

## Research Methods in Weed Science: Statistics

Christian Ritz, Andrew R. Kniss, and Jens C. Streibig\*

**Nomenclature:** Bromoxynil; ioxynil; oxitril.

**Key words:** Analysis of covariance, analysis of variance, linear regressions, logistic regressions, nonlinear regressions.

### Introduction

There are various reasons for using statistics, but perhaps the most important is that the biological sciences are empirical sciences. There is always an element of variability that can only be dealt with by applying statistics. Essentially, statistics is a way to summarize the variability of data so that we can confidently say whether there is a difference among treatments or among regression parameters and tell others about the variability of the results. To that end, we must use the most appropriate statistics to get a “correct” picture of the experimental variability, and the best way of doing that is to report the size of the parameters or the means and their associated standard errors or confidence intervals. Simply declaring that the yields were 1 or 2 ton ha<sup>-1</sup> does not mean anything without associated standard errors for those yields. Another driving force is that no journal will accept publications without the data having been subjected to some kind of statistical analysis.

Weed science deals with controlling and managing unwanted plants in crops and other habitats. Thus, most of the research in weed science is about how to control plants mechanically, chemically, or by other means. The research questions encompass the array of complexity from simple site-of-action studies in a highly controlled environment to complex field and regional studies of vegetation management; from basic chemistry and genetics at the cellular level to whole-plant physiology studies; and community-level field studies, in which few

environmental factors can be managed, let alone managing the behavior of plant stands.

Because weed science is such a broad discipline, there are a variety of statistical methods researchers can choose, ranging from simple *t* tests to advanced multivariate statistics. It would be nearly impossible to discuss all the statistical methods used in our discipline in an article of a reasonable length. We hope this article can whet your interest in the use of proper statistics with weed science data. We are fully aware that we have offered only a preview of the numerous methods available, and we concentrate on the two largest groups of statistical methods employed in weed science research: ANOVA and regression techniques.

Although traditional ANOVA and linear regression share a common mathematical background, the clear difference between regression and ANOVA is quite evident based on the experimental design; regression deals with quantitative, explanatory variables, whereas ANOVA deals with categorical factors. One of the more underused statistical methods is the combination of ANOVA and regression, called *analysis of covariance* (ANCOVA).

Because the relationship between crop production and weed density is often quantified as the number of plants per unit area or as the relative cover, the relationship between crop yield and weed infestation calls for regression analysis. Likewise, herbicide studies often use a range of doses with a goal of determining the effect on weed control or crop yield and, therefore, should also be analyzed with regression analysis. Herbicides are designed to kill plants, and their selectivity is a matter of a complex suite of more-or-less interconnected phenomena related to the plant, the chemical and physical properties of the herbicide, and of course, the herbicide dose. ANOVA is central to many applications in agricultural research; indeed, agricultural examples are still used in most statistics textbooks to illustrate the utility of ANOVA. ANOVA has been, and will continue to be, an

DOI: 10.1614/WS-D-13-00159.1

\* Associate Professor, Department of Nutrition, Exercise and Sports, University of Copenhagen, Nørre Allé, DK-2200 Copenhagen N, Denmark; Associate Professor, Department of Plant Sciences, University of Wyoming, 3354 1000 E. University Avenue, Laramie, WY 82071; Professor, Department of Plant and Environmental Sciences, University of Copenhagen, Højbakkegaard, DK-2630 Taastrup, Denmark. Corresponding author's E-mail: jcs@plen.ku.dk

important method in weed science. However, ANOVA is not well suited to many weed science research questions, like unambiguously defining herbicide selectivity or crop yield loss in response to weed infestation. ANOVA is limited to separate factors and categories, such as unquantifiable control methods, e.g., weeds being exposed to some predefined management strategies, chemically, mechanically, or biologically.

Many other statistical methods are descriptive. Several of those descriptive statistics are multivariate statistics, such as principal component analysis, factor analysis, and correspondence analysis. These multivariate methods are useful in summarizing numerous response variables, for example, weed surveys and weed flora competition. In a way, many of these methods can be classified as hypothesis-generating methods, which only in rare instances can be used as hypothesis-testing methods. Many scientists and statisticians are well aware of the problems with testing normality of residuals and the assumptions of linearity in two- or three-dimensional space; those problems are exacerbated as additional dimensions are added.

The weed science literature has many examples of experiments that should have been analyzed with regression models but were instead analyzed with ANOVA, which is a statistical method for establishing whether there are any differences overall using the global  $F$  test: If the  $F$  test is significant, there are significant differences between some of the treatments. However, *post hoc* pairwise comparisons will be needed to sort out which differences are significantly different from zero. If, however, the significance is between responses affected by densities of weeds or doses of herbicides, the sheer differences of means can only be compared with other experiments with the very same weed densities or herbicide doses. If we do the experiment at different places or years, we might get completely different results, mainly because the dose in one location may not be the same effective rate at other locations. This fact is caused by the variability of responses in biology and is not a problem specific to weed science.

This article will not deal with the theoretical aspects of statistics or describe various alternative methods that can be used. There is an ample supply of articles along those lines (Cousens 1988; Onofri et al. 2010), but few on learning by example.

Based on the current status for use of statistics in weed science *sensu lato*, we have structured this article to fit our view on data analyses and their

importance in the discipline of weed science. All examples are reproducible using the R scripts provided in the Supplemental Appendix.

We only consider parametric methods because they enable quantification of effects through parameter estimates, which, we believe, is essential for evaluating any experiment and because they allow inclusion of additional information through additional explanatory variables. This is not easy to do with nonparametric methods (Stroup 2014).

**Experimental Design and Statistical Analysis.** To answer a research question in weed science you must, in most cases, be able to show data supporting your claim. To collect empirical evidence, various experimental designs are the cornerstones in agricultural research. Although the statistical method is a critical component of that process, no particular statistical method is the driving force of the experimental design; the process should be in the opposite direction. A clear research question should be developed, an experimental design should then be chosen that allows the research question to be tested, and that experimental design will then dictate the proper statistical analysis method or methods. Statistics should not be separated from the scientific inquiry process; rather, it should be included and considered in the initial stages of planning.

In addition to being designed in such a way that the research question of interest can be answered, usually through falsification of a null hypothesis, experiments have to be planned carefully to ensure that results reflect a general systematic feature or trend, that is, that they will be reproducible in the future, in similar experiments carried out under similar conditions. This often implies replicating the experiment at least twice, e.g., in two different growing seasons or at two different locations.

Well-designed experiments and replication over space or time may be well-known rules among weed researchers, it is much less common or well understood that the subsequent statistical analyses should be aligned with the experimental design used to generate those data. Often, but not always, the more complex the experimental design, the more complex is the corresponding statistical model. For example, you cannot analyze data from a randomized complete-block design without including a block factor in the statistical model considered. Simplifying the statistical analysis by some kind of preprocessing (e.g., taking averages) or subsetting of the data is *at best* a waste of data (and money/time

spent on collecting data) by making some conclusions unavailable and *at worst* results in misleading or noninformative conclusions driven more by the manipulations than by the data themselves. In many cases, data-reduction results in unreliable estimates of variability: smaller standard errors, smaller confidence intervals, and therefore, a greater likelihood of finding differences where none actually exist (type I error). With few exceptions, the statistical model should be based on the original raw data. If applicable, any data reduction should take place as part of (be motivated strongly by) the subsequent statistical analysis; rarely, will that be the case with ANOVA.

### Visualization

Whatever the design of an experiment, before statistical analysis, it is informative to do a graphical presentation of the data. Visualization of data is helpful to get a first impression of the quality of the data, but is particularly useful in the regression context to make sure that the relationships that are going to be analyzed support the initial idea of a linear or a nonlinear relationship. The choice of graphs could be based on the initial idea of the experiment and can just be ordinary plots with or without associated statistics, such as boxplot and bar plots. In the nonlinear case, it is important to see whether the intended nonlinear relationship initially considered is appropriate. The same applies to the response range and the predictor range in regressions.

### Regression Analysis

Without doubt, the most important method of summarizing research in weed science *sensu lato* is to use regression analyses: We use rates of herbicide, intensities, or depth for physical control, and we score the degree of weed infestation on yield loss. Figure 1 illustrates some of the important regression relationships in relation to weed science: three nonlinear regression curves and the linear regression line ( $x$  denotes the predictor;  $y$  is the response).

One important feature of regression analysis is that the values of the predictor as such are not of interest because they only serve as a means of estimating the regression line or curve. However, what is important is to try to ensure a reasonable spread in the  $x$  values to ensure that the line or curve can be determined from the data. Moreover, you do not need to have the same predictor values ( $x$ ) in

each treatment group to be able to compare them because it is not the mean response level at any given  $x$  that is of interest. Instead, the parameter of interest is the rate of change or some other quantity that is an intrinsic part of the regression line or nonlinear regression curve being estimated (Kniss et al. 2011; Ritz and Streibig 2005, 2008).

**Linear Regression.** Figure 2 illustrates two individually fitted regression lines showing the effect of herbicide rate on crop yield in two experiments. The slopes of these regression lines indicate how much the crop yield changes per unit increase in herbicide rate.

It is apparent from Figure 2 that the herbicide rates are not all the same for the two experiments. From a practical point of view, you can design experiments with different herbicide rates and still be able to compare slopes and intercepts. These comparisons of the rates of change would not be possible using ANOVA.

If one wishes to compare the slopes of the two regression lines to substantiate that one unit of herbicide rate results in the same increase in yield in the two experiments, one can compare the difference between slopes in view of the corresponding standard error of the difference. In this particular instance, the difference between the slopes is  $2.43 - 1.64 = 0.79$ , and the corresponding standard error of the difference is  $\sqrt{0.31^2 + 0.17^2} = 0.35$ . Obviously, there is small difference in the slopes of the two lines. From a weed-control point of view, this means that the herbicides do not give the same yield increase, when adding one unit of herbicide in both experiments. We leave it to the reader to calculate the difference in the intercept and its associated standard error and carry out the subsequent statistical inference. Doing regression analysis always implies the assumption that the relationship is linear. The best way to check that is so is to make graphs, such as in Figure 2, assessing the agreement between observed data points and the fitted regression line. Figure 2 confirms that the assumption of linearity holds (keep in mind that all model-checking procedures aim to establish that the statistical model used provides an adequate description; we are not looking for an exact match).

The data from a herbicide-efficacy experiment with common duckweed (*Lemna minor* L.) and a mixture of glyphosate and metsulfuron is illustrated in Figure 3 (Cedergreen et al. 2007). The linear relationship seems appropriate within a narrow dose range (Figure 3a). There is no systematic pattern

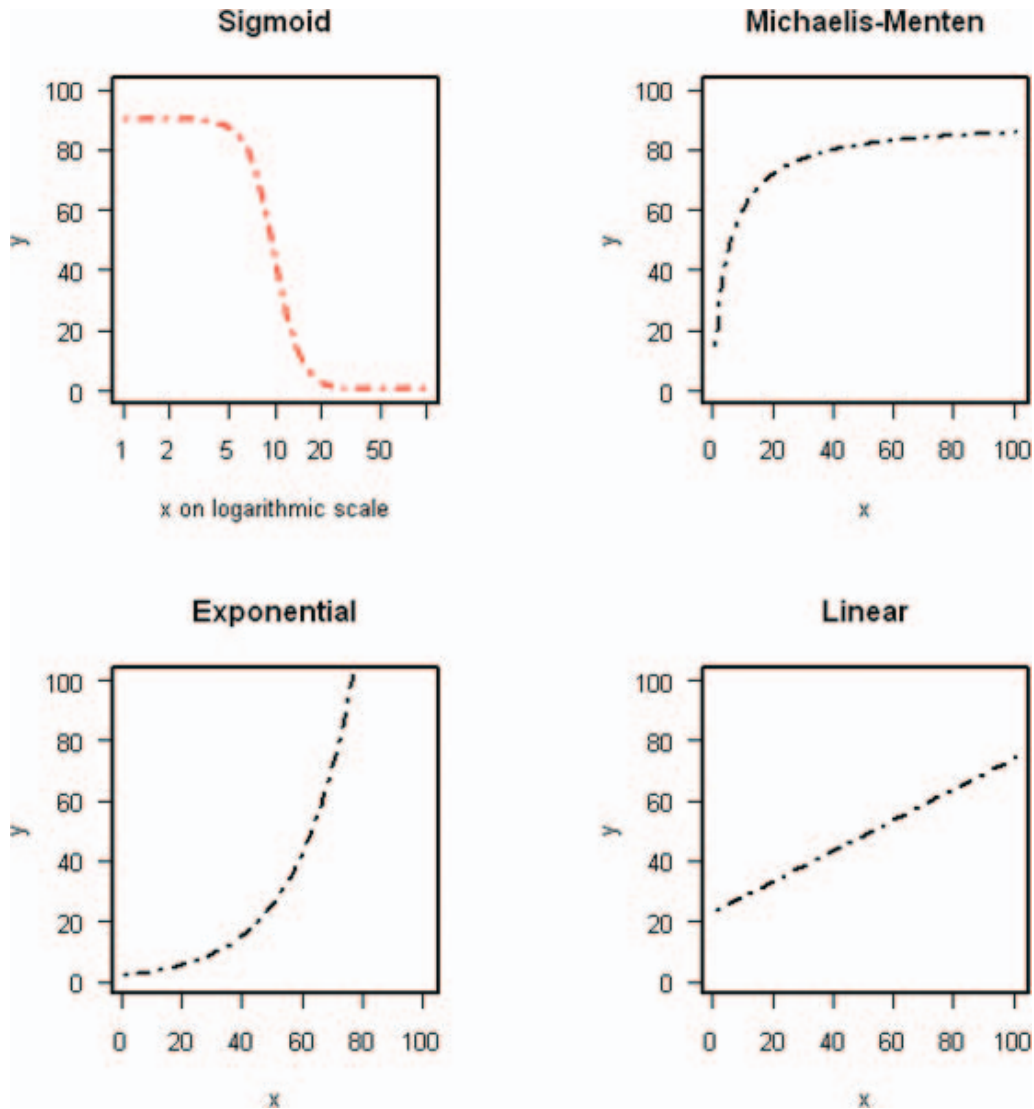


Figure 1. Four common regression relationships in weed science.

in the distribution of the residuals around the horizontal line of zero (Figure 3b). A residual is the observed response value minus the predicted response value from the fitted-regression model. Residuals should be randomly scattered around the  $x$ -axis. Because we have replicates of herbicide doses, we can make a test for lack of fit, which compares the residual sums of squares from the regression analysis with the residual sums of squares of the more-general one-way ANOVA model, where the herbicide rates are treated as categorical variables and not as continuous variables. The test for lack of fit of the regressions in Figure 3 is shown in Table 1, indicating that the data (Figure 3a) can be described properly by the regression. However, when the range of rates is extended, the test for lack of fit becomes highly significant (Table 1; Figures 3c and 3d). Consequently, we must reject the hypothesis

of linear relationships in Figure 3c. The test for lack of fit, however, can only be used if there are replicates (in Figure 3, there are two replications per herbicide rate).

When simple linear regression involves only a single predictor, the analyses are straightforward. In principle, the same considerations are applicable to the case of multiple linear regression involving two or more predictor variables. For these models, it is imperative to use residual plots to ensure that the model is describing the variation in data adequately. It may be possible to display the fitted regression plane in three dimensions, but for more than two predictors it is not possible. For a multiple linear regression with two predictors, the interpretation of the parameters is as follows: The regression slope for the first predictor,  $x$ , expresses predicted changes of  $y$  per unit  $x$  when the other predictor,  $z$ , is fixed, and



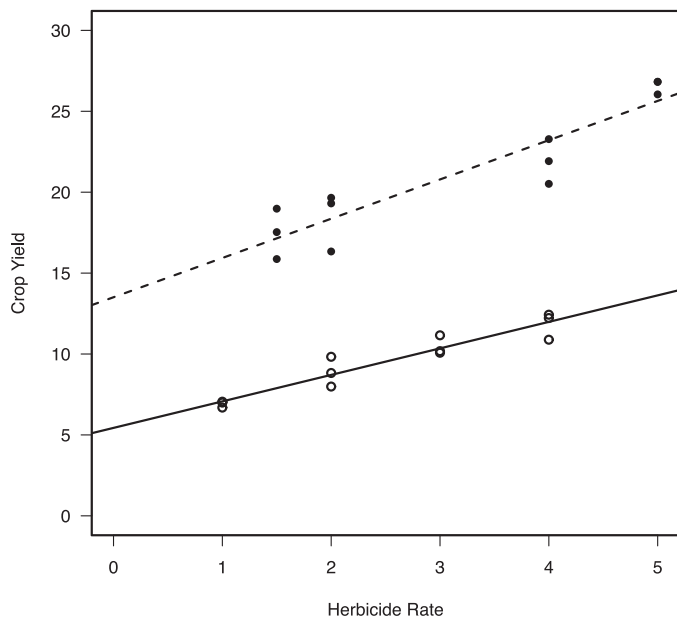


Figure 2. Two regression lines for crop yield in response to the herbicide rate from two independent experiments with the same herbicide. The slope of the continuous line is  $1.79 (\pm 0.31)$ , and the broken line is  $2.23 (\pm 0.43)$ . The intercepts are  $5.4 (\pm 0.9)$  and  $13.5 (\pm 1.4)$  (artificial data).

vice versa with the predictor  $z$ . The tricky point here is that we have to ensure that the two predictors affect the response independent of each other; that is, that there is no (or at least negligible) collinearity.

A field experiment with barley in response to increasing tribenuron-methyl and nitrogen rates was analyzed by a multiple linear regression analyses. The research question was to find out whether the herbicide interacted with the nitrogen supply (Figure 4).

The regression intercept of  $4.2 \text{ t ha}^{-1}$  was the yield without nitrogen or herbicide. An increment of one unit of nitrogen at any fixed herbicide dose increased the yield by  $0.03 \text{ t ha}^{-1}$ , and the increase in one unit of herbicide at any fixed nitrogen level increased the yield by  $0.18 \text{ t ha}^{-1}$ .

*Polynomials.* Sometimes linear regression falls short of describing the relationship in the data. In such cases, polynomials may be considered as a last resort because they can flexibly capture any curvilinear trend; it is simply a matter of choosing a sufficiently high order. The dose–response data in Figure 5 cannot be satisfactorily described by either a second-order polynomial ( $y = \text{Intercept} + ax + bx^2$ ) or a third-order polynomial ( $y = \text{Intercept} + ax + bx^2 + cx^3$ ) (Streibig 1983). Usually, you have to be careful with augmenting the degree of polynomials. A general rule is to never go beyond a third-degree

polynomial. When using a polynomial, make sure that interpolations between measurements make sense; sometimes, high-order polynomials produce a strange behaviors within the data, between predictor values, for instance. Even a second-order polynomial can produce illogical predictions, as seen in Figure 5. Another problem with polynomials is that you cannot always easily extract biological meaning from the model parameters, as compared with a linear regression. Inverse polynomials were sometimes preferred over ordinary polynomials for describing data in weed science (Nelder 1966), but they have become less common because personal computing power allows nonlinear regression models to be used.

Figure 5a shows that neither a second-order nor a third-order polynomial is very good to describe the data between observations, and interpolation of predicted fresh weights becomes negative at some point. Note also that, when using polynomial regression models, the residuals plots may look all right, but interpolation may still be a problem because that interpolation is usually carried out for data that were not used to fit the model. The behavior of the assumed regression model between observations must be biologically meaningful. Another way to fit the curve is to use  $\log(\text{Fresh wt})$  as the response in a second-order polynomial regression model. In Figure 5b, that fit is obviously much better for predicting fresh weight within the Oxytril interval, but extrapolation beyond the dose range is not advisable.

**Nonlinear Regression.** Figure 1 illustrates three nonlinear regression models: exponential, Michaelis-Menten, and the sigmoid curves. There is a Babylonian confusion of nonlinear regression models, but generally, there are few of interest and in current use, and those few are indispensable for weed scientists.

*Exponential Curves.* An exponential function could take the following form:

$$y = ce^{kx} \quad [1]$$

Where  $c$  is the initial response  $y$  when  $x$  is zero, and  $k$  is the rate constant. For the logarithm of  $y$ , the functions turn into straight-line relationships:  $\log(y) = \log(c) + kx$ .

Depending on the sign of the rate parameter  $k$ , the curve either increases or decreases, characterizing the growth of plants (positive  $k$ ) or degradation

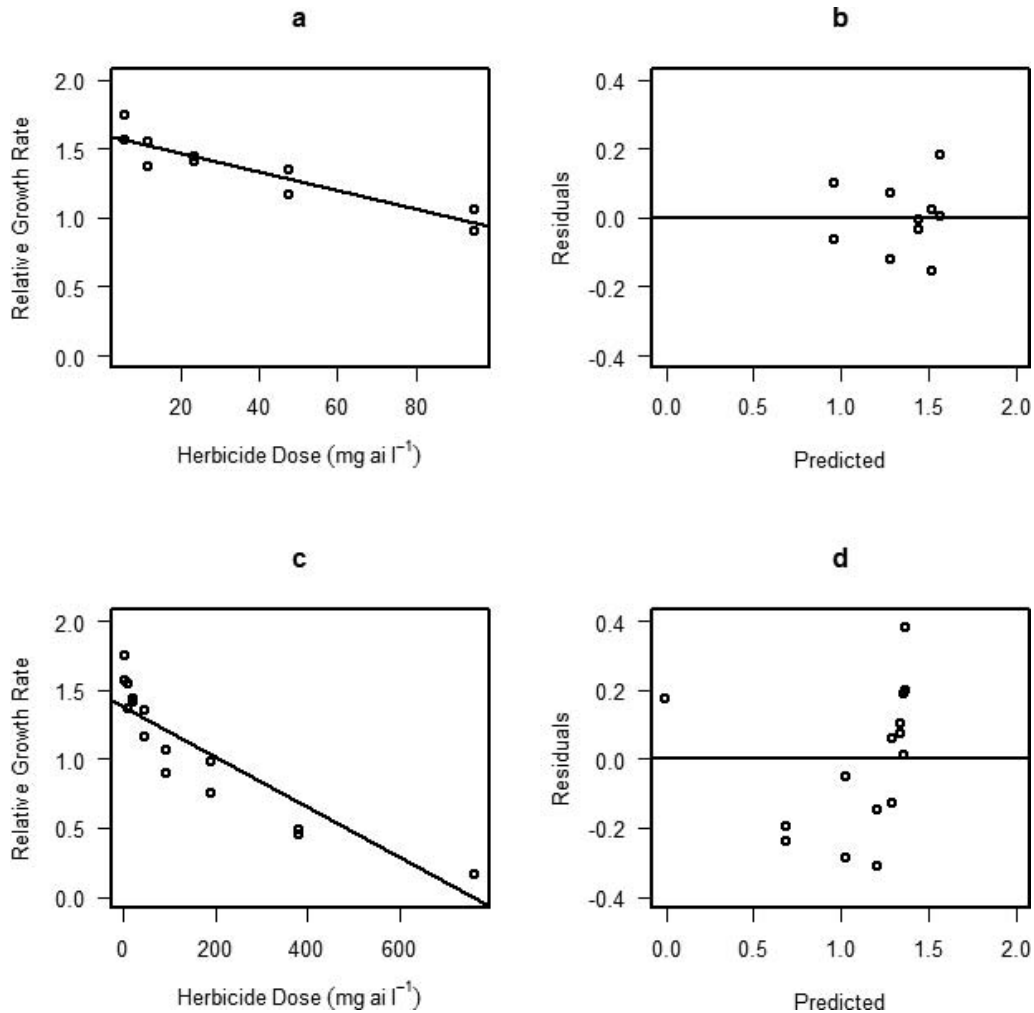


Figure 3. The relative growth rate of common duckweed in a standardized medium, according to standard guidelines from the Organisation for Economic Co-operation and Development for toxicity tests, in response to a mixture of glyphosate and metsulfuron. Data are satisfactorily described by the linear model (intercept =  $1.60 \pm 0.05$ ; slope =  $-0.007 \pm 0.001$ ), and the distribution of residuals was not a systematic distribution around the horizontal line of zero in *b*. By extending the dose range, the linear model felt short of describing the data within an acceptable limit, and the residual plot first shows an underestimate and then overestimates the data. (Cedergreen et al. 2007).

(negative  $k$ ) of compounds in living organism or in the environment. The parameter  $k$  denotes the relative rate of change, which is an important biological parameter and can be compared among

experiments. It is used to estimate the half-life of a degrading compound in soil, plants, and animals.

The exponential growth curve in Figure 1 does not have an upper limit; the lower limit  $c$  at  $x$  equals

Table 1. The test for lack of fit. The ANOVA model is the most-general model and does not assume either a linear or nonlinear relationship between the relative growth rate and the herbicide dose. The linear regression does assume a linear relationship, and if the lack of fit is significant, then the relationship is not linear.<sup>a</sup>

Combined residuals	Regression <sup>b</sup>		Regression <sup>c</sup>	
	df	Sum of squares	df	Sum of squares
Regression Residual	8	0.092	14	0.61
ANOVA Residual	5	0.064	8	0.09
Lack of fit	3	P = 0.56 (NS)	6	P = 0.006**

<sup>a</sup> Abbreviations: df, degrees of freedom; NS, not significant.

<sup>b</sup> The data for this regression comes from Figure 3a.

<sup>c</sup> The data for this regression comes from Figure 3c.

\*\* P ≤ 0.01.

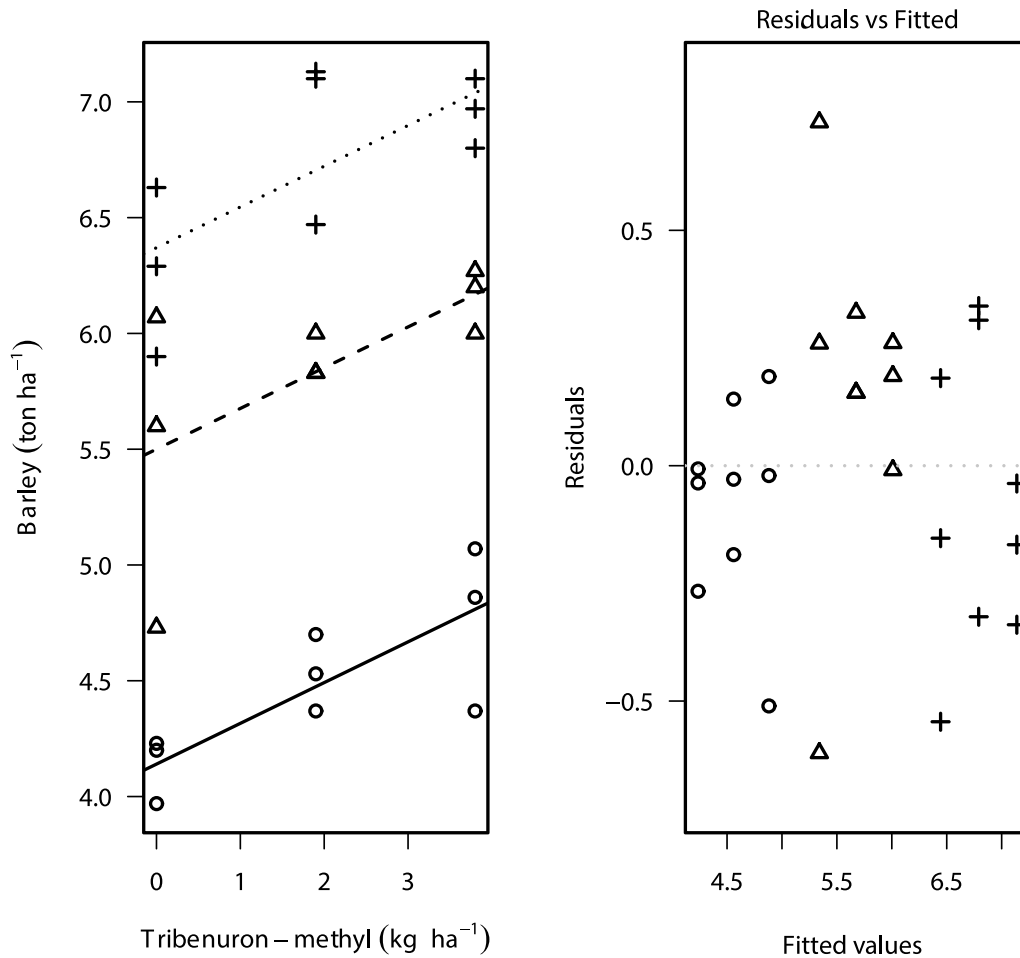


Figure 4. Illustration of a multiple regression with increased doses of a herbicide and nitrogen (left, continuous line = no nitrogen; broken line = 50 kg N ha<sup>-1</sup>; punctured line = 100 kg N ha<sup>-1</sup>) and the residual plot (right). Intercept was 4.2 ( $\pm$  0.12), the slope for N was 0.03 ( $\pm$  0.002), and the slope for the herbicide was 0.18 ( $\pm$  0.04). The results are shown by fixing the N in the right-hand graph. A model with the interaction between N and the herbicides did not show any significant lack of fit ( $P = 0.55$ ) (KAB Davidsen, unpublished data).

zero. That means that extrapolation beyond the range of the predictor for plant growth is unwarranted, exactly the same way it was for the extrapolation in Figure 5.

The decay of herbicides can usually be described by an exponential decay curve (as in Equation 1), but with a negative  $k$ . The result of such a fit is shown in Figure 6. Of particular interest for the degradation of herbicides is the half-life ( $t_{1/2}$ ). The half-life refers to the time when  $y$  is halved, relative to the level of the initial concentration  $c$ . It means that  $0.5 = e^{-kx}$  and hence  $x = [\log(0.5)]/-k$ .

*Michaelis-Menten Models.* The Michaelis-Menten function is commonly used in enzymology and biochemistry. In weed science, it is better known as the *inverse hyperbolic curve*, which describes the yield loss of a crop as a function of the weed infestation in terms of the density or leaf cover (Cousens 1985).

The Michaelis-Menten function comes in many disguises, but basically, Equation 2 describes a relationship where the response  $y$  starts out of the origin (0, 0) and increases until an upper limit  $d$  is reached;  $x$  is the independent variable, and  $e$  is the  $x$  value where  $y$  is halfway between zero and the upper limit  $d$ :

$$y = (dx)/(x + e) \quad [2]$$

If  $e$  is small, the initial slope of the curve will be steep and vice versa. Figure 7 illustrates the fit of the growth rate as a function of nitrogen.

Equation 2 is commonly linearized, which is a relic from the era of low-speed computers, as shown in Equation 3:

$$(1/y) = (1/d) + (e/d) \times (1/x) \quad [3]$$

As was the case with the exponential, we have to derive the parameters and their associated standard

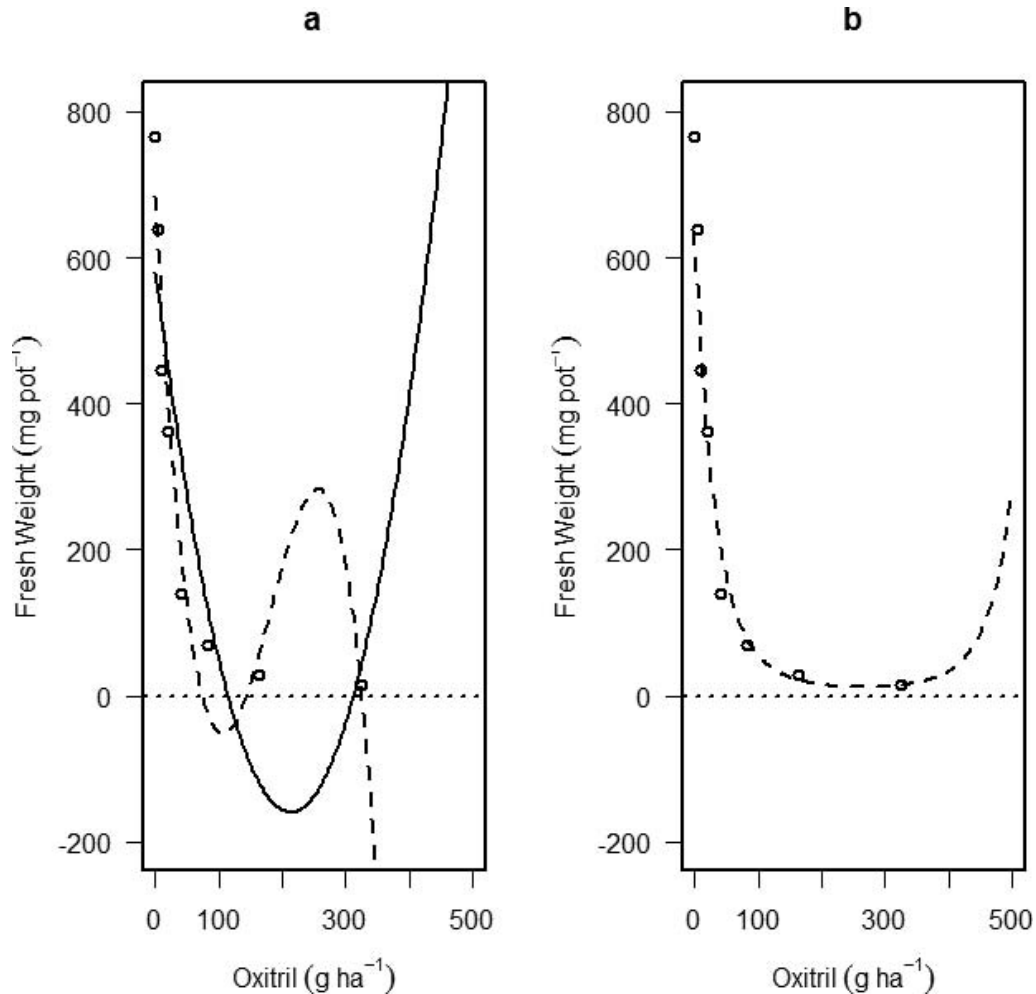


Figure 5. A dose–response curve fitted with a second-degree (continuous line) and a third-degree polynomial (broken line) (a), and with a second degree polynomial (b) on the logarithm of the fresh weight with the predicted values. The data from both were back-transformed (Streibig 1983).

errors. The nonlinear and linear fits are seen in Figure 7.

There is some discrepancy between the parameter estimates for the original Equation 2 and the linearized Equation 3. By comparing the rate of nitrogen uptake in response to the nitrogen concentration in a hydroponic solution (Cedergreen and Madsen 2002), the curve on the original scale and the one after linearization show that the latter did not catch the upper limit very well (Figure 8). That probably reflects the discrepancy between the concentration of nitrogen necessary to yield 50% uptake ( $e$ ) of  $17 \pm 3$  for Equation 2 and an  $e = 26$ , as derived from Equation 3.

The model is used in weed–crop competition studies and in yield–loss descriptions with a well-known parameterization:

$$y = Ix / [1 + (Ix/d)] \quad [4]$$

Where  $d$  denotes the maximum yield loss as a percentage of the yield in the weed-free environ-

ment  $y$ , and  $I$  is the initial slope of the straight-line part of the curve close to 0,0.

The function of Equation 4 is to describe the yield loss as a percentage of the yield in weed-free plot. However, weed-free environments are difficult to find in practice, so, depending on the distribution of actual yields in response to weed density, an estimated weed-free yield can be obtained by using the following reparameterization of Equation 4:

$$y = Y_0(1 - \{Ix/100[1 + (Ix/d)]\}) \quad [5]$$

where  $Y_0$  denotes the predicted yield for the weed-free environment. An example of this reparameterization is shown in Figure 9 (Streibig et al. 1989).

A recent literature review found that herbicide-absorption studies in the weed science literature have been analyzed using a variety of disparate (and sometimes inappropriate) methods, making comparisons between studies difficult (Kniss et al. 2011). After reanalysis of many herbicide-absorption studies, another parameterization of the Michaelis-



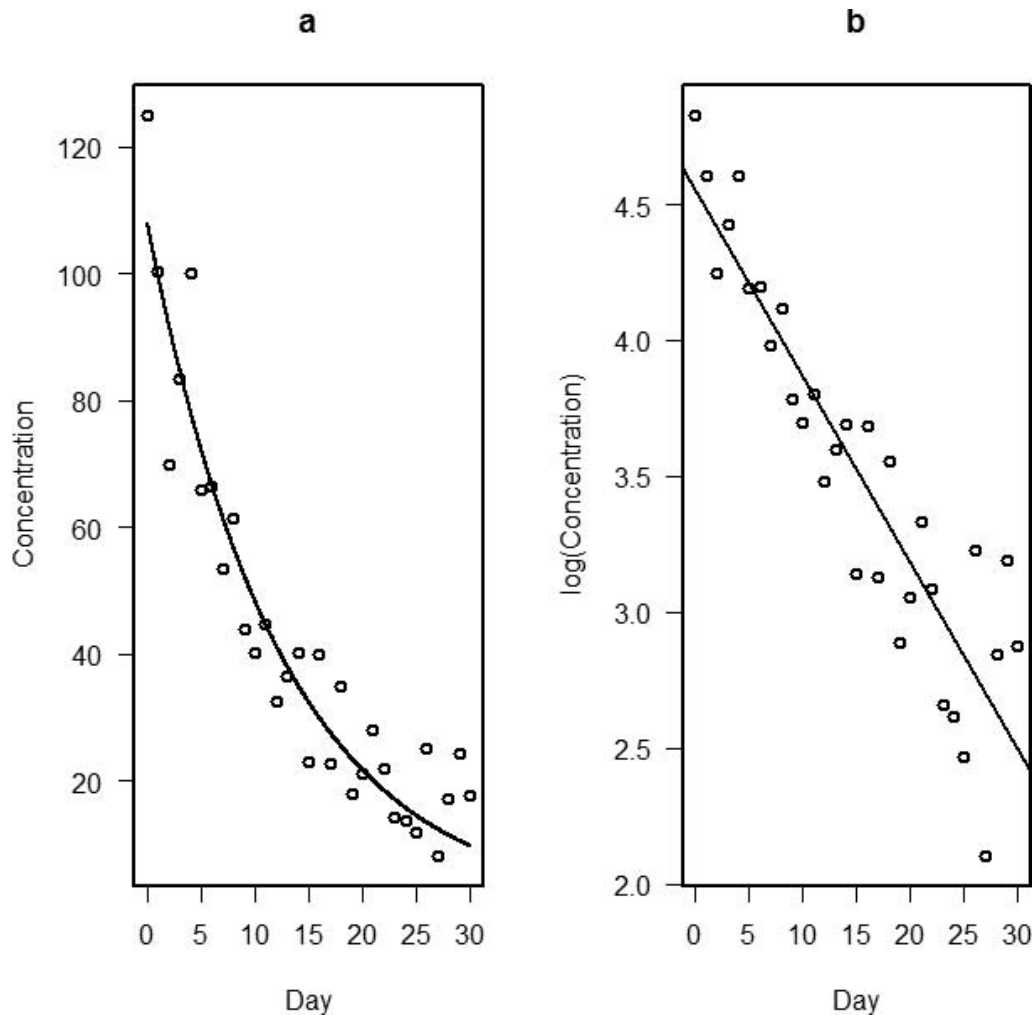


Figure 6. The decay curve for degradation of a herbicide in the soil from Equation 1. The initial concentration was  $108 (\pm 5)$ ,  $b = -0.08 (\pm 0.006)$ , and the  $t_{1/2}$  was 8.7 d. The curve is fitted with the nonlinear function in R, which means that a fairly accurate guess on the initial parameters must be estimated (in the Appendix, the R code it is shown as an example of how to approach this problem). For the linear fit,  $c = \exp(4.55) = 94.6$  and the  $t_{1/2}$  was 10 d. Please note the calculation of the standard errors is rather involved when using the linearization. (a) In the fit, we can use a so-called delta method to calculate the standard errors (see the R code for Figure 6).

Menten model was proposed to standardize the analysis of herbicide-absorption studies. The model is a simple modification of Equation 2, where a constant was added to the denominator, which provides a more meaningful parameter estimate in the herbicide-absorption context (Equation 6):

$$y = dx / (0.11 \times t_{90} + x) \quad [6]$$

where  $d$  is the maximum asymptote or maximum expected herbicide absorption, and  $t_{90}$  is the time required for 90% of maximum absorption to occur. R code (R Foundation for Statistical Computing, Wien, Austria) for this model is provided in the original article (Kniss et al. 2011).

*Sigmoid Curves.* Sigmoid curves are widely used in toxicology, ecotoxicology, pharmacology, and pesticide

science, particularly the log-logistic model (Equation 7) introduced to many weed scientists by Seefeldt et al. (1995) and Streibig et al. (1993) (Figure 1). In a way, sigmoid curves can be considered an extension of the exponential model and the Michaelis-Menten model considered above. In practice

$$y = c + [(d - c) / (1 + \exp\{b[\log(x) - \log(ED_{50})]\})] \\ = c + \left\{ (d - c) / \left[ 1 + (x / ED_{50})^b \right] \right\} \quad [7]$$

where  $c$  is the lower limit of  $y$ ,  $d$  the upper limit,  $ED_{50}$  is the  $x$  that yields response  $y$  halfway between  $d$  and  $c$ . Finally,  $b$  is the relative slope. A thorough description of the nonlinear models mentioned above is given elsewhere (Ritz 2010). The models in Equation 7 are similar and are only defined for doses greater than zero.

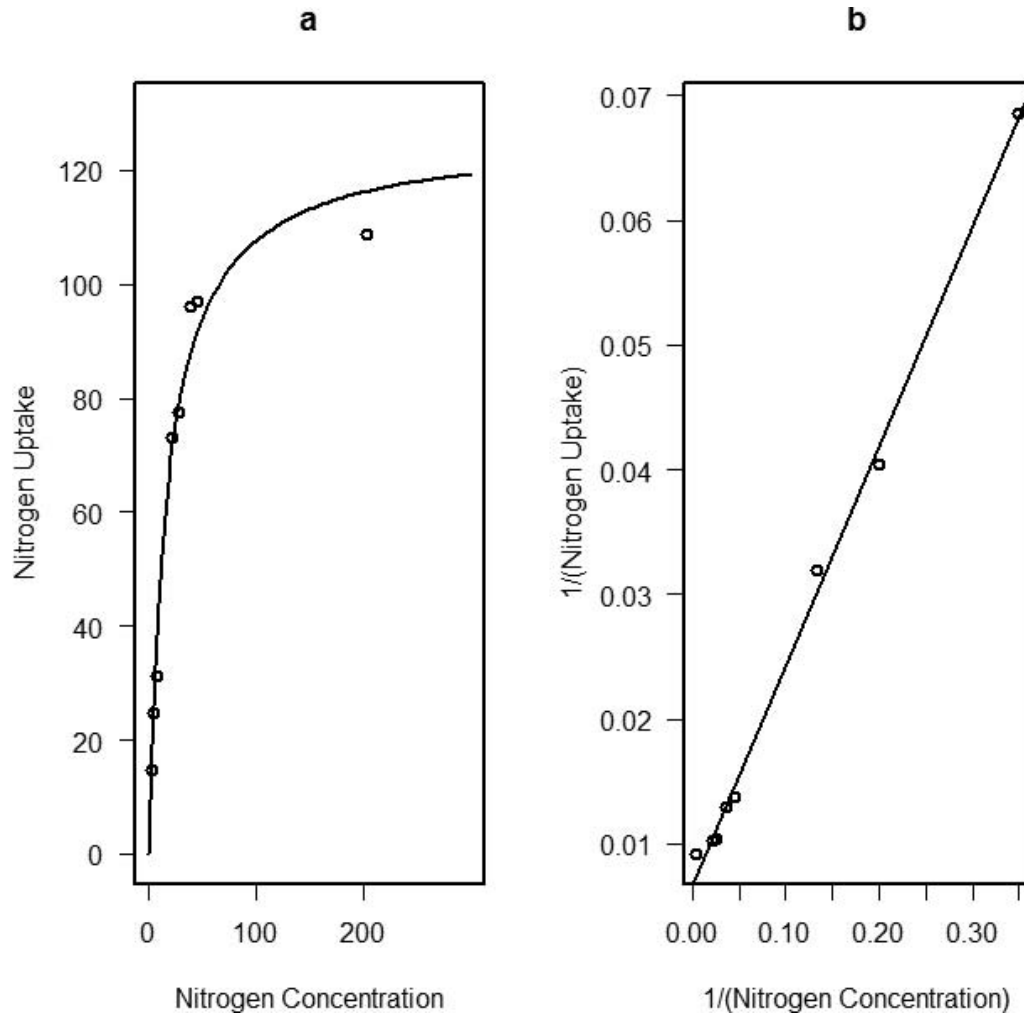


Figure 7. Uptake of nitrogen as a function of the initial concentration of nitrogen in the growth medium for common duckweed (Cedergreen and Madsen 2002). (a) The parameters for Equation 2) are  $d = 126 (\pm 7)$ , and  $e = 17 (\pm 3)$ . (b) For the linearization using Equation 3, the derived parameters are  $d = 1/0.0067 = 149$ , and  $e = 148 \times 0.175 = 26$ . As mentioned in Figure 6, the standard errors are rather difficult to obtain from these linearized curves.

The log-logistic model has been extensively used to describe herbicide effects, but it is also appropriate for germination curves and numerous other biological phenomena. In contrast to polynomial linear regression, the parameters of the log-logistic curves have biological meaning that can be used to compare independent experiments.

The log-logistic model (Equation 7) is symmetric around its inflexion point, which is at the  $ED_{50}$ . Sometimes symmetry does not apply, as explained by Ritz (2010), and there are asymmetric alternatives, where the point of infection is not halfway between the upper and lower limit; in such cases, the  $ED_{50}$  must be derived from the curve after the fitting process.

The parameters in Figure 10 (using same data as in Figure 5) have biological meaning, and the  $ED_{50}$  provides the potency of the herbicide. The relative slope  $b$  is more intangible, but it can be useful when

comparing several response curves. Our experience is that if the relative slope  $b$  is larger than 7 for biomass responses, it would be a good idea to check the fit. *Large*, however, in this context, is rather vague if you do not associate it with standard errors. The lower limit of  $-6.5$  (with a huge standard error of 32) in Figure 10 is not realistic because fresh weight cannot be negative. In such cases, it would be more reasonable to use a three-parameter log-logistic curve where the lower limit is zero:

$$y = d / (1 + \exp\{b[\log(x) - \log(ED_{50})]\}) \quad [8]$$

By doing this, we get the same  $ED_{50}$  ( $15.6 \pm 2.22$ ) as we did with Equation 7, but sometimes, if the lower limit in the four-parameter model is negative, the  $ED_{50}$  and its associated standard error may change dramatically.

$ED_{50}$  has an important role in toxicology and is the most-precise parameter for comparing potencies

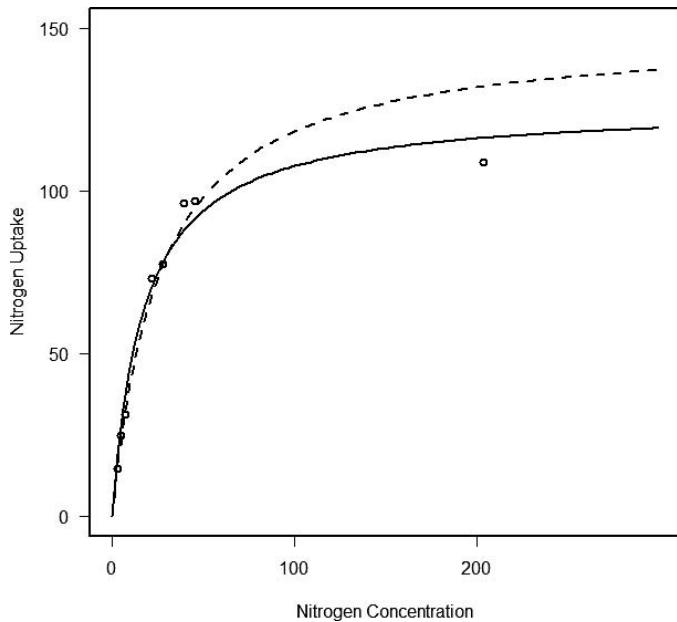


Figure 8. The continuous line is the regression fit from Equation 2; the broken line is derived from the parameters in Equation 3. The upper limit is much higher than the original one, and that has implication for the concentration yielding half-uptake rate (see text) (Cedergreen and Madsen 2002).

of toxic compounds. In weed science, however, we are not often interested in the  $ED_{50}$ , but instead, in the  $ED_{10}$  to assess crop tolerance, which is also an important parameter in toxicology, and in the  $ED_{90}$  to assess weed control. These  $ED$  levels can be found after fitting (Ritz 2010). For the three-parameter

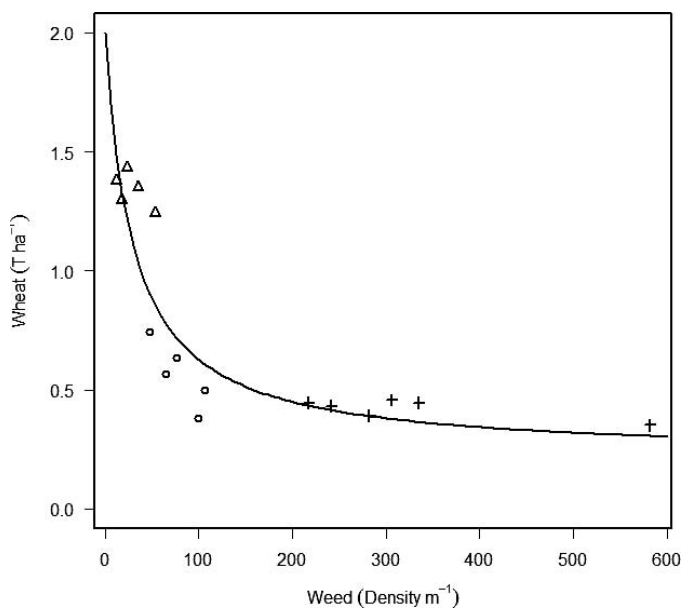


Figure 9. For a fit using Equation 5, the intercept is 2.0 ( $\pm 0.5$ ), the upper limit of the yield loss percentage  $d$  is 89 ( $\pm 5.6$ ) at infinite weed density, and the initial slope  $I$  is 0.03 ( $\pm 0.02$ ). Different symbols denotes different fields (Streibig et al. 1989).

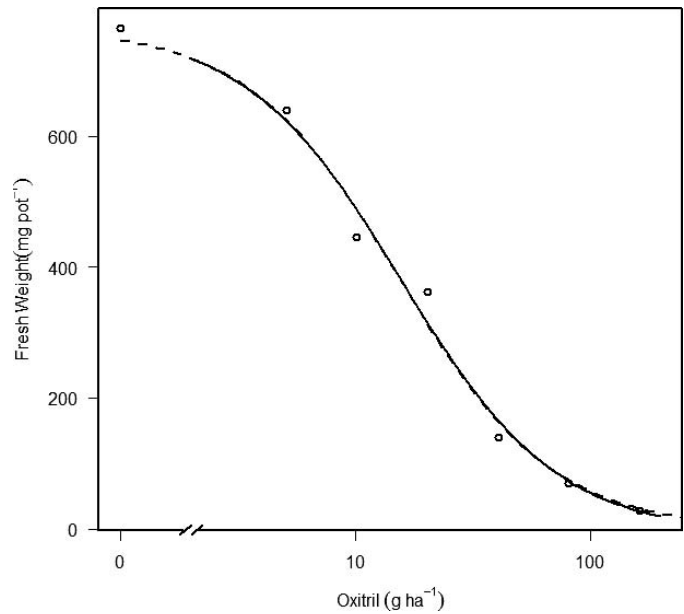


Figure 10. The four-parameter dose-response curve (continuous line) fitted to the data from Figure 5. The upper limit  $d$  is 766 ( $\pm 35$ ), the lower limit  $c$  is  $-6.4$  ( $\pm 32$ ), the slope around  $ED_{50}$   $b$  is 1.31 ( $\pm 0.21$ ), and  $ED_{50}$  is 15.6 ( $\pm 2.22$ ). The dose axis is on a logarithmic scale to illustrate that the curve is symmetric around  $ED_{50}$ .

log-logistic curve (Equation 8) the  $ED_{90}$  is 79 ( $\pm 13$ ). If the illogical negative lower limit  $c$  of  $-6.5$  is kept, the  $ED_{90}$  would have been 82 ( $\pm 26$ ), which is not a dramatic difference in this particular instance.

*Dose-Response and Selectivity.* Herbicides are designed to kill plants and their selectivity is a matter of dose. In herbicide research and development, the selectivity is often measured by the various potency estimates of the compound in various species. Often, we wish to compare various herbicides on the same plant species or the same herbicides on different plant species. That means we have to compare dose-response curves. Herbicide action and their selectivity in already emerged crops is a matter of a complex suite of more-or-less interconnected phenomena related to the plant, the chemical, and physical properties of the herbicides. The principles in toxicology and pharmacology do apply to herbicide research by classifying toxicity according to some standardized methods with sigmoid dose-response models. After World War II, there were several articles on how to assess the toxicity of herbicides in plants (Blackman et al. 1951; Woodford 1950), and now, some of those models are being applied easily, thanks to the seamless fitting of nonlinear response curves with high-speed computers.

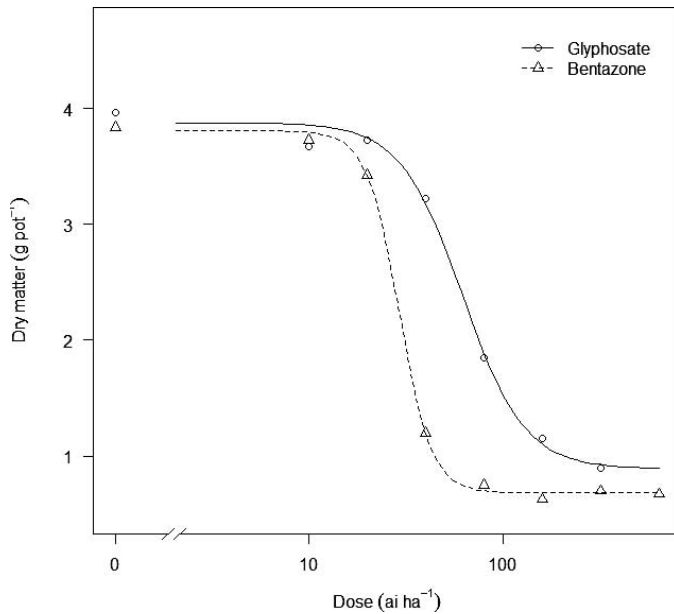


Figure 11. Four-parameter dose–response curves for two herbicides and white mustard (*Sinapis alba* L.) (Christensen et al. 2003).

A bioassay with one plant species and two herbicides—bentazon and glyphosate—is illustrated in Figure 11 and in Table 2 (Christensen et al. 2003). Comparison of the two regressions can be performed by comparing the  $ED_{50}$  values, which can be summarized in one figure as the relative potency ( $R = [ED_{50}(\text{bentazon})/ED_{50}(\text{glyphosate})]$ ) between the two herbicides, which is 2.12 ( $\pm 0.27$ ). The relative potency denotes the biological exchange rate between the compounds, meaning we need 2.12 times more glyphosate to get the same effect as we do with bentazon. The 95% confidence interval is between 1.6 and 2.7. As mentioned earlier, we might be more interested in the relative potency at  $ED_{90}$ , which is 3.1 and with the 95% confidence interval lying between 1.3 and 4.9. Note the relative potency is *not* independent of the response level because the slopes are different from

Table 2. Regression parameters for the log-logistic fit for white mustard biomass (see Figure 11).<sup>a</sup>

Parameters	Estimate	SE
Relative slope, glyphosate	2.72	0.75
Relative slope, bentazon	5.13	1.13
Lower limit, glyphosate	0.89	0.19
Lower limit, bentazon	0.68	0.10
Upper limit, glyphosate	3.88	0.11
Upper limit, bentazon	3.81	0.11
$ED_{50}$ , glyphosate	62.09	6.61
$ED_{50}$ , bentazon	29.27	2.24

<sup>a</sup> Abbreviations: SE, standard error;  $ED_{50}$ , 50% effective dose.

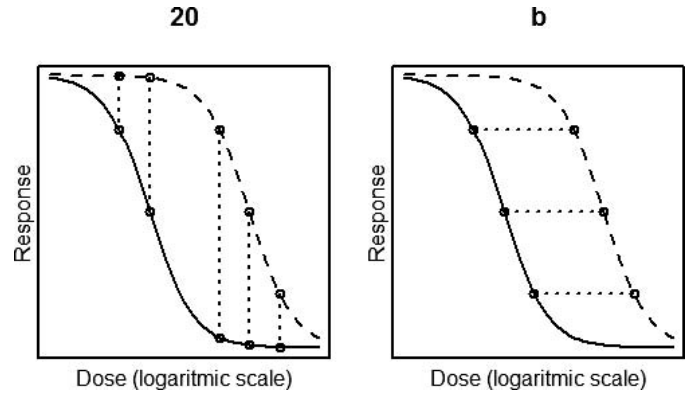


Figure 12. The vertical and horizontal reading of the dose–response curves for two herbicides sprayed on the same species. (a) The vertical reading used ANOVA. (b) The horizontal reading is based on regression analyses. A vertical reading of the means can only be compared at the same doses. That restriction does not apply to regressions because the curves can be compared, even though the dose ranges differ among the herbicides (from Streibig 1988).

each other, as are the upper and lower limits (Table 2).

In some rare instances the dose–response curves have common upper and lower limits and the same relative slopes, and then the curves are similar in all parameters, except for their relative displacement along the dose axis, expressed as the relative potency, which is independent of response level as seen in Figure 12 (Ritz et al. 2006; Streibig 1988). With relative potencies, we compare the biological exchange rate between two herbicides, an important biological characteristic that can be used to develop new herbicides or new formulations of herbicides, as well as the importance of the relative potency in assessing the joint action of herbicide mixtures (Streibig and Jensen 2000). In whole-plant assays, the assumption of similar curves is rarely met because of several unknown factors. In test tube site-of-action studies and other well-defined simple-enzyme systems, the likelihood of similar curves with herbicides with similar sites of action is probable. Biologically, in site-of-action studies, similar curves are a necessary but insufficient condition for unambiguously determining similar action at the target site (Streibig 1984).

Figure 12 illustrates the principle of vertical and horizontal assessment of the effect of herbicides. The vertical assessment is used to compare effects at preset doses with an ANOVA (discussed further later). Comparing the two methods of analysis, the vertical assessment is much more dependent on the dose; at small doses, there are almost no differences between the effects of the two herbicides at either high nor

Table 3. Summary of ANOVA table of a factorial experiment of the effect on weed biomass in oat and barley crops, with and without fertilizer. The block effect could be deleted because it was not significant; however, it could also be analyzed with a mixed model as described in the mixed-model section of text.

Factors	df	Probability (> <i>F</i> )
Nitrogen, kg	1	0.001**
Crop, type	1	< 0.001***
Nitrogen × crop	1	0.01*
Block	2	0.09
Residuals	6	

\*  $P \leq 0.05$ .

\*\*  $P \leq 0.001$ .

\*\*\*  $P \leq 0.001$ .

small doses. With horizontal assessments, there is no interest in the particular preset dose, which could be different for each curve, but rather, on preset response levels:  $ED_{10}$ ,  $ED_{50}$ , or  $ED_{90}$ . As already mentioned, when we use regressions, we can compare predicted response levels among bioassays with different dose ranges.

### Analysis of Variance

When treatments cannot be described by continuous predictors, regression analysis cannot be used; ANOVAs must be used to test differences among means. The following example is from a small field experiment with two crops, barley (*Hordeum vulgare* L.) and oat (*Avena sativa* L.), and two fertilizer levels, 0 and 100 kg N ha<sup>-1</sup>, and the objective was to determine how the crops and N affected weed biomass. The experimental design was a randomized complete block, and the ANOVA results are shown in Table 3. Obviously, the main effects of nitrogen and type of crop were significant as was the interaction between the type of crop and nitrogen; the block effect was not significant. The significant interaction makes the main effect irrelevant because, to understand the interaction, we have to look at all the combinations of treatments, namely, the simple effects (Figure 13).

Technically, when we have an interaction, we can combine the interaction effects by merging together the two factors, type of crop and nitrogen level, into one-factor combinations and perform a one-way ANOVA. In this instance, the result is that all factor combinations are significantly different from each other, except the weed biomass for barley, with or without nitrogen. The calculations for the estimated differences and adjusted P values are shown in Table 4. The adjusted P values are an important

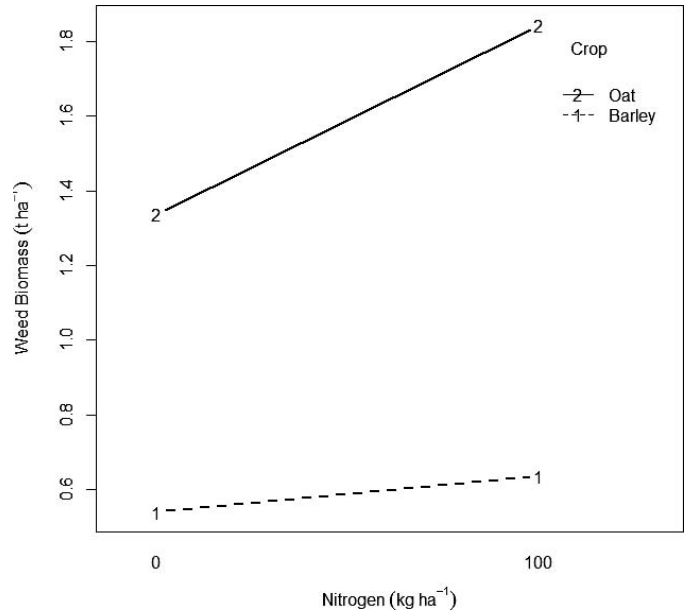


Figure 13. An interaction plot shows the differences between weed biomass in barley and oat are dependent on the nitrogen level. With no nitrogen, the differences between the two crops are significantly smaller than at 100 kg nitrogen, and that difference is significant according to the data in Table 4. The only nonsignificant difference is between barley at 0 and 100 kg nitrogen.

detail because considering more than a single P value simultaneously, as is commonly practiced, inflates the risk of false-positive findings; that is, the family-wise error rate becomes larger than 5%. There are a number of procedures for adjusting P values. The commonly used procedure is the Bonferroni adjustment where all P values considered are inflated through multiplication by the number of P values considered, which is a conservative procedure, and the P values may be inflated too much because the correlations between the tests are not incorporated. We recommend using a less-conservative approach: the flexible, single-step procedure, proposed by Hothorn et al. (2008), because it incorporates the correlation between tests when calculating the adjusted P values; it is applicable for arbitrary ANOVA and regression (mixed-effects) models.

Neglecting the interaction and only reporting the main effect of nitrogen and type of crop will forfeit important information about the difference in the behavior of the weed infestation in the two crops at various nitrogen levels. Interactions make interpretation of results difficult and require detailed examinations of simple effects that are a combination of effects. In the next example, there are no interactions between light and stress level on



Table 4. Estimated differences in weed biomass, and *t* test of differences between all simple effects and adjusted P values. Note that weed biomass was not affected in barley at any nitrogen (N) level.

Simple effects	Difference	Probability ( $>  t $ )
Oat (no nitrogen) – barley (no nitrogen) = 0	0.80	< 0.001***
Barley (100 kg N ha <sup>-1</sup> ) – barley (no nitrogen) = 0	0.10	0.740
Oat (100 kg N ha <sup>-1</sup> ) – barley (no nitrogen) = 0	1.30	< 0.001***
Barley (100 kg N ha <sup>-1</sup> ) – oat (no nitrogen) = 0	-0.70	< 0.001***
Oat (100 kg N ha <sup>-1</sup> ) – oat (no nitrogen) = 0	0.51	0.003**
Oat (100 kg N ha <sup>-1</sup> ) – barley (100 kg N ha <sup>-1</sup> ) = 0	1.21	< 0.001***

\*\* P ≤ 0.05.

\*\*\* P ≤ 0.001.

soybean [*Glycine max* (L.) Merr.]; therefore, the lines in Figure 14 are parallel, and it is safe to report only the main effects (Samuels et al. 2004).

In weed science literature, ANOVAs are probably the most used statistical analyses because they are integral parts of the analyzing field experiments in agriculture. In this section, we have only worked with two main effects. If more than two main effects are analyzed, the interpretation of multiway interactions becomes much more involved.

In contrast to regression, where we describe relationships, ANOVAs only test the differences of means. However, regression and ANOVA are tied together, and comparison between the results of an ANOVA and the results of a regression is used to test for lack of fit, as mentioned earlier.

**Analysis of Covariance.** In an ANOVA, the interest lies in the differences of the means, and in

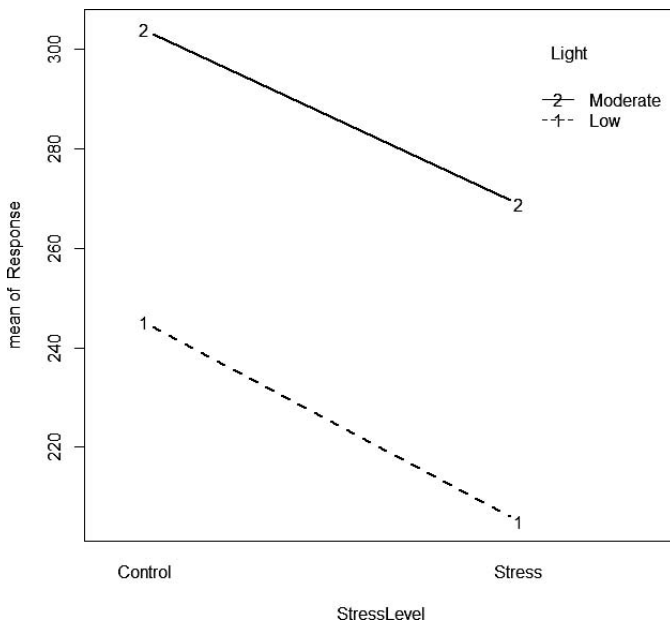


Figure 14. An experiment studying the combination of light and stress on the growth of soybeans. The two lines are parallel, which means there is no interaction (Samuels et al. 2004).

regression analysis, the interest lies in the rate of change, slope, or perhaps some other parameters of biological interest, e.g., asymptotes and  $ED_x$  (e.g.,  $ED_{50}$ ) levels in the nonlinear case. The ANCOVA combines the two approaches and can give strong and elegant tests for differences of means, intercepts, and slopes. A mechanical weed control experiment (Rasmussen et al. 2008) in Figure 15 illustrates the result of the intensity of harrowing, either along or across the direction of sowing (factors), and the cover of green leaves left after the harrowing (continuous predictor). To make the relationship between leaf and intensity of harrowing linear, we must use a logarithm transformation of the cover of

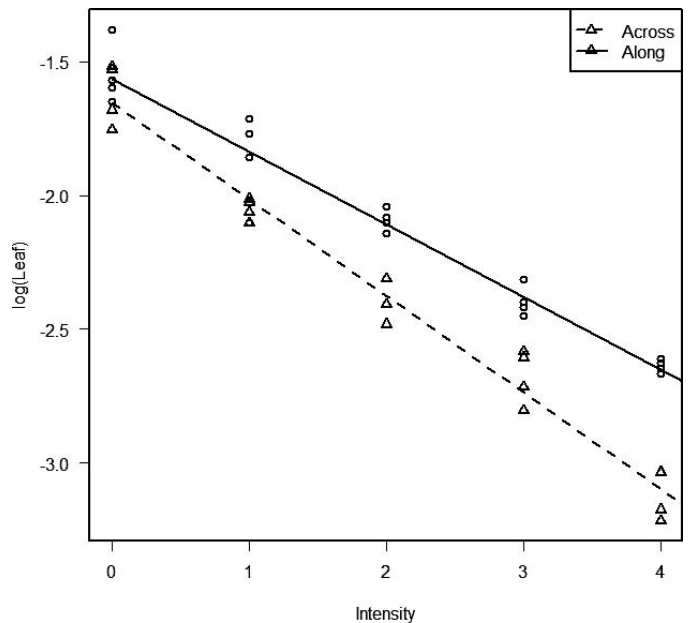


Figure 15. Graphic representation of the analysis of covariance of harrowing along or across the direction of sowing and the intensity of harrowing (number of tines with which the plot was harrowed). The slope for harrowing across the plot is  $-0.36 (\pm 0.01)$ , whereas the slope for harrowing along the plot is  $-0.27 (\pm 0.01)$ . The slopes are significantly different from each other, whereas  $\log(\text{leaf})$  at zero intensity is  $-1.56 (\pm 0.04)$  and  $-1.83 (\pm 0.05)$  for the along plot and the across plot, respectively (Rasmussen et al. 2008).

green leaves. This design calls for an ANCOVA to test whether the two regression slopes are different.

First, we perform an ordinary ANOVA, where *intensity* is defined as a factor. This is the most general model and the one we use to test whether we can assume linearity of logarithm of the green cover on harrowing intensity in the ANCOVA model (Figure 15). The slopes are different, because the test-for-lack-of-fit and the test yields an insignificant result ( $P = 0.7$ ). Therefore, we can tentatively assume that the linear regression describes the systematic variation in data satisfactorily. In the second step, we assume the slopes of the two harrowing directions, in response to intensity, are similar. This test is highly significant ( $P < 0.001$ ). Consequently, the best model to describe data is the one in which the regression slopes are different for the two lines (Figure 15). The two regressions lines are not similar, and in Figure 15, the individual slopes must be reported. If there were no interaction, we would only need to report one regression slope. The results in Figure 15 illustrate a logarithmic-transformed leaf cover on harrowing intensity, similar to the one shown in Figure 6.

Currently, ANCOVA is generally used for linear models, but the concept extends also to nonlinear models. The data in Figure 11 and Table 2 can also be run like an ANCOVA, for example, by testing whether the upper and lower limits could be the same for both curves. A sequential test shows that the upper and lower limit of the regressions curves in Figure 11 are probably the same for both curves. The next and final step could be to test whether the slopes of the two curves were the same. We get a significant result ( $P < 0.0001$ ) by comparing the model with the same upper and lower limits against the one with the same slope for the two curves. Consequently, we cannot assume the curves to be similar.

### Analysis of Independent Experiments

To make sure that the results of an experiment do not just depend on mere chance, we must always repeat our experiments, particularly if we wish to publish. Most journals require that experiments be independently replicated at least twice. That does not necessarily mean that the experiment should use exactly the same predictor values, e.g., density of weeds or crops or herbicide rates. By using regression, it is possible to change the predictor values, perhaps from results from previous experiments, and improve the range of the predictors to acquire better estimates. Therefore, regression is the

most versatile method to summarize experiments with quantitative predictors. The best way to summarize replicated experiments is to use mixed-model versions of linear regression, ANOVA, and ANCOVA, or, with some loss in power, to show results from each experiment individually.

With an ANOVA, there is no such flexibility because we implicitly state that it is the mean of a specific treatment of interest or, more specifically, the difference between treatments (Table 4). This is understandable when working with different species, cultivation methods, etc., but it puts heavy restrictions on the ANOVA. Today, the handling of series of ANOVA experiments with the same design and treatments is done by mixed models.

**Linear Mixed Models.** When performing an ANOVA in a randomized complete-block design, everything, except the intentional (fixed) treatment(s), reflects random variation. We consider the selection of the blocks to take place by random, in the sense that we cannot tell in advance whether or not the next block will exhibit a low, high, or more-moderate response level. We can describe the *random effect* as effects that cannot be controlled and, therefore, cannot be explained by the treatment structure of the experiment. Consequently, we are not inherently interested in these block effects, but in an adequate analysis, we need to acknowledge the variation that they reflect, variation introduced as part of how the experiment was designed.

The following example is a field experiment that was conducted at two sites over two yr for a total of four experimental environments. The study included two sugarbeet (*Beta vulgaris* L.) genetic backgrounds (cultivar pair) and 12 different weed-control treatments (Kniss et al. 2004). The study was a randomized complete-block design, and therefore, we have two sets of random effects (sites, and blocks within environment) as well as two fixed effects (cultivar and herbicide). Including block and environment as random effects allows us to account for the variability those factors contribute to the results, but to focus our attention primarily on the fixed effects, in which we are often most interested.

For the response variable, gross sucrose production per hectare, there was no interaction between the sugarbeet cultivar and the herbicide (Kniss et al. 2004). We can draw that conclusion by fitting the mixed-effect model with and without the cultivar by herbicide interaction term, and then conduct a test-for-lack-of. A nonsignificant test indicates that we can remove the interaction term and focus on

the main effects of cultivar and herbicide. The contribution of random effects (as well as interactions between random effects and fixed effects) can also be quantified in a similar fashion, by fitting a model with and without the effects of interest and conducting a lack-of-fit test. This may be valuable if the researcher is interested in the effect of the random effect on the response variable but has chosen the levels at random (such as crop cultivars). The structure of the original experiment lends itself to several nonorthogonal contrasts to compare glyphosate-resistant and conventional versions of each cultivar, but, for simplicity, we focus here simply on the main effect of herbicide treatment (Table 5). Results of the contrasts can be found in the original article (Kniss et al. 2004).

The original analysis (done in SAS software [SAS Institute]) can be nearly replicated in R using the lme4 add-on package, and mean separation can be conducted using the LSMeans add-on package. If all pairwise comparisons are made and the P values are not adjusted for multiple comparisons, many differences between herbicide treatments are significant at  $\alpha = 0.05$  (Table 5). If Tukey's method is used to adjust the P values in the LSMeans function, significant treatment differences are mostly limited to comparisons with the nontreated control. The glht function in the "multcomp" package (Hothorn et al. 2008) also controls the type I error rate when conducting multiple comparisons and provides similar results to that of the LSMeans package.

When it comes to the mixed-model version of linear regression analyses, the same principles as outlined above apply. The complexity becomes much greater when running nonlinear regressions with upper and/or lower limits. Our experience is that, for the time being, the best method for comparing two nonlinear regression models is to present each independent experiment (Mennan et al. 2012). There are ways of running mixed nonlinear models, but they are not always easy to handle (Ritz and Streibig 2008).

### Model Diagnostics

To make sure that the prerequisites for the statistical analysis are fulfilled, we have to address the assumptions:

1. That the correct regression model (mean function) is used,
2. That the variance homogeneity is constant at any response level,

3. That measurement errors are normally distributed,
4. That measurement errors are mutually independent.

Substantial departure from these assumptions could result in inaccurate estimates of parameters or distorted standard errors or both. However, the first assumption does not apply to ANOVA because with ANOVAs, we only model the means of treatments; otherwise, the other three assumptions do apply.

Model validation can be detected graphically, as already shown in Figures 3, and by formal statistics, such as the test for lack-of fit, mentioned earlier (Table 1). Graphical presentation of the fit, using a residual plot and a quantile–quantile (Q–Q) plot will often suffice. Particularly in a regulatory framework, such as for toxicity testing, formal statistics tests are also considered (Van der Vliet and Ritz 2013).

The correct regression model is often difficult to determine. The wider the range of the independent variables, the more likely it is that linearity will disappear. Our experience is that the higher number of an independent variable, the more difficult it may be to make sure that the model is a genuine linear or a log-logistic curve. Often, the choice of the numbers of  $x$ 's is based on the tradition in the laboratories, and sometimes, several sigmoid curves will fit equally well, for example, the symmetric log-logistic curves or the asymmetric Weibull functions (Ritz 2010). Generally, models with easily understood, biologically meaningful parameters should be preferred.

The variance homogeneity is rarely met, when the difference between the largest response and smallest response differ by a factor 10. The data in Figure 11 are a good example: Minimum response is 0.5, and the maximum response is 4.7. The residual plot in Figure 16 clearly shows a funnel-type distribution of residuals; at low predicted dry matter, the scatter is small, and at high predicted dry matter, the scatter is large.

### Transformation to Ensure Variance Homogeneity.

The question is how much does the nonhomogeneous variance matter? Numerous reviews of research articles in crop-protection literature show that several authors routinely transform data, for example, percentage data, using the arcsine square-root transformation, and for count data authors use the log transformation. When that is done in the context of an ANOVA, it does not affect the underlying basic model because only the means or

Table 5. Summary of gross sucrose production in sugarbeet cultivars conducted over 2 yr and four locations with 12 different weed-control treatments in a randomized complete-block design. There was no interaction between sugarbeet cultivars and type of herbicide. For interpretation of the results, see the text; a full report can be found in Kniss et al. (2004).

Comparison <sup>a,b</sup>	Estimate	P value
LSMeans unadjusted for multiple comparisons		
01_Handweeded.RR – 02_Nontreated	5,499	< 0.0001
01_Handweeded.RR – 05_Glyphosate3	-1,459	0.01764
02_Nontreated – 03_Glyphosate1	-5,274	< 0.0001
02_Nontreated – 04_Glyphosate2	-6,508	< 0.0001
02_Nontreated – 05_Glyphosate3	-6,959	< 0.0001
02_Nontreated – 06_Handweeded	-6,235	< 0.0001
02_Nontreated – 07_PreConv2	-5,465	< 0.0001
02_Nontreated – 08_Conv2	-4,802	< 0.0001
02_Nontreated – 09_PreMicro	-5,862	< 0.0001
02_Nontreated – 10_Micro	-4,695	< 0.0001
02_Nontreated – 11_PreConv3	-5,104	< 0.0001
02_Nontreated – 12_Conv3	-5,396	< 0.0001
03_Glyphosate1 – 04_Glyphosate2	-1,233	0.04451
03_Glyphosate1 – 05_Glyphosate3	-1,684	0.00621
04_Glyphosate2 – 07_PreConv2	1,042	0.08923
04_Glyphosate2 – 08_Conv2	1,705	0.00561
04_Glyphosate2 – 10_Micro	1,812	0.00327
04_Glyphosate2 – 11_PreConv3	1,404	0.02234
04_Glyphosate2 – 12_Conv3	1,111	0.07011
05_Glyphosate3 – 07_PreConv2	1,493	0.01515
05_Glyphosate3 – 08_Conv2	2,156	0.00048
05_Glyphosate3 – 09_PreMicro	1,096	0.07408
05_Glyphosate3 – 10_Micro	2,263	0.00025
05_Glyphosate3 – 11_PreConv3	1,854	0.00262
05_Glyphosate3 – 12_Conv3	1,562	0.0111
06_Handweeded – 08_Conv2	1,432	0.01977
06_Handweeded – 10_Micro	1,539	0.0123
LSMeans adjusted for multiple comparisons using the Tukey method		
01_Handweeded.RR – 02_Nontreated	5,499	< 0.0001
02_Nontreated – 03_Glyphosate1	-5,274	< 0.0001
02_Nontreated – 04_Glyphosate2	-6,508	< 0.0001
02_Nontreated – 05_Glyphosate3	-6,959	< 0.0001
02_Nontreated – 06_Handweeded	-6,235	< 0.0001
02_Nontreated – 07_PreConv2	-5,465	< 0.0001
02_Nontreated – 08_Conv2	-4,802	< 0.0001
02_Nontreated – 09_PreMicro	-5,862	< 0.0001
02_Nontreated – 10_Micro	-4,695	< 0.0001
02_Nontreated – 11_PreConv3	-5,104	< 0.0001
02_Nontreated – 12_Conv3	-5,396	< 0.0001
05_Glyphosate3 – 08_Conv2	2,156	0.024
05_Glyphosate3 – 10_Micro	2,263	0.01324
GLHT adjusted for multiple comparisons using the Tukey method		
02_Nontreated – 01_Handweeded.RR = 0	-5,903	< 0.01
03_Glyphosate1 – 02_Nontreated = 0	5,241	< 0.01
04_Glyphosate2 – 02_Nontreated = 0	7,084	< 0.01
05_Glyphosate3 – 02_Nontreated = 0	7,652	< 0.01
06_Handweeded – 02_Nontreated = 0	6,385	< 0.01
07_PreConv2 – 02_Nontreated = 0	5,517	< 0.01
08_Conv2 – 02_Nontreated = 0	4,835	< 0.01
09_PreMicro – 02_Nontreated = 0	6,083	< 0.01
10_Micro – 02_Nontreated = 0	5,196	< 0.01
11_PreConv3 – 02_Nontreated = 0	5,236	< 0.01
12_Conv3 – 02_Nontreated = 0	5,226	< 0.01



Table 5. Continued.

<sup>a</sup> Abbreviations: Conv, conventional; GLHT, generalized linear hypothesis test; LSMeans, least-square means RR, glyphosate resistant.

<sup>b</sup> Treatments: PRE, ethofumesate; conventional (conv2 at cotyledon and two to four true leaf; conv3, same as conv2 plus at four to six true leaf), phenmedipham (0.19 kg ha<sup>-1</sup>) + desmedipham (0.19 kg ha<sup>-1</sup>) + triflusaluron (0.02 kg ha<sup>-1</sup>) + clopyralid (0.10 kg ha<sup>-1</sup>); microrate, phenmedipham (0.048 kg ha<sup>-1</sup>) + desmedipham (0.048 kg ha<sup>-1</sup>) + triflusaluron (0.005 kg ha<sup>-1</sup>) + clopyralid (0.025 kg ha<sup>-1</sup>); glyphosate, 0.84 ae ha<sup>-1</sup>, numbers indicate number of times.

the transformed means matter. Regression analysis is a different matter. One cannot assume a linear relationship, and then, based on variance heterogeneity, make a log transformation of the dependent variable. Transformation of the response can dramatically change the whole regression relationship. Figure 6 is a good example, an exponential relationship changes to a linear one, just as it did in Figure 15.

A general method for transforming regression data is to use the most-general model, an ANOVA and use a Box–Cox analysis to find out which power exponent to use on both sides of the regression (Streibig et al. 1993). The optimal Box–Cox analysis strives to find an exponent that achieves homogenous variance, but it requires that the responses be greater than zero, or alternatively, the response can be augmented initially by some positive constant.

With the data from Figure 11 a transform-both-sides Box–Cox analysis provides an estimated exponent close to the logarithm (0.101). By using that information and doing the transform-both-sides regression, we get the following parameter estimates for the *ED*<sub>50</sub> of glyphosate, 62 (± 5.4),

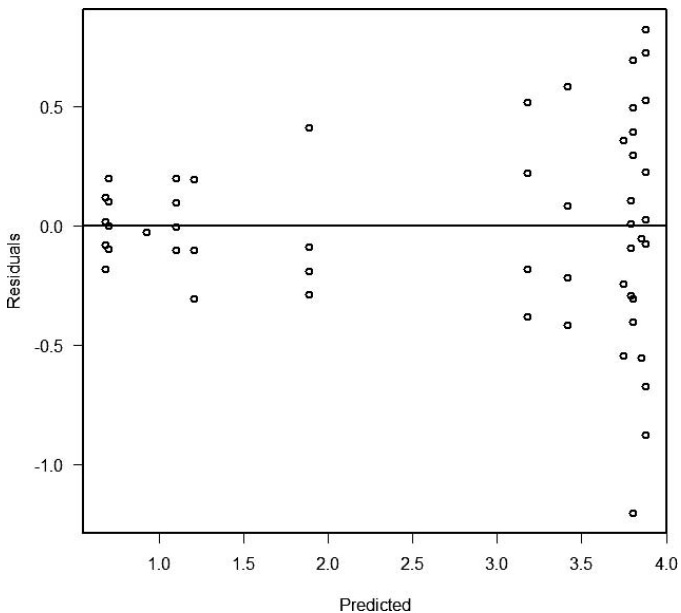


Figure 16. The residual plot for the white mustard regression fit in Figure 11 and Table 1.

and bentazon, 29 (± 2.1); there are almost no differences between the two fits, and the same result applies to the other parameters. Usually, one gets almost the same parameters, whether or not a transformation of both sides of a Box–Cox approach was used. The real difference lies in the estimated standard errors, which may change. That means that, when comparing parameters, the bias in the standard error will affect subsequent tests and conclusions when the variation is heterogeneous.

In some instances, the transformation of both sides is crucial to determine the proper results. The response curve in Figure 17 is steep, but there are really no data to support the fitted continuous line between the upper and lower limits. Furthermore, the variation in the data is very heterogeneous (Figure 17). The residual plot in Figure 18 shows the funnel-like distribution of residuals, as also shown in Figure 16. The Box–Cox analysis suggested a transformation of both sides close to the logarithm. The fit of the transform- both- sides was almost the same as the upper and lower limits

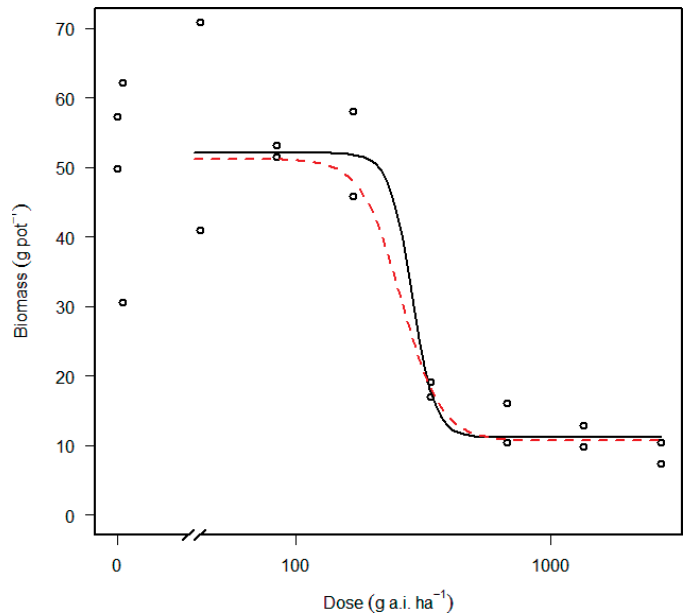


Figure 17. The dose–response for the biomass of barley using glyphosate. There is a huge variation in the biomass close to the control, and very small variations at high doses of glyphosate do matter. The continuous line is an ordinary fit, and the red broken line is the fit after Box–Cox transformation (see Table 6).



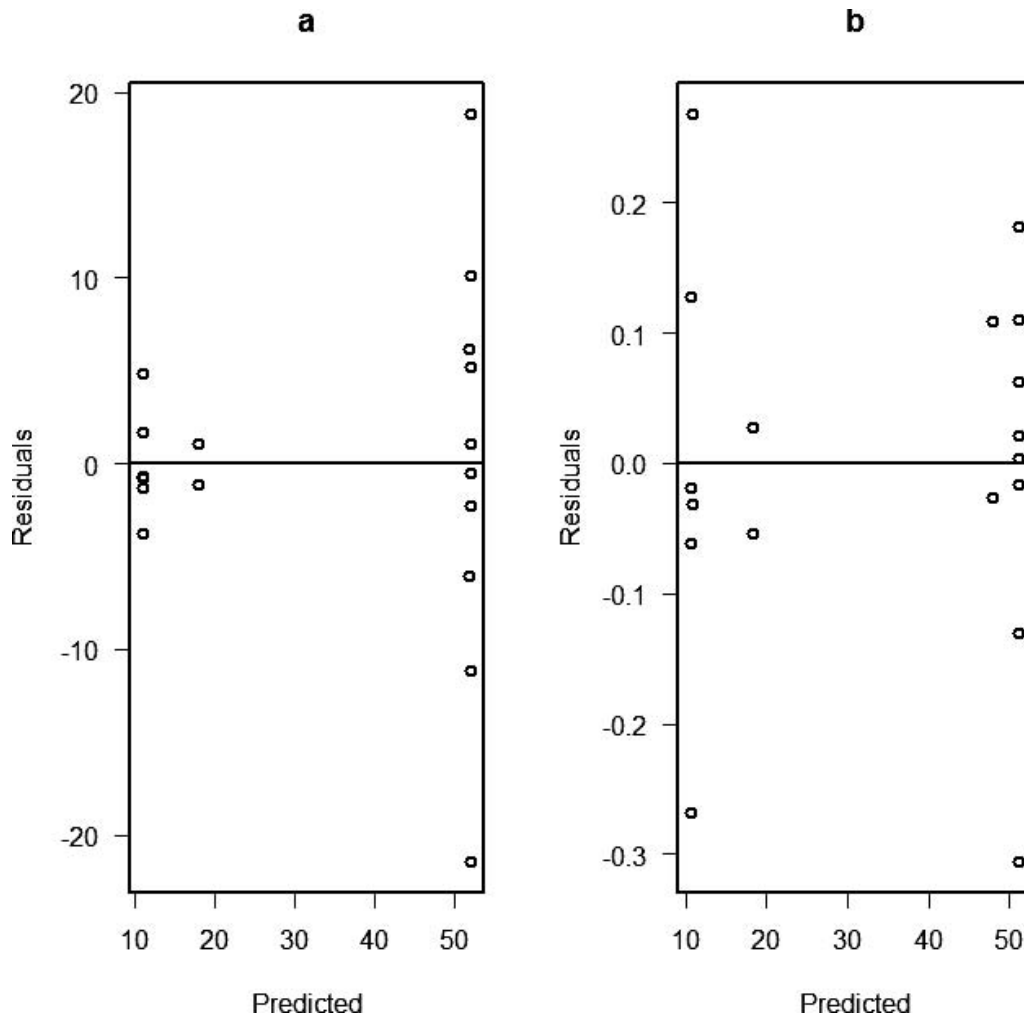


Figure 18. Residual plots with no transformation (a) and with transform-both sides (b).

(Figure 17 and Table 6). The  $ED_{50}$  and the relative slope and their standard errors changed dramatically. Without the transform- both- sides, the  $ED_{50}$  was not different from zero, but with a regression using transform- both- sides the standard error of the  $ED_{50}$  is reduced dramatically (Table 6). Interestingly, the relative slopes  $b$  are never different from zero because there are almost no observations to support it. This illustrates that a transformation cannot remedy a poor distribution of responses or doses. A regression with or without transform-

both- sides often finds almost the same parameters, provided the responses cover the whole response range. However, the “correct” standard errors are usually based on transform- both- sides. The distribution of residuals in Figure 18 shows how transform- both- sides changes the funnel-like distribution to a more evenly distributed pattern without any systematic trends.

### Binomial Data

A binomial response is defined as a sum of a series of independent experiments in which there are 2 outcomes: dead or alive, or germinating or not. Basically, a binomial response means that each seed or plant in an experiment is an experimental unit of its own. Consequently, to analyze these kinds of data, we need to know the total number of seeds or plants and the number of germinating seeds or dead plants.

**Logistic Regression.** There is an ensuing discussion about the use of dose-response curves on biomass

Table 6. Comparison of parameters of the fitting of the data in Figure 17.<sup>a</sup>

	Without Box-Cox		With Box-Cox	
	Estimate	SE	Estimate	SE
Slope	9.71	42.92	5.60	5.69
Lower limit	11.13	3.78	10.67	1.07
Upper limit	52.05	3.25	51.21	4.93
$ED_{50}$	286	210	260	60

<sup>a</sup> Abbreviations: SE, standard error;  $ED_{50}$ , 50% effective dose.

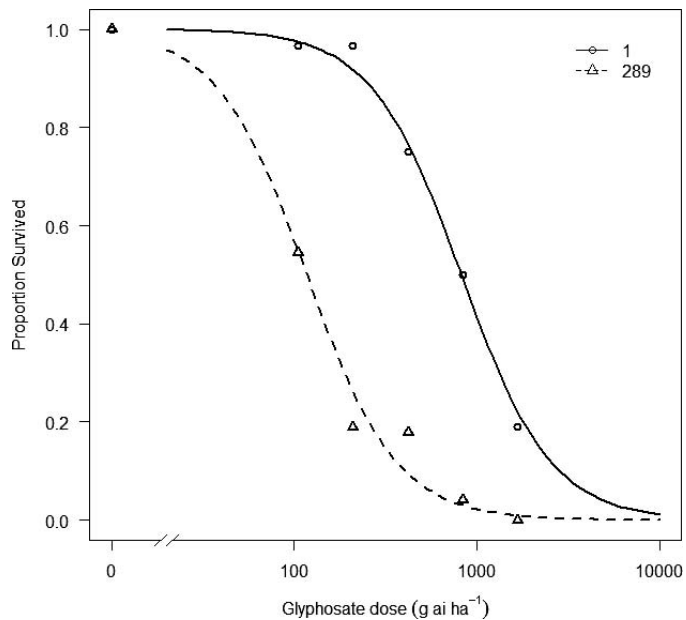


Figure 19. The survival of plants at increasing doses of glyphosate is shown. The curves are similar, except for their relative displacement on the dose axis. The relative potency was  $7.0 (\pm 1.7)$ .

or on survival, a discussion that is highly relevant for curtailing the spread of resistance genes. In herbicide-resistance studies, plants highly injured by herbicide may produce seeds. From a practical point of view, a dead weed will not multiply. Fortunately, the analysis of the dose–responses for binomial data is almost the same as for continuous data, except that the total number of plants in the pot and the number of dead plants in the pot are both required for defining the response.

Often, binomial data are converted to percentage of survival or mortality and analyzed as though the data were continuous. Simulation studies have shown that whatever the numbers of test plants used, approximations will yield standard errors that are too narrow. That has implication when comparing say several  $LD_{50}$  values.

By assuming that the herbicide kills all plants at high doses and no plants die at the untreated control, we have a response span from 0 to 1. The binomial distribution and a two-parameter log-logistic curve can be used:

$$y = 1 / \{1 + \exp[\log(x) - \log(LD_{50})]\} \quad [9]$$

The parameter of interest is the  $LD_{50}$ . The literature on this kind of experiments is comprehensive because these types of experiments are instrumental to classifying chemical substances into hazard groups. Sometimes, there will also be some lethality in the control group, and Equation 9 will have an upper limit  $d$  that is smaller than one, and

we must use Equation 8 to estimate the survival/mortality in the nontreated control. The logistic regression based on Equation 9 will usually give results that are similar to what a probit analysis would produce, but the former seems to be the preferred model nowadays because of the computational advantages and ease of interpretation (Finney 1971).

In the experiment in Figure 19, two biotypes of a weed species were exposed to glyphosate, and their survival (the opposite of their lethality) was fitted with Equation 9. The two regression curves look like slope  $b$  is similar. Based on a test, that assumption was not rejected. Therefore, the relative potency between the curves, whatever the response level, is  $7.0 \pm 1.7$ , which means that sevenfold more glyphosate must be used to obtain the same survival level for biotype 298 as for biotype 1.

**Germination.** Measurement of germination is an integral part of weed, seed, and crop science research. Weed scientists are particularly interested in the speed of germination: Does the weed or the crop germinate with the same speed or not? Experiments to investigate germination over time can be classified into two approaches.

In the first approach, the germinated seeds in a petri dish at a specific time are counted, and afterward, the petri dish is discarded. That means that, at each inspection time, a fresh petri dish was inspected. This is probably the least-common method because it requires a lot of space and is, therefore, rather costly. That experimental design can be analyzed using well-established models for binary/binomial data, such as logistic regression or other generalized linear models, mentioned in the section above.

The second approach is more common and less costly. The petri dishes are inspected over time; the same seeds are observed repeatedly for some prespecified duration of the experiment. In other words, what is observed is the waiting time until the event of interest occurs, and the resulting data are often referred to as event–time or time-to-event data. Although we have not done a thorough analysis of the literature on germination studies in biological sciences, it would be close to the truth to say that most germination experiments follow inspection of the same petri dishes, but data are analyzed with logistic regression. In doing so, the standard errors are too narrow, for example, to 50% germination time. An analysis of the two methods on germination data of the same petri dishes is given elsewhere (Ritz et al. 2013).

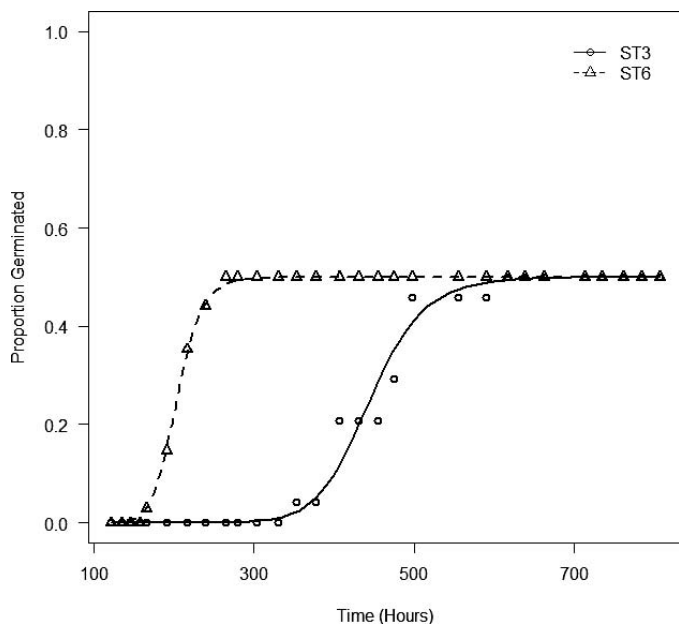


Figure 20. Germination curves for nettlespurge at different temperatures is fitted with an event-time model. The slope of the germination curve at time  $T_{50}$  is proportional to  $b$  (with a proportionality constant that depends on  $d$ ). In this instance, the two curves are identical, meaning that the only difference between the slopes is the  $T_{50}$  ( $203 \pm 6.7$  for ST3 and  $445 \pm 7$  for ST6). The interpretation of the parameters is the same as for the other dose-response curves (upper limit =  $0.50 \pm 0.07$ ; slope =  $-13 \pm 2$ ) (C Andreasen, unpublished data).

The analysis presented takes into account the inspection intervals. The shorter the intervals, the more precise will be our determination of germination time of a seed. The key result is that use of appropriate statistical models is important to get a better appreciation of the uncertainty that is present in germination experiments.

Figure 20 illustrates the germination, in response to time, of nettlespurge (*Jatropha* L. spp.), a nondomesticated oil plant; the two curves represent different germination temperatures. The analysis is done with Equation 8 because the upper germination proportion is different from 1.

### Concluding Remarks

Applied statistics is the collection, analysis, and interpretation of statistical parameters in such a way that others can see the quality of the data and the link between the experimental design, the objective (research questions) of the work, and the statistical methods used. We have only unveiled a tiny part of the possibilities and methods available for weed scientists. We do hope we have whetted your interest in the use of proper statistics on weed science data.

### Literature Cited

- Blackman GE, Templeman WG, Halliday DJ (1951) Herbicides and selective phytotoxicity. *Annu Rev Plant Physiol* 2:199–230
- Cedergreen N, Kudsk P, Mathiassen SK, Streibig JC (2007) Combination effects of herbicides on plants and algae: do species and test systems matter? *Pest Manag Sci* 63:282–295
- Cedergreen N, Madsen TV (2002) Nitrogen uptake by the floating macrophyte *Lemna minor*. *New Phytol* 155:285–292
- Christensen MG, Teicher HB, Streibig JC (2003) Linking fluorescence induction curve and biomass in herbicide screening. *Pest Manag Sci* 59:1303–1310
- Cousens R (1985) An empirical model relating crop yield to weed and crop density and a statistical comparison with other models. *J Agric Sci* 105:513–521
- Cousens R (1988) Misinterpretations of results in weed research through inappropriate use of statistics. *Weed Res* 28:281–289
- Finney DJ (1971) *Probit Analysis*. 3 edn. London: Griffin. 333 p
- Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. *Biometrical J* 50:346–363
- Kniss AR, Wilson RG, Martin AR, Burgener PA, Feuz DM (2004) Economic evaluation of glyphosate-resistant and conventional sugar beet *Weed Tech* 18:388–396
- Kniss AR, Vassios JD, Nissen SJ, Ritz C (2011) Nonlinear regression analysis of herbicide absorption studies. *Weed Sci* 59:601–610
- Mennan H, Streibig JC, Ngouajio M, Kaya E (2012) Tolerance of two *Bifora radians* Bieb populations to ALS inhibitors in winter wheat. *Pest Manag Sci* 68:116–122
- Nelder JA (1966) Inverse polynomials, a useful group of multi-factor response functions. *Biometrics* 22:128–141
- Onofri A, Carbonell EA, Piepho HP, Mortimer AM, Cousens RD (2010) Current statistical issues in Weed Research. *Weed Res* 50:5–24
- Rasmussen J, Bibby BM, Schou AP (2008) Investigating the selectivity of weed harrowing with new methods. *Weed Res* 48:523–532
- Ritz C (2010) Toward a unified approach to dose-response modeling in ecotoxicology. *Environ Toxicol Chem* 29:220–229
- Ritz C, Cedergreen N, Jensen JE, Streibig JC (2006) Relative potency in nonsimilar dose-response curves. *Weed Sci* 54:407–412
- Ritz C, Pipper CB, Streibig JC (2013) Analysis of germination data from agricultural experiments. *Eur J Agron* 45:1–6
- Ritz C, Streibig JC (2005) Bioassay analysis using R. *J Stat Softw* 12:1–22
- Ritz C, Streibig JC (2008) *Nonlinear Regression with R*. New York: Springer. 1:1–144
- Samuels M, Witmer J, Schaffner A (2004) *Statistics for the Life Sciences*. 4th edn. London: Pearson. 672 p
- Seefeldt SS, Jensen JE, Fuerst EP (1995) Log-logistic analysis of dose-response relationships. *Weed Technol* 9:218–227
- Streibig JC (1983) Fitting equations to herbicide bioassays: using the methods of parallel line assay for measuring joint action of herbicide mixtures. *Ber Fachg Herbol* 24:183–193
- Streibig JC (1984) Measurement of phytotoxicity of commercial and unformulated soil-applied herbicides. *Weed Res* 24:327–331
- Streibig JC (1988) Herbicide bioassay. *Weed Res* 28:479–484
- Streibig JC, Combellack JH, Pritchard GH, Richardson RG (1989) Estimation of thresholds for weed-control in Australian cereals. *Weed Res* 29:117–126

- Streibig JC, Jensen JE (2000) Action of herbicides in mixtures. Pages 153–180 *in* Cobb AH, Kirkwood RC, eds. *Herbicides and Their Mechanisms of Action*. Sheffield: Sheffield Academic
- Streibig JC, Rudemo M, Jensen JE (1993) Dose–response curves and statistical models. Pages 29–55 *in* Streibig JC, Kudsk P, eds. *Herbicide Bioassays*. Boca Raton: CRC
- Stroup WW (2014) Rethinking the analysis of non-normal data in plant and soil science. *Agron J.* 106:1–17
- Van der Vliet L, Ritz C (2013) Statistics for analyzing ecotoxicity test data. Pages 1081–1096 *in* Blaise C, Férard JF, eds. *Encyclopedia of Aquatic Ecotoxicology*. Berlin: Springer. 1221 p
- Woodford EK (1950) Experimental techniques for the evaluation of selective herbicides. *NAAS Q Rev* 9:1–10

*Received November 1, 2013, and approved July 9, 2014.*