Use of the MOSAIC web platform
Practical exercises - Answer elements

Sandrine CHARLES
sandrine.charles@univ-lyon1.fr

FCS Rovaltain, 17-18 March 2021

Introduction

You will find in this document some answer elements to the practical exercises with the MOSAIC web platform.

1   Quantal (or binary) data analysis

Step 1.1 Go to `http://pbil.univ-lyon1.fr/software/mosaic/` and choose Menu `surv`, then submenu `Standard`.

Step 1.2 Try first with example `chlordan`, what do you notice after the data have been automatically downloaded? Look at the entire dataset by extending the field or using the scroll bar.

> Empty fields are automatically filled in.
> By stretching the data field from the bottom-right corner, it is possible to see the whole dataset. It is also possible to use the scroll bar, especially when there are a lot of data.

Step 1.3 Then click `run` and go through the following items:

- While waiting for the fit results, how many concentrations have been tested?

> 6

- How many replicates per concentration have been done?

> 10

- Once fitting results have been obtained, what do you observe when changing the scale for the x-axis of the fitting plot?

> Control data are not visible any more.

- What is the meaning of parameter e? What di you about its estimate?

> Parameter e is in fact the LC_{50}, that is the 50% lethal concentration: the concentration that kills 50% of the organisms. From parameter estimates, we get $LC_{50}(chlordan) = 2.69\,[1.52; 5.3]$ (expressed in $\mu g.L^{-1}$). This value can be read from the fitting plot at the intercept of the dose-response curve with the horizontal line $y = 0.5$ (without log-scale).

- What is the meaning of parameter b? What di you about its estimate?

> Parameter b represents the intensity of the effet; the more b is high, the more the compound has a strong effect. From parameter estimates, we get $b(chlordan) = 1.18 [0.5; 2.1]$ (dimensionless). Regarding the fitted curve, the decrease fon the control to 100% of effect is rather slow, denoting a medium effect of chlordan on *D. magna* survival.

- What do you notice when looking at LC_x estimates?

> When x increases, the LC_x increases too.
> Note that if x could tend to infinity, the LC_x would converge towards what is classical called the "incipient" LC_x.
> We can check that $LC_{50} = e$.

- Regarding the PPC plot, does the fit seem correct or not?

> No, the fit cannot fully be considered as correct because 100% of the prediction uncertainty limits intersect with line $y = x$, while around 95% (namely between $[94\% - 96\%]$) is expected.

- Go back to the page top. Print the MOSAIC survival report. Is it in accordance with what you expected? If not, make some suggestions.

Step 1.4 Try now with example `cadmium1`. Which differences do you notice within the results?

> With `cadmium1` data we get an estimate for parameter d; this means that the control survival rate is not 100%, but rather $d = 0.929[0.881; 0.963]$.
> We get a warning for parameter e. Indeed, e is estimated at a value higher than the highest tested concentration. Hence, this estimate must be considered with caution. The consequent conclusion is that *D. magna* is slightly sensible to cadmium.
> Nevertheless, if we accept to consider that $LC_{50,cadmium} = 4.7[4.29; 6.47]$, given that $LC_{50,chlordan} = 2.69[1.52; 5.3]$ then we could conclude that *D. magna* is more sensible to chlordan than to cadmium, even if 95% credible intervals sligthly overlap. But, also considering parameter b for both data sets ($b_{chlordan} = 1.18[0.487; 2.12]$ and $b_{cadmium} = 6.43[2.07; 59.5]$) and the whole shape of the dose-response curves (see Figure 1), we notice that the curvature for cadmium is much higher than for chlordan. Hence, at a given concentration, *e.g.* 2 $\mu g.L^{-1}$, daphnids survive better under a cadmium exposure (at $\sim 90\%$) than under a chlordan exposure (at $\sim 60\%$). this illustrates the importance of considering tha whole dose-response curve, not only a summary statitics such as the LC_{50}.

Step 1.5 If available, load your own data set (pay attention to the data format) and run the survival analysis. Pay attention to choose the appropriate concentration unit and the correct Y-axis label. Save the report and discuss your results.

Step 1.6 If you tried a third example, note below your remarks on the provided results.

2 Count (or discrete) data analysis

Step 2. 1 Go to http://pbil.univ-lyon1.fr/software/mosaic/ and choose Menu `Repro`.

Step 2. 2 Try first with example `chlordan`. Compare the results with those obtained on survival data. What about parameter e? Does it have the same meaning? The same value?
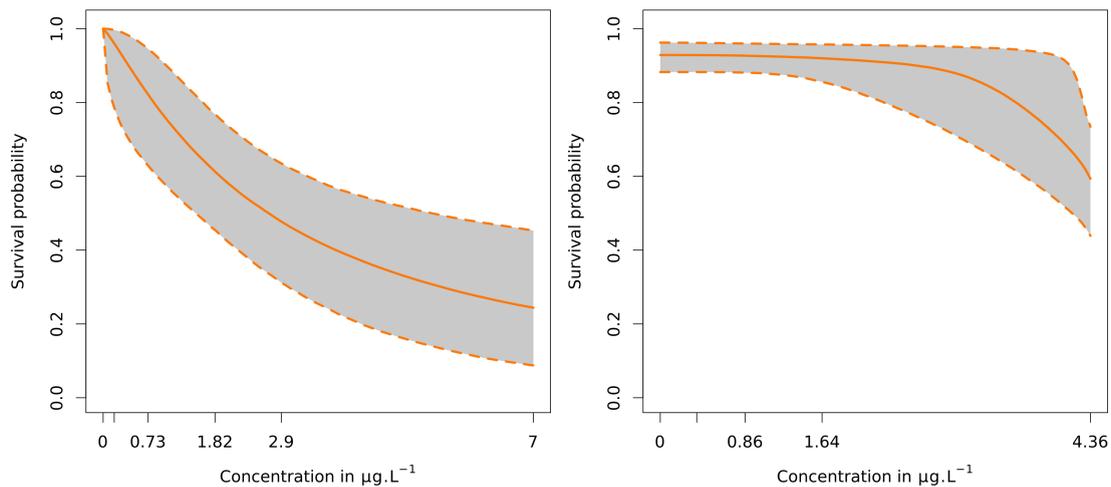
Figure 1: (left) Dose-response fit for *D. magna* survival exposed to chlordan; (right) Dose-response fit for *D. magna* survival exposed to cadmium.

> Parameter e has a different meaning here: $e = EC_{50}$ ($\neq LC_{50}$), that is the concentration that causes a 50% reduction of the reproduction rate compared to the control one.
> We get $EC_{50} = 1.77\,[1.37; 2.25]$ (expressed in $\mu g.L^{-1}$) what means that the reproduction endpoint is more sensible than the survival one to chlordan, despite the slight overlap of credible intervals for both parameter e (the LC_{50} with survival data, the EC_{50} for reproduction data.

Step 2. 3 Try then with example `cadmium1`. Compare the results with those obtained on survival data. Any remark?

> No warning here with parameter estimates, even for parameter e.
> *D. magna* was almost not sensible to cadmium when considering survival, but it is sensible for the reproduction endpoint.

Step 2. 4 When comparing sensitivity of *D. Magna* reproduction to chlordan and cadmium, do you get the same difference than the one obtained when comparing survival data?

> With reproduction data in data set `cadmium1`, we got $EC_{50} = 3.22\,[2.6; 4.08]$ meaning that the *D. magna* reproduction is more sensible to chlordan than to cadmium (as for survival).

Step 2. 5 If available, load your own data set (pay attention to the data format) and run the reproduction analysis. Pay attention to choose the appropriate concentration unit and Y-axis label. Save the report and discuss your results.

Step 2. 6 If you tried a third example, note below your remarks on the provided results

3 Quantitative continuous data analysis

Step 3.1 Go back to `http://mosaic.univ-lyon1.fr/` and choose menu `Growth`. This is the way to use MOSAIC at your office. During this training course, please use the pedagogical server which ensures that all participants can perform exercises at the time:

3

`http://umr5558-shiny.univ-lyon1.fr/mosaic-growth/`.

Step 3.2 If you wish, you can first visualize our demo video (duration: `08:39`), by simply clicking on the video link or directly here: `http://umr5558-shiny.univ-lyon1.fr/mosaic-growth/` `MOSAICgrowth.mp4`

Step 3.3 Data visualisation

- Try first with example `plant07.txt`. Notice that either a plot and a table are provided to check data have been correctly uploaded. Note that the user may personalize X- and Y-axis labels and choose the appropriate units for exposure, measurements and time.

> For example data set that are provided within MOSAIC, X- and Y-axis labels as well as units for exposure, measurements and time are automatically filled in. There is no need to change anything here, nevertheless you can personalize fields as you want.

- An innovative feature in this module is the possibility to upload several files **at the same time**. Try to upload these additional examples `chlordan-daphnia.txt`, `cadmium-daphnia.txt`, `subst01-lymnaea.txt`, `plant03.txt` and `plant10.txt`. What do you notice? Which data set is displayed? Try to change file.

> The list of selected files appears within the light grey box, while a warning is displayed below standing that some data files have no common time with the others. This means that a similar dose-response analysis (that is for the same exposure duration, also called the target time) will be not possible for these files simultaneously.
> Not selecting `subst01-lymnaea.txt` solves the issue; indeed, for this file the exposure duration is 56 days, while for all the others it is 21 days.
> Notice that by default the last uploaded data file is displayed, so here `plant10.txt`. The scrolling menu allows you to choose the data you want to visualize.

Step 3.4 Dose-response analysis

- Keeping several files uploaded, go to sub-menu `Dose-response analysis`. Run the analysis for example `plant07.txt`. What is the estimate of the EC_{50} (denoted ER_{50} within the module). Provides the median and the 95% credible interval.

> We get $ER_{50} = 685.05\,[511.67; 949.79]$.

- From goodness-of-fit results, get the correlation plot. What do you notice? Don't hesitate to read the help text below the figure.

> Parameter σ is not correlated to the three others (potatoid shapes of contour lines), while a more or less strong correlation appears between b, d and e (left-leaning ellipses).

- Ask for the calculation of ER_{25}, ER_{50} and ER_{75}. What do you notice in terms of precision of the estimates?

> When x increases, the precision of the ER_x estimates decreases, underlined by an increasingly large credible interval.

- Run additional analyses for examples `plant03.txt` and `plant10.txt`. Look at the ER_{50} estimate from file `plant10.txt`. What do you notice? How could you explain such a result?

> The ER_{50} estimate from file `plant10.txt` is very high, far from the highest tested concentration equal to 2000), and the estimate is very imprecise; a way to quantify the precision is to calculate a kind of coefficient of variation as $CV = \frac{Q97.5 - Q2.5}{median}$ that is expected to be less than 0.5: here it is equal to 1.3. This means that this ER_{50} estimate must be considered with caution, as it corresponds to an extrapolation far from the concentration range of the experiment.
>
> Such a disappointing result is due to the fact that the concentration range has not been chosen accordingly to the sensitivity of the plant species towards the substance to which it has been exposed. Indeed, it is a priori expected, when designing a toxicity test, to go from 0 to 100% of response or effect within the chosen concentration range.

- Is the PPC plot suitable enough to trust provided results for further ERA?

  > This question stands for file `plant10.txt`, where 100% of segments are green what translates a not enough correct fit of the model on these data.

- Ask the calculation of ER_{50} for `plant03.txt` and `plant10.txt`. Ask for the table displaying all ER_{50} for files `plant03.txt`, `plant07.txt` and `plant10.txt`. Which plant species is the most sensitive?

  > Regarding the ER_{50} estimates from the three files, `plant07` is the most sensitive.

Step 3.5 Downloads

- Go to sub-menu `Downloads`.
- For example `plant07.txt`, download the single report (choose the Word format).
- Download the text file of the joint posterior distribution.

  > A file entitled `outputPosterior_plant07.txt` is downloaded within the appropriate folder on your computer. This file is a text file (.txt) with four columns corresponding to the joint posterior distribution of the four model parameters: *b*, *d* and *e* for the deterministic part (the three-parameters log-logistic model), and σ for the stochastic part (here a normal distribution).

Step 3.6 Prediction tool

- Go to sub-menu `Prediction tool`.
- Enter a concentration range, for which parameter value of the three-parameters log-logistic model will be available. For example: `40; 80; 160; 320; 640`.

  > Note that fields `Target Time` and `Unit` ; the user should fill in them to keep in mind that the provided simulation will be made for a given set of parameter values that have previously been estimated at a given target time (namely the exposure duration, expressed in a given unit). Hence, the prediction is valid for a similar exposure duration.

- **As a first try**, choose non distributed parameters (default option).
- **Option (1)**: enter a single value for each parameter, based on your experience, based on expert knowledge or coming form the literature for example. Keep in mind that these values where obtained for a given exposure duration, chosen by the experimenter, usually equal to the duration of the experience (e.g., the standard duration

of 21 days for most of the toxicity tests performed under an OECD guideline). Try option (1) with parameter values entered by default as they are; please note that parameter b and e are expected to be given in log_{10}. What do you notice?

> With default parameter values, we can notice that the prediction (given as a median curve) goes from 0 to 100% of effect, meaning the the chosen concentration range will be suitable to test the toxicity of the species/substance combination of interest.

- Change the previous value of parameter e from $log_{10}(e) = 2.2$ to $log_{10}(e) = 3$. What do you notice?

> Changing parameter e from $log_{10}(e) = 2.2$ to $log_{10}(e) = 3$ provides a prediction for which the chosen concentration range leads to less than 50% of effect at the highest concentration. Hence, if this concentration range is kept as it is, it will not be possible to precisely estimate the ER_{50} once data will have been collected. It is thus strongly recommended to choose another concentration range.

- **Option (2)**: get median parameter estimates from a DR analysis that has already been performed on a data set (either with MOSAIC$_{growth}$ or with any other software), under experimental conditions that are close enough to those for which you want to make the prediction; in particular pay attention to simulate for a similar exposure duration. Try option (2) with parameter estimates obtained from `plant07.txt` data set. What do you notice?

> Parameter values should be around $b = 0.63$, $d = 3.46$ and $e = 685.05$; use a dot to separate decimals. Don't forget to convert b and e in log_{10}-scale.
> We obtain a predicted curve very similar to the median curve of the fitting plot obtained during the DR analysis from the `plant07.txt` data set; the difference is in the concentration range.

- Try again option (2), but with parameter estimates obtained from `plant10.txt` data set. Keep again the same concentration range. What do you notice?

> Parameter values should be around $b = 0.73$, $d = 8.63$ and $e = 6797.48$; use a dot to separate decimals. Don't forget to convert b and e in log_{10}-scale.
> As previously, we obtain a predicted curve very similar to the median curve of the fitting plot obtained during the DR analysis from the `plant07.txt` data set; the difference is in the concentration range.

- Based on the previous simulation, find a concentration range that would be more appropriate.

> A possible choice could be the following series: `800; 1600; 3200; 6400; 12800`, that allows to predict more than 50% of effect at the highest concentration.

- **As a second try**, choose **distributed parameters**. This necessary means that you previously downloaded a joint posterior distribution of the DR model parameters, what is today only possible with the MOSAIC$_{growth}$ module, or that you previously performed a DR analysis with MOSAIC$_{growth}$.

- **Option (1)**: coming back to the initially proposed concentration range (namely `40; 80; 160; 320; 640`), upload file `outputPosterior_plant07.txt` you may have already downloaded earlier. If not, go to option(2). Which difference(s) do you observe with previous predictions?

> Since parameters are now distributed according to the joint posterior distribution, the prediction is associated with a 95% credible band, giving an idea about the precision around the prediction.

- **Option (2)**: choose option `from a previous DR analysis`, that is supposed to have been performed on data set `plant07.txt` earlier. What do you notice?

> We get the same prediction as previously.

- Change the concentration range and see what happens.

> For example, taking the same concentration range obtained earlier, namely `800; 1600; 3200; 6400; 12800`, we predict a 50% of effect at concentration 12800 associated here with the precision of the prediction (quite large).

For further details on MOSAIC$_{\text{growth}}$, you can download and read the Tutorial available on line from here: `http://umr5558-shiny.univ-lyon1.fr/mosaic-growth/Tutorial.pdf`
A vignette also accompanies the Tutorial, with details on the modelling framework used in MOSAIC$_{\text{growth}}$: `http://umr5558-shiny.univ-lyon1.fr/mosaic-growth/vignette.pdf`.
Its reading is optional for the use of the application.