PROBIT GUIDELINES WITH R!!!

By: Juan Silva

Email: js3274@cornell.edu

1. Arrange your data before using R:

All data set can be stored in the provided Excel worksheet. However, keep in mind that all data sets must have these seven columns (as shown in Figure 1):

-insecticide: Always mark the insecticide tested for each bioassay

-strain: Label each strain with their corresponding name.

-dose: Write each of the corresponding doses tested on the bioassays and only use periods. Always use **controls with no insecticide**, and add them into the worksheet and use a dose value of zero ("0"). Look at the first (upper dose) shown in Figure 1.

-dead: Number of dead individuals tested per dose.

-total: Number of total individuals tested per dose

-date: Add the date you run the probit.

-color: Each strain must have a different number which results in a color code that help distinguish each strain in the probit plots.

Considerations:

- Notice that **all the column names are in lower case (including the first letter)** and R will only recognize them like that. Any change to this style will prevent R to use that column.
- It is recommended to "copy & paste" the information added to the first cell into all the other corresponding cells per column for the <u>insecticide</u>, <u>strain</u>, <u>date</u> and <u>color</u>. This is to be sure that all the information is exactly the same for each column and this avoids computing errors when running the script in R.
- Make sure you use periods "." instead of commas "," for numerical values because R will NOT understand commas as numerical values but as categorical.
- Make sure there are no spaces in any cell otherwise
- **Save as Tab delimited ".txt" file**

Figure 1. Example of data organization in Excel before using R.

2. Downloading R

You can download R Studio for Windows (Vista, 7, 8 & 10), Mac (OS X 10.6+ (64-bit)) and Ubuntu from the following link: (https://www.rstudio.com/products/rstudio/download3/).

3. Opening the Script for Probit Analysis

The script is fundamental to run the probit analysis with R because it contains all the commands required to obtain parameters such as LD_{50}, slope, etc. and statistical analysis (X^2 to determine if the data provided fit the probit model). To use this script, you need to open the file *Script for Probit R Analysis v2* located in the directory and open it with Rstudio. You can check the script is open in Rstudio as shown on Fig. 2.

Make sure the script has the extension ".R" (R File) before opening with Rstudio. If this extension is missing then Rstudio will not recognize this file as a script ("Run" button will not appear in Rstudio).
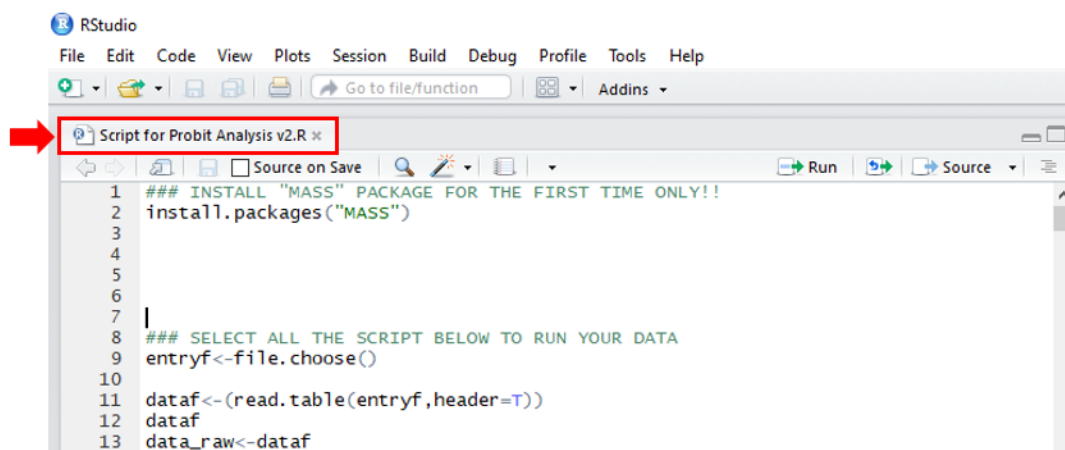

Figure 2. View of the script within Rstudio.

4. Running Probit analysis with R

When you run your data for the **first time, it is important to install the Modern Applied Statistics with S (MASS) Package**. This package uses the function *"dose.p()"* which allows

you to estimate each of the Lethal Concentrations (LC) after the program automatically runs the probit model with the input worksheet. Note that the calculations for LC_{50} or LD_{50} are the same and their meaning is interchangeable in this guideline.

You can easily install MASS directly from the script by selecting the text on row 2 (see purple arrow on Figure 3) and clicking the "Run" button (or pressing Control + Enter) as shown in Figure 3.
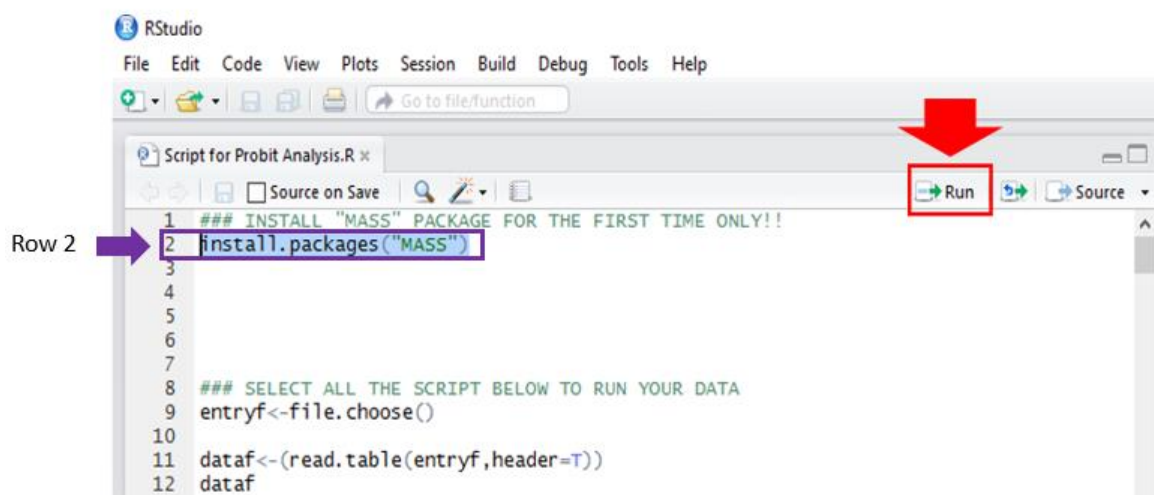

Figure 3. Installing MASS package.

After installing MASS, you can proceed to **SELECT ALL** the script from row 9 (see purple arrow and box on Fig. 4) down to row 230 and click the "Run" button (or press Control + Enter) as shown in Figure 4. The software will automatically open a window for you to select the tab delimited worksheet (".txt") you want to analyze. **Notice that for future probit analysis you only need to use this part of the script instead of installing MASS again.**
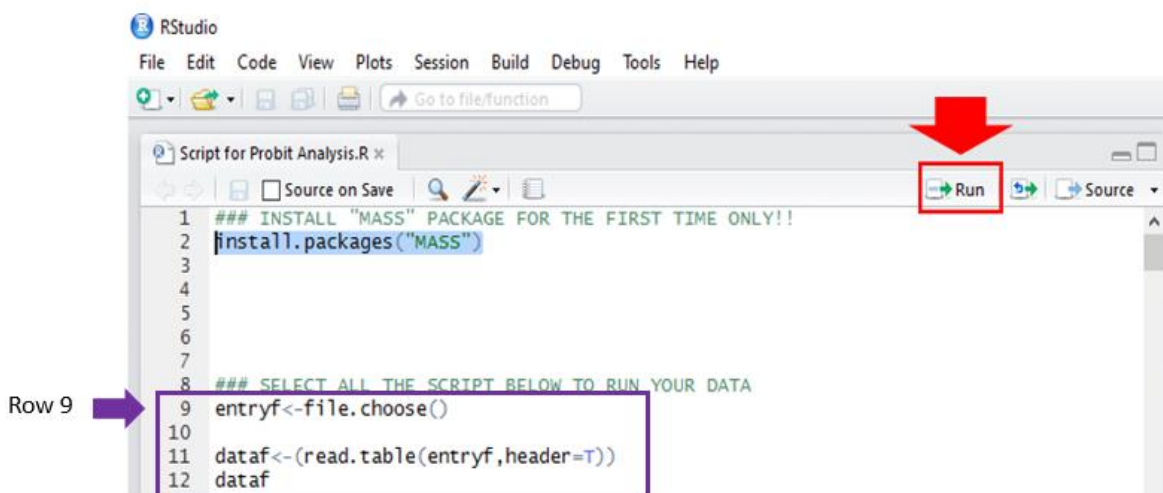
Figure 4. Running the probit analysis

Do not worry if a warning message appears in R once you run your data (see Figure 5). This is just a default warning to notify that matrices saved in the output file will include column names… (Yeah, that's weird).

```
> write.table(output,append = T, exporttab, sep = "\t", col.names=T,row.names = F,quote=
F)
Warning message:
In write.table(output, append = T, exporttab, sep = "\t", col.names = T,  :
  appending column names to file
```
Figure 5. Warning message after running the script.

After running your data with R, a new file will appear exactly in the same location where you saved your original worksheet ".txt". This new file will have the same name as the original worksheet but with the suffix "=Results" which means that your data has been analyzed. **Notice the analyzed data is saved as ".txt" as well.**

The output will include the following information:
- Control mortality
- Intercept, Slope and Slope Standard Error (SlopeSE)
- Equation of the line
- X^2 value and p-value (data fitted to the regression line)
- Dose, Mortality correction (%), Probit, Total, Dead, Dead expected, X^2 contribution
- LC (5, 10, 50, 90 & 95) and the corresponding Confidence Intervals at 95%.
- There is a new "Extreme values" column on the right side of X^2 that labels with a star ("*") those dead expected values which are dangerously low (below 5, based on X^2 assumptions) or high (expected mortality ≥ 95%).

Considerations:
- ➢ Output file now displays original data and % mortality before Abbott's correction
- ➢ There is now a warning message when mortality in the controls is equal 10%. ("WARNING: CONTROL MORTALITY EXCEEDING 10%"")
- ➢ Goodness of fit equation (Finney, p.66) applied for X^2 estimation is:

$$X^2 = \frac{\sum(r-nP)^2}{nPQ} \; ; \text{where}$$

r = Number of observed dead individuals
n = Total number of individuals tested
P = Proportion of dead expected
Q = Proportion of survivors expected
- ➢ Based on the X^2 value, a message notifies if data fit or do not fit the linear regression ("DATA REPRESENTED BY A LINE" or "DATA **NOT** REPRESENTED BY A LINE)

> ➢ There is now a "*** PLEASE ADD CONTROLS ***" warning message when controls are missing in the input worksheet.
> ➢ There is now an "INVALID POINT: MORTALITY LOWER OR EQUAL THAN CONTROL" warning message when a dose response is equal or lower than control mortality

---

Example for Probit Analysis

This example includes data from bioassays tested to a fictional strain of a certain insect that is kryptonite-susceptible.

```
EXAMPLE Data set Probit - Notepad
File  Edit  Format  View  Help
insecticide      strain  dose   dead   total   date        color
Kryptonite       SUPER   0      0      80      8/27/2016       1
Kryptonite       SUPER   0.98   3      80      8/27/2016       1
Kryptonite       SUPER   1.95   13     80      8/27/2016       1
Kryptonite       SUPER   3.9    54     80      8/27/2016       1
Kryptonite       SUPER   7.8    79     80      8/27/2016       1
```
Figure 6 Data saved as ".txt" before running the probit analysis.

After running the script, (as shown on Figure. 7).

```
EXAMPLE Data set Probit=Results - Notepad
File  Edit  Format  View  Help
                   SCOTT LAB - PROBIT ANALYSIS

Strain:         SUPER
Insecticide:    Kryptonite
Date:           8/27/2016


Control Mortality (%) = 0 ( 0 / 80 )

Original Data

Dose     Dead    Total    Observed Mortality (%)
0.98     3       80       3.75
1.95     13      80       16.25
3.9      54      80       67.5
7.8      79      80       98.75


Interc. Slope   SlopeSE
2.872   4.464   0.614

Equation of the line
Y = 2.872 + 4.464 * X

p(X2 = 4.606 ,df =  2 )= 0.1    DATA REPRESENTED BY A LINE


Dose    Mort.Corr(%)    Probit  Total   Dead    Dead Expected   X2      Extreme values
0.98    3.75            3.22    80      3       1.209           2.694   *
1.95    16.25           4.016   80      13      16.187          0.787
3.9     67.5            5.454   80      54      55.612          0.153
7.8     98.75           7.241   80      79      77.452          0.971   *


                Conf. Interv. 95%
LC      Value   Min     Max
5       1.283   0.979   1.683
10      1.548   1.229   1.949
50      2.998   2.584   3.478
90      5.807   4.597   7.335
95      7.004   5.323   9.216
```
Figure 7. Result output.

Acknowledgements

I want to thank Drs Jeff Scott, Michel Raymond and Pierrick Labbé for their assistance in developing this script.

References

Finney D. 1971**.** Probit Analysis, 3rd ed. Cambridge University Press, Cambridge, UK.

Milesi P, Pocquet N & Labbé P. 2013. BioRssay: a R script for bioassay analyses. http://www.isem.univ-montp2.fr/recherche/equipes/genomique-de-ladaptation/personnel/labbe-pierrick/

Ripley B, Venables B, Bates D, Hornik K, Gebhardt A & Firth D. 2014. Support functions and datasets for Venables and Ripley's MASS. R package version 7, pp. 3-35. http://www.CRAN.R- project.org/package=MASS