significant test (Mauchly's test) for sphericity. However, the time effects are clearly significant even after the $\epsilon$-corrections. Analysis with error correlation structures (SAS or Stata) show no strongly declining correlations with time distance, but some evidence of a lower variance at time 0 than the other time points.

the analysis for the time effects must be taken with some reservation. The hierarchical model can be written,

$$\begin{aligned}\ln(y_{ij}) &= \mu + \alpha_{\text{drug}(i)} + \beta_{\text{onset}(i)} + (\alpha\beta)_{\text{drug}\times\text{onset}(i)} + A_i \\ &\quad +\gamma_j + (\alpha\gamma)_{\text{drug}(i)j} + (\beta\gamma)_{\text{onset}(i)j} + (\alpha\beta\gamma)_{\text{drug}\times\text{onset}(i)j} + \varepsilon_{ij},\end{aligned}$$

where the $A_i$'s are assumed i.i.d. and N(0,$\sigma_A^2$), and the $\varepsilon_{ij}$'s are assumed i.i.d. and N(0,$\sigma^2$).

```
Analysis of Variance for lnconc, using Adjusted SS for Tests

Source           DF   Seq SS   Adj SS   Adj MS      F      P
drug              1   5.5083   5.3217   5.3217   2.61  0.132 x
hist0             1  15.5282  14.6592  14.6592   7.19  0.020 x
time              3  12.7690  12.8307   4.2769  59.87  0.000
drug*hist0        1   5.7477   5.9056   5.9056   2.90  0.114 x
drug*time         3   2.2022   2.0582   0.6861   9.60  0.000
hist0*time        3  13.3465  13.3187   4.4396  62.14  0.000
drug*hist0*time   3   2.9037   3.0112   1.0037  14.05  0.000
dog(drug hist0)  12  24.6098  24.6098   2.0508  28.71  0.000
Error            35   2.5004   2.5004   0.0714
                                                 x Not an exact F-test.

Variance Components, using Adjusted SS
Source         Estimated Value
dog(drug hist0)    0.50537
Error              0.07144
```

Due to the one missing value in the data, some of the tests are no longer exact $F$-tests, but this small unbalancedness should have no major impact (and is no reason of concern). The ANOVA table shows all effects involving time to be strongly significant, including the 3rd order interaction. At this point of the analysis this is no real surprise; we've already seen that the response patterns over time for the 4 treatments are far from additive. The estimated variance components show most of the variation to be between dogs; the intraclass correlation between two measurements on the same dog is estimated at $\hat{\rho} = \hat{\sigma}_A^2/(\hat{\sigma}_A^2 + \hat{\sigma}^2) = 0.505/(0.505 + 0.071) = 0.88$ (which is a huge value).

The residual plots shown earlier indicated a possible strong outlier: the first observation of dog no. 3 which is unexpectedly small. Its deletion residual is -4.67, which is clearly significant ($P = 2 \cdot 63 \cdot 0.000023 = 0.0029$). Note that an outlier for dog no. 3 was already indicated in the analysis of the ranges. However, I would be reluctant to remove this observation, because it does not appear very extreme when looking at the data value or the plots, and because it has only little impact on conclusions (an analysis with and without the observation must be undertaken for such a statement!). The residuals in the analysis of dog least square means (=ordinary means, except for dog no. 6 with a missing observation) do not show any particular problems with the assumption of normality for the random effects. Again, the variation is smallest in the depleted dogs, but overall the normal distribution seems to be a reasonable assumption.

Due to the significant 3-factor interaction the results can not be summarized by simple main effects or 2-factor interactions. The strong interaction between the two treatment factors suggests to combine them into a single factor (with 4 levels). Then the results can be summarized by the interaction between the combined treatment and time; the values are listed below and shown in an interaction plot on the next page.

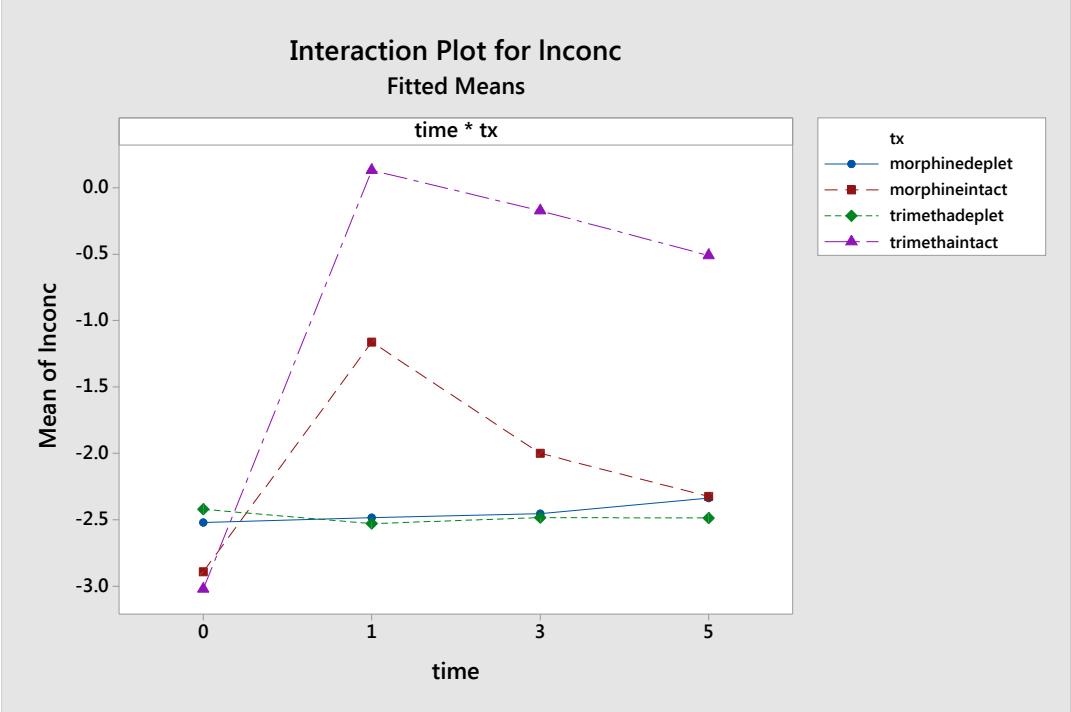
```
Least Squares Means for lnconc

drug*hist0*time      Mean  Inverse ln(Mean)
morphine  deplet  0  -2.519  0.0805
morphine  deplet  1  -2.483  0.0835
morphine  deplet  3  -2.454  0.0860
morphine  deplet  5  -2.336  0.0967
morphine  intact  0  -2.891  0.0555
morphine  intact  1  -1.162  0.3129
morphine  intact  3  -2.000  0.1353
morphine  intact  5  -2.323  0.0980
trimetha  deplet  0  -2.420  0.0889
trimetha  deplet  1  -2.529  0.0797
trimetha  deplet  3  -2.483  0.0835
trimetha  deplet  5  -2.486  0.0832
trimetha  intact  0  -3.020  0.0488
trimetha  intact  1   0.131  1.1400
trimetha  intact  3  -0.174  0.8404
trimetha  intact  5  -0.510  0.6005
```

The backtransformed means have been computed manually and added to the list; they are estimated *medians* for the treatment by time levels on original scale. Correct standard errors for the means and their differences would only be available from another software, but as a simple approximation we can use the formulae for a balanced design:

- means $\overline{\ln(y_{\text{drug}\times\text{onset},j})}$ : SE $\approx \sqrt{(\hat{\sigma}_A^2 + \hat{\sigma}^2)/4} = 0.38$,
- time differences $\overline{\ln(y_{\text{drug}\times\text{onset},j})} - \overline{\ln(y_{\text{drug}\times\text{onset},j'})}$ : SE $\approx \sqrt{2\hat{\sigma}^2/4} = 0.19$,
- treatment differences $\overline{\ln(y_{\text{drug}\times\text{onset},j})} - \overline{\ln(y_{\text{drug}\times\text{onset}',j})}$ : SE $\approx \sqrt{2(\hat{\sigma}_A^2 + \hat{\sigma}^2)/4} = 0.54$.

With only a slight unbalancedness from a single missing observation, these values can be expected to approximate the standard errors quite well (an analysis using SAS confirmed this supposition).




The plot suggests that there is no difference to be seen between the effect of the two drugs when histamine levels are depleted at onset, and that in this case the values are constant over time. Pairwise comparisons within each time are easily within two SEs (even within one SE), and also the comparisons over time within each treatment are within two SEs. There is a slight indication of an increase over time for the morphine group, but it does not seem significant (based on the pairwise comparisons).
Other questions of interest are comparisons between drugs and within drug for dogs with intact histamine levels. First, the two treatment groups (in fact, all four treatment groups) are within 2 SEs at time zero, as we would expect prior to innoculation. Second, the differences between drugs for intact dogs are above 2 SEs at times 1, 3 and 5. However, the difference after 1 minute is only slightly above 2 SEs ($t = (0.131 - (-1.162))/0.54 = 2.39$) so that a strict analysis using Bonferroni corrections for multiple comparisons might fail to declare this one significant. The differences for 3 and 5 minutes are much larger. The within-drug comparisons are more straightforward from the plot. There is very clear evidence of a spike in histamine levels at time 1 for both drugs, and there is also evidence of a drop from 1 to 5 minutes. The morphine group does not quite reach its initial value after 5 minutes.