TRIAL SETUP AND STATISTICAL ANALYSIS

Andreas Büchse, Pawel Krajewski, Kristian Kristensen and *Wiesław Pilarczyk*

1. Introduction

1.1. Background

The response from field trials is subject to random variation. This means that two neighbouring plots grown with the same variety and treated in the same way will always yield differently. This also applies to all other recordings made on a continuous scale. The size of the differences will depend on several circumstances such as the variability in the soil, variability in the applied fertilizer, historical events and uncertainty in the recording process. This means that a recorded difference between e.g. two varieties may be due to either a true difference in the response of the two varieties or may be due to random variations. In order to help decide whether the difference is caused by the different varieties or by random variation it is necessary to apply some statistical methods in order to estimate the actual size of the random variation in the field and compare the measured difference with the size of the random variation. In order to do that properly it is necessary to use properly designed trials and the correct way of analysing the recorded data.

This chapter gives some information on how to design the trials in such a way that the part of random variation that determines the uncertainty in the difference between varieties (treatments) is as small as possible. In designing the experiment it is essential to take into account the size of the difference that the researcher wants to become significant in order to design the trial with the number of replicates that is considered to be appropriate for the level of random variations expected in the trial.

This chapter also gives some information on how to analyse the most common types of measured variables in variety trials under organic and low input systems and the conclusions that can and cannot be drawn from the analyses.

The random variability may in some cases be much larger for organic grown trials than for similar conventional grown trials. In two series of comparable trials with spring barley in Denmark and Sweden, the random variability was largest in the organic grown trials in 19 out of 34 pairs of trials in Sweden and in 3 out of 4 trials in Denmark. On average the random variability was approximately: 5.0 $(hkg/ha)^2$ in the conventional grown trials and 7.2 $(hkg/ha)^2$ in the organic grown trials, but the maximum random variability was 2-3 times higher for the organic grown trials than for the conventional grown trial. This indicates that it may be necessary to have more replicates in organic grown trials than in conventional grown trials if one wishes to maintain the same precision.

It may be expected that the competition between neighbouring plots may increase when diseases are uncontrolled. This may be handled by increasing the guard areas between plots. However, increasing the guard areas too much will usually increase the random variation. Alternatively one could compensate for the increased competition by modelling the competitions (see the text on plot size and shape in section 2.2) or by increasing the number of replicates.

The validity of the statistical analyses depends on some basic assumptions.
Therefore, some information is given on how to check that the assumptions are fulfilled and how to proceed if they are not.

The text in this chapter tries to describe the principles and methods to be used together with the most important assumptions that are needed for the methods to work correctly. Details on how to do the calculations are not provided. More details on that subject can be found in the references and in the documentation of the statistical software that can be used to do the calculations such as Genstat (Payne, 2006), SAS (SAS Institute, 2006), R (http://www.r-project.org/) and others. Examples on the applications may be found in the references and in the documentation of the used software.

1.2. Definitions

In the following we will define a plot as the units to which the varieties are allocated. A plot may contain several plants from the same variety. In some cases a plot can be subdivided and each part of such a plot will be called a sub-plot and in such cases the plot that is subdivided is usually called main-plot (or whole-plot). Sub-plots may be used either for applying different treatments to each of these (as in a split-plot design) or for taking more samples in each plot (e.g. samples of plants for determination of dry matter). A block is a collection of plots within which the plots are randomised. If many varieties (treatments) are to be included in the design, a block with all varieties (treatments) may be so large that it will be difficult (impossible) to find blocks that are sufficiently homogeneous. In such cases the varieties are collected in sub-blocks, which are randomised within each block and subsequently the plots within each sub-block are also randomised. This is the case in the recommended types of incomplete blocks (see 2.1).

In order to describe the level and the variability of a given variable, e.g. yield, some measures are usually calculated. The most frequently used measure is the mean, which is given by:

$$\bar{y} = \frac{1}{n}(y_i + y_2 + ... + y_n).$$

The median is given by the value that separates the ordered observations in two groups of equal size. The median is more robust than the mean, but has a larger uncertainty than the mean if the data are normally distributed. The most frequently used measures to describe the variability are the variance and the standard deviation given by:

$$\text{Variance: } s^2 = \left(\frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + (y_3 - \bar{y})^2 + ... + (y_n - \bar{y})^2}{n-1} \right)$$

$$\text{Standard deviation: } s = \sqrt{s^2}$$

In the recommended statistical methods it is assumed that the recorded plot values for a variable are independent, which means that the observation made in one plot does not give any information on the observation in another plot. One feature of independent observations is that the variance of the mean is inversely proportional to the number of observations used for forming the mean. So if s^2 is the estimated variance on single observations then the variance and standard error of the mean based on n independent observations is given by:

Variance on the mean: $s_{\bar{y}}^2 = \frac{s^2}{n}$ and the standard error of the mean: $s_{\bar{y}} = \frac{s}{\sqrt{n}}$.

This can be used to calculate the variance of a difference between two means, e.g. for variety A and variety B:

$sed^2 = s_{\bar{y}_a}^2 + s_{\bar{y}_b}^2$ and equivalently the standard deviation of the difference: $sed = \sqrt{s_{\bar{y}_a}^2 + s_{\bar{y}_b}^2}$.

If the variances can be assumed identical for the varieties (which they most often can) this can be written as $sed = s\sqrt{\frac{1}{n_a} + \frac{1}{n_b}}$ where n_a and n_b represent the number of independent observations for the two varieties. This quantity can be used to calculate the minimum distance that must be found between two varieties in order to prove that the varieties are significantly different for the observed character. Assuming that the distribution of the variable in question is normal, this minimum distance can be calculated as $LSD^F = sed \times t_{f,1-\alpha/2}$, where $t_{f,1-\alpha/2}$ is the $1-\alpha/2$ fractile of a t-distribution with f degrees of freedom, where f is $(n_a + n_b - 2)$.

The purpose of doing statistical analysis is usually both to estimate the parameters of interest, such as the mean yield of each variety and the mean difference between pairs of varieties as well as to test whether some hypotheses can be accepted or have to be rejected. In a simple situation such as a randomised complete block design without missing values the estimates of the mean yields of a variety are simply the averages over all observations on that variety. The estimated difference is simply the difference between the averages of the varieties. In more complicated designs or when some observations are missing the estimation is more complicated as it is necessary to use methods that take into account other factors such as the blocks in which a given variety is present.

Statistical tests are performed in order to know whether a hypothesis can be accepted or has to be rejected. Examples of such hypotheses could be the hypothesis that all varieties have the same yield, that the difference between variety A and B is zero or that all varieties react in the same way to nitrogen. The tests are made on some predefined significance levels, usually called α (alpha). There is a strong tradition to take α equal to either 5% or 1%. If α is 5%, it is said that the test is performed on the 5% level of significance. A significant result means that the hypothesis has to be rejected, i.e. the difference between variety A and B is different from zero at the 5% level of significance. A non-significant result means that the hypothesis can be accepted, i.e. the difference between variety A and B is not different from zero at the 5% level of significance. Note that this does not mean that the difference is zero; it only means that with the used number of replicates, the chosen design and the actual random variation there is no reason to conclude that the difference is not zero.

The application of statistical tests always implies some risks of making wrong decisions. These are usually separated into two types of risks. They are called Type I error and Type II error, respectively. The type I error is the error that arises when we decide the varieties to be distinct, when they are in reality identical. The type II error is the error that arises when we decide the varieties to be identical, when they are in reality different. The risk of type I error can be controlled easily as the risk here is α, whereas the risk of type II error, usually called β (beta), is more difficult to control as it depends on the size of the real difference between the varieties, the random variability, *s*, and the chosen design (number and replicates and lay-out in the field).

2. Experimental Designs

2.1. Type of designs

Randomised complete block design (CBD)

The experimental field is divided into blocks according to the number of replicates. Each Block is divided into a number of plots according to the number of treatments. The treatments are then assigned randomly to the plots. Each treatment occurs one time per block.
A benefit of block designs over completely randomised designs is, that differences between blocks (e.g. due to soil quality) do not influence the estimates of treatment differences and can be separated from the experimental error when performing analysis of variance.
One drawback of the CBD is that only soil differences in one direction can be modelled. Possible extensions of the block design for two directions are the Latin square, allowing for row and column effects.

A CBD is a good choice when there are no technical aspects that restrict the randomisation. Simple block designs are mostly used for one-factorial trials but two or more factors are also possible. The layout of blocks on the field has to be chosen in such a way, that soil differences between blocks are maximised and within blocks are minimised. Homogeneity of conditions within blocks requires that the treatment number and therefore the dimension of the blocks have an upper limit. Depending on plot size and soil conditions block designs are recommended for trials up to 20 treatments. In block designs the assumption is usually made that there are no interactions between treatments and blocks.

Fig. 1. A randomised complete block design with 5 treatments in 4 complete blocks.

block 1	A	E	B	D	C
block 2	C	D	A	E	B
block 3	E	B	D	C	A
block 4	E	D	A	B	C

Incomplete block design (IBD)

In trials with high treatment numbers, e.g. variety trials, complete blocks are too large to give a good control of the experimental error due to soil heterogeneity. In these cases designs with incomplete blocks are useful. Every block only contains a fraction of the total number of treatments and is therefore *incomplete*. Several incomplete blocks form one complete replication. One type of such designs is the *lattice design*. The blocks of an incomplete block design can be arranged in any way that is useful for controlling soil heterogeneity.

With an IBD the arithmetic mean of a treatment is not the best estimator for the expected mean value. Treatment means have to be adjusted according to the linear model used for data analysis. One should use powerful software for the analysis (ALPHA+, GenStat and SAS).

Specialist software is also needed for the construction of the design (e.g. Alpha+ or CycDesigN, http://www.cycdesign.co.nz).

There are several types of lattice designs:

a. Square Lattices need a quadratic or cubic number of treatments (9, 16 and 25). The number of plots per block (k) has to be the square root of the number of treatments (v). For example 36 treatments in 6 blocks of 6 plots per replicate.

b. Rectangular Lattices: The number of treatments has to equal k(k+1) with k= number of treatments per block. This algorithm allows for treatment numbers like 12 or 20.

c. Alpha-designs: More flexibility is reached with the new class of alpha designs or generalised lattices (Patterson & Williams 1976, Patterson et al. 1978). The following requirements have to be met: (1) The number of plots per Block (k) has to be smaller or equal to the square root of the number of treatments (v). (2) The number of replicates has to be smaller or equal to the ratio v/k. (3) The number of treatments has to be a multiple of k. Where the number of treatments does not meet these conditions, a design for the next possible number is developed and the redundant treatments are discarded.

Fig. 2. Example of an incomplete block design with 16 treatments in 3 complete replications. The replications are divided into 4 incomplete blocks with 4 plots each.

```
Blocks of the design printed in rows

 rep  1 ----------------
 plot      1   2   3   4
 block +----------------
   1   |  15   3   1   8
   2   |  11   9   5  16
   3   |   7  14  12  13
   4   |  10   6   2   4

 rep  2 ----------------
 plot      1   2   3   4
 block +----------------
   1   |   3   8   6  16
   2   |   2  10   1  13
   3   |   7   4  12   5
   4   |  14  15  11   9

 rep  3 ----------------
 plot      1   2   3   4
 block +----------------
   1   |   4  13  11   3
   2   |   1  12   6   9
   3   |  10  14   8   5
   4   |  15   7  16   2
```

Split plot design

This type of design is often advantageous for factorial trials when one factor can not be allocated to small plots for technical reasons or when the factors should be tested with different precision.

Imagine a two factorial trial (tillage 1 and 2 and varieties A, B, C, D, E) with three replicates. First each block is divided into two main plots. The factor, tillage, is then allocated randomly to the plots. Each main plot is then divided into as many sub-plots as the second factor has levels, here 5. Then the levels of the second factor are allocated randomly to the sub-plots within the main plots.

In the analysis of variance the main plot factor has to be tested against the interaction main plot factor x block (the main plot error), whereas the sub plot factor is tested against the residual. Because the main plot factor is tested with less precision and with only a low number of degrees of freedom for the error term, usually only large differences become significant. A difference in sub plot factor means normally show much smaller standard errors. Since more than one error term occurs in split plot designs, the analysis should be performed in a mixed model framework. A description of the analysis of split plot trials is given in 3.3.

Fig. 3 A split plot design with 2 treatments for the main plot factor (1 and 2), 5 treatments for the sub plot factor (A-E) and 3 complete blocks.

block 1	1-A	1-E	1-B	1-D	1-C	2-C	2-D	2-A	2-E	2-B
block 2	2-E	2-B	2-D	2-C	2-A	1-E	1-D	1-A	1-B	1-C
block 3	2-B	2-C	2-A	2-E	2-D	1-D	1-A	1-C	1-B	1-E

2.2. Trial set up and design

What type of design to choose?

Depending on the plot size and soil conditions complete block designs are recommended for trials up to 20 treatments. With higher treatment numbers incomplete block designs will normally give results with a lower standard error. Because of their great flexibility we recommend to use alpha-designs.

Complete blocks, incomplete blocks and split plot design can be combined in different ways to meet the technical and statistical requirements. The chosen structure may not be covered by examples in statistical textbooks. The only requirement is that the principles of replication and randomisation are kept in mind and that the model used for analysis is based on the randomisation structure of the trial (see Piepho et al. 2003 for details).

Number of replicates

For single trials four replicates are often recommended. But four replicates may not be enough to give results with a standard error of mean that is small enough to distinguish interesting treatment means significantly. Compared with randomised greenhouse or laboratory experiments, field trials utilise an extremely small numbers of replicates due to practical restrictions. Table 1 presents the detectable difference Δ as a k-fold of the standard deviation for the two-sided t-test for different numbers of replicates with a maximum false negative rate of 20% (Type II error) and the common false positive rate of 5% (Type I error). In a field trial with replicate or plot size of four, only effect differences larger than 2.02*SD will be detected with a maximum false negative rate (β) of 20% and a maximum false positive rate (α) of 5%.

Table 1. Detectable relative difference (Δ = Diff / SD) for various numbers of replicates with nominal values of α=0.05 and β=0.20 for type-I and Type-II experimental error respectively.

number of replicates	2	3	4	5	6	10	20
Δ (for α=0.05; β=0.20)	3.07	2.38	2.02	1.80	1.70	1.33	0.91

In trial series in different environments estimating the genotype x environment interaction is much more interesting than exact results in single trials. Therefore two or three replicates per location will be sufficient when the number of locations is high enough.

Block size and shape

The optimal block size and shape depends on the heterogeneity of the experimental field. If no additional information is available, a quadratic shape of the blocks is the best choice. The larger the blocks are, the higher the experimental error will be due to differences in the soil conditions. With more than 20 treatments, a lattice design (e.g. square lattice, generalised lattice) is recommended.

Plot size and shape (and guard areas)

A plot size larger than 20 square metres is seldom reasonable in variety trials. When the total experimental area is fixed, many small plots give a better control of the experimental error than a few large plots. Differences in the soil quality will be distributed more evenly on the different treatments. The minimum size of the plots also depends on the dimensions of the machinery to be used. A plot size between 5 and 20 square metres is commonly recommended. Variety trials are mostly performed in narrow plots. This has some technical advantages. For example if the harvester has a working width of 150 centimetres, it is practical to use plots of 150 centimetres wide.

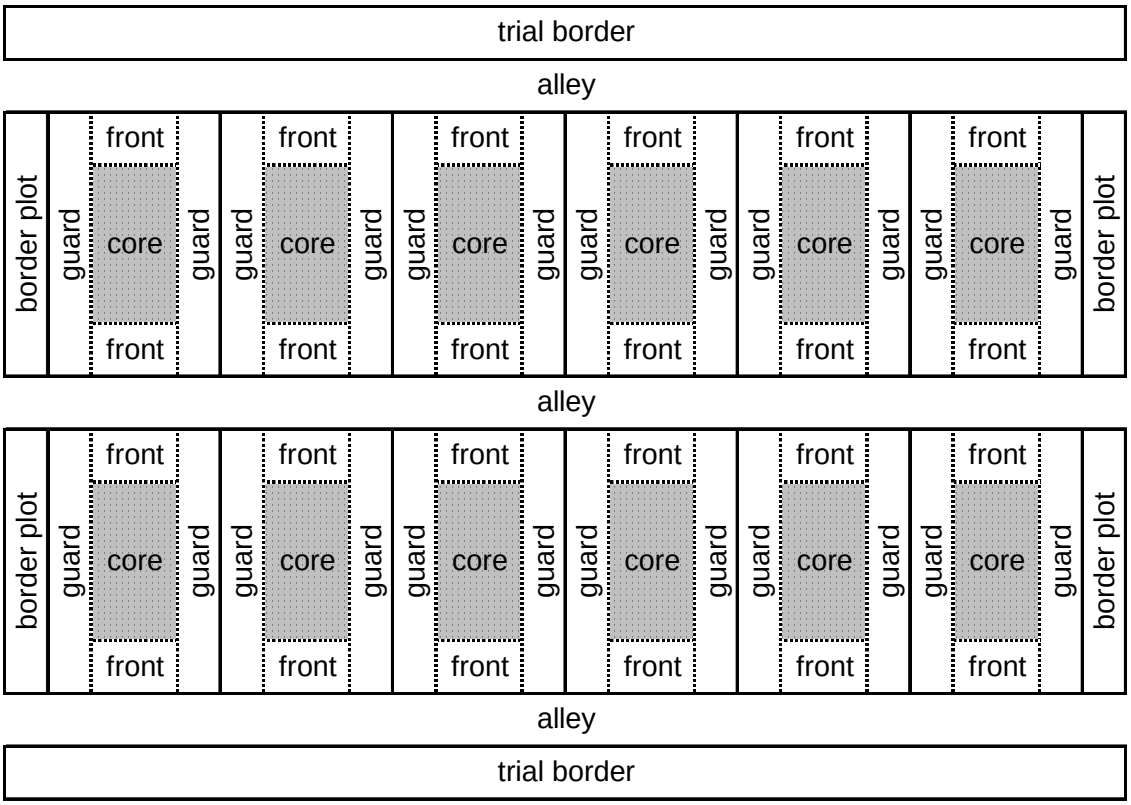
Interplot interference can affect estimates of yield, quality and disease resistance due to differences in competitiveness of the tested genotypes (Talbot et al. 1995, Clarke et al. 1998). Interference may be caused by differences in plant height with consequent competition for light, and also by differences in disease resistance.

How can we reduce interference?

a. Sow wider plots: harvest only the core of the plot, discard the guard rows.
b. Grouping of cultivars: e.g. cultivars can be divided into a “short”, “intermediate” and “tall” groups (if height is an issue) and tested in a split plot like design avoiding tall cultivars neighbouring short cultivars (see David & Kempton 1996).
c. Use of covariates: correlated traits (plant height!) can be used as covariates, which could have an adjustment for competition (see Goldringer et al. 1994).
d. Modelling of neighbour-effects: fit linear model with additive effects for cultivar and neighbour and calculate adjusted estimate for pure stand

An additional problem can occur at the front of plots. The plants located at the edge of plots have much better conditions to grow, because of the additional amount of light and nutrients available at the alleys between blocks. If genotypic differences are expected in using these better growing conditions, the front area of the plots should also be discarded (see Fig. 4).

Fig. 4. Plots with core areas and different types of borders and guard areas.



3. Analysis of data

3.1. Evaluation of data

3.1.1. Check for errors and assumptions

Every statistical analysis of trial data needs some assumptions to be fulfilled, otherwise the conclusions may be false. Among these assumptions the most common (for analysis of variance) are:

- independence of observations,
- normality of distribution,
- additivity of treatment and block effects,
- homogeneity of variances,
- lack of outliers.

Independence. In the majority of statistical methods used for analysis of trial data, the independence of observations is a key assumption. On the other hand, it is commonly known that – for example in field trials – observations from adjacent plots are likely to be more similar than observations distant from each other. So, usually observations are correlated. Luckily a proper randomisation prevents statistical analysis from giving biased results. There are some statistical tools to detect correlations (lack of independence) between observations but for above-mentioned reasons there is no need to present them here.

Normality. All the tests used in analyses of variance and analyses of regression are based on normality assumption. Normality means that the distribution of observations is "bell shaped" for all treatments under comparison. Mead et al. (1983) say "in most situations it is impossible to decide by examining the data whether the assumption of normality is reasonable and one has to rely on common sense in arguing whether the assumption is biologically likely". So this assumption is rather difficult to be verified unless the sample size is very large. There are some tests for checking this assumption but all of them are rather weak (in the sense that they very rarely reject the null hypothesis) when sample sizes are small and even moderately large. So they can be applied only for large sample sizes (sample size tending to infinity). As in routine experimentation the number of replicates is small (usually smaller than 6) and the sample size for a particular treatment is of the same order, the use of such a test is not possible. Graphical presentation of data can provide a visual inspection for lack of normality. Luckily the tests used in the analysis of variance (as well as regression), namely the F-test and t-test, are resistant against moderate deviations from normality. A method that is often used to check normality is the Shapiro-Wilk test, which is recommended for sample sizes not larger than 50 (Shapiro and Wilk, 1965).

Additivity. In the analysis of variance of block trials (CBD or IBD, see section 2.1) it is assumed that there is no interference between blocks and treatments. In practice this means, that differences between any two treatments are the same in all blocks in which they appear together and that possible fluctuations are caused solely by experimental error. This assumption is usually fulfilled if the differences between blocks are not very large. When blocks differ considerably, e.g. an average yield of 20 kg/plot in one block and of 50 kg/plot in the other block, it is not reasonable to expect that the difference between two varieties of 4 kg in the first block will be of the same magnitude in the second block. A simple test for non-additivity in a CBD design was proposed by Tukey (1949), known as "one degree of freedom for non-additivity". In this approach the sum of squares for error is subdivided into two parts. One is

attributed to non-additivity, the other to the residual. Then, using the usual Fisher F-test with one degree of freedom for the numerator, the hypothesis that there is lack of additivity is tested. In the case of multiplicative effects, a logarithmic transformation can improve the situation.

Homogeneity. The typical assumption in an analysis of variance is that the treatments do not influence the variance of experimental error, in other words that the variance is the same for all treatments. This assumption is likely to be fulfilled when levels of expression are similar for all treatments. When levels of expression (mean values) differ considerably between treatments, normality and additivity as well as homogeneity of variances can be violated. This assumption can be verified using Bartlett's or the Cochran test. In both tests, the estimates of variances are calculated for all treatments and next the hypothesis of equal variances is tested against the alternative that some of them (at least one) are different.
If the variances (standard deviations) are related to the level of expression (mean values) of the characteristic that is analysed, a logarithmic (or square root) transformation can improve the situation.

Outliers. All the statistical analyses of trial data are carried out (possibly after checking all underlying assumptions) assuming that all collected data is correct. However, this is not always the case. Errors can occur when recording, copying or preparing data for computer processing. When such an error observation is out of the expected range of observations it is easily detected by a visual inspection of the data. Sometimes it can be detected after preliminary analysis, for example if such an observation "produces" an extremely high residual. In general, an observation is considered as an outlier if its value differs considerably from all other observations. If the value of one (or more) observation is far from the cloud of all other observations it is likely to be an outlier. The easiest statistical method to detect outliers is as follows:

a. order all *n* observations in ascending or descending manner,
b. temporarily remove the 'suspected' observation from your sample (it is either the smallest or the largest observation),
c. calculate the (1-α) confidence limits for single observations by using the rest (*n*-1) of the observations (see footnote [1])
d. if the 'suspected' observation is out of the calculated confidence limits, it is considered as an outlier and the reason for this should be checked. If the deviation is caused by a simple typing error, the error should be corrected. For other reasons such as damage to the plot caused by external factors independent of the treatment, the observation should be permanently rejected from the sample and be treated as missing data. If no reason can be found for the deviation the observation should be kept unless the deviation is so large that it will make the analysis unreliable (in such cases it may be wise to run the analysis twice – both with and without the outlying observation to see if the conclusion will change).

[1] For normal distribution, the lower X_* and upper X^* confidence limits are of the form:

$$X_* = \overline{X} - t_{\alpha/2}s \qquad X^* = \overline{X} + t_{1-\alpha/2}s,$$

where $\overline{X}$ is the mean value calculated over (n-1) observations, $t_{\alpha/2}$ and $t_{1-\alpha/2}$ are the Student t-distribution table values with (*n*-2) degrees of freedom and *s* is the standard deviation calculated over (*n*-1) observations. If there are more such "suspect" observations, the whole procedure can be repeated.

3.1.2. Usefulness of the data for investigations

The choice of the most appropriate data to answer the question put in the investigation is not always simple and straightforward. In most cases, the investigator is forced to accept a compromise between the precision of the conclusions and the cost of the data. Clearly, cheap data is often sufficient to answer simple questions. However, this does not mean that expensive data will guarantee better reliability and accuracy of results and conclusions.

In general, the data will be useful for the investigations if the experiment in which they have been collected was properly designed. If the experimental design is faulty, no data cleaning, filtering, outlier detection or other processing techniques will be helpful. Also, no statistical method of data analysis is going to help to make proper conclusions. Statistical handbooks are full of recipes of how to properly plan experiments. It is noteworthy that rules as old as the ones given by Finney (1953, p. 173) are still valid.

Data used in statistical analyses are observations of random variables. The statistical procedures work only if there is a variability of the observations. The source of this variability must be known to the investigator if the conclusions are to be sensible. Thus, data obtained from carefully designed experiments are more valuable than data from observational studies or extracted from databases with an incomplete description of origin. A helpful discussion of this problem in the context of regression is given by Gomez and Gomez (1984, p. 417).

3.1.3. Transformations.

When one or several of the mentioned assumptions is violated, the performed analysis is incorrect and decisions may be false. If, using some statistical tool or just after visual inspection of the data, deviation from the standard situation is detected, it is sometimes possible to 'improve' the situation by transforming data and then analysing the transformed data. Depending on which assumption is violated, several transformations may be applied. The most commonly used transformations are:

Logarithmic transformation. This transformation is appropriate for data in which there is proportionality between mean values and standard deviations or when effects are multiplicative. Typical examples of such data (see Gomez and Gomez, 1984) are data concerning the number of insects per plot or the number of egg masses per plant (or per unit area). To transform original data (X), into a logarithmic scale (Y), simply take $Y = \log(X)$. When some observations are small (smaller than 10), the transformation $Y = \log(X+1)$ is suggested. The fulfilling of all assumptions must be again checked for the transformed data. If there are no serious deviations from assumptions, all the analyses and tests are applied to the transformed data. After performing treatment comparisons the mean values can be re-transformed into original scale.

Square root transformation. This transformation is useful and effective for data for which variance tends to be proportional to the mean. This transformation can be applied for data consisting of small whole numbers. Such data appear when rare events are counted (in limited time or space). Typical examples are the numbers of weeds per plot (or per square metre) or the number of insects caught in traps. The square root transformation may also be appropriate for percentage data where all observations are in the range from 0% to 30% or in the range from 70% to 100%. For intermediate data (all observations between 30% and 70%) this transformation is usually not necessary. To apply this transformation simply calculate $Y=\sqrt{X}$ when all observations are in the range between 0 and 50 and calculate $Y=\sqrt{100-X}$ when all observations are in the range between 50 and 100. Again all the analyses are performed using

transformed data. The final results (treatment means) can be presented on transformed scale or can be re-transformed to original scale.

The arcsine transformation. To perform this transformation, use the formula $Y=\arcsin\sqrt{X/100}$. This transformation is appropriate for data concerning fractions and expressed as percentages. Usually data obtained by dividing two counts (e.g. number of deceased plants and total number of plants) can be transformed using this transformation. The extreme values of 0% and 100% are to be substituted by (1/4n) and (100-1/4n) respectively before using *arcsine* transformation. This transformation can be easily performed using a computer or the tables of C.I. Bliss (1934) reproduced in many statistical textbooks. Because percentage data can also be transformed using other transformations, the practical advice is as follows (Gomez and Gomez, 1984):

- for percentage data from the range between 30% and 70 % no transformation is needed,
- for data sets where all data are in the range between 0% and 50% (or between 50% and 100%) the square root transformation is appropriate (see the text above),
- for the data that do not belong to any of above-mentioned ranges the *arcsine* transformation is to be used.

The logit transformation. This transformation is applicable for percentage data expressed as fractions. To perform this transformation simply apply formula Y=ln(X/(1-X)), where X is the fraction to be transformed. Please note that this transformation is undefined for X=0 and X=1. This transformation is much simpler but almost indistinguishable from *probit* transformation described by Bliss [1934]. The logit transformation may be used to analyse the relative number of insects killed by different doses of an insecticide.

Additional remark. If there is lack of homogeneity of variance in a data set and there is no relationship between means and variances (standard deviations), a possible solution is to split treatments into groups with similar (homogeneous) variances and perform independent analyses of variance for each of these groups or apply more advanced methods such as weighted analyses of variance or methods that allow the variance to be different (by using some approximations).

Instead of applying transformations to the recorded observations some characteristics e.g. percentages and counts may alternatively be analysed using generalised linear mixed models (see section 3.5).

3.2. Methods for analysis

3.2.1. Analysis of variance, F-tests, LSD-values

3.2.1.1. Randomised complete block design

Analysis of variance (ANOVA) is the main tool used for statistical interpretation of agricultural trial data. The analysis of variance is based on linear model of observation. For experiments performed in a randomised complete block design (CBD), the linear model is of the form

$$y_{ij} = \mu + \tau_i + \beta_j + e_{ij} \qquad (1)$$

where y_{ij} denotes the value of observed trait for the i-th treatment (i=1,2,…, t), received in the j-th block (j=1,2,…,r) with a total number of observations $n = rt$; τ_i is the fixed effect of the i-th treatment, β_j is the effect of the j-th block and e_{ij} is an experimental error associated with observation of the i-th treatment in the j-th block.

Different assumptions can be made on the block effects β_j.
If the assumption is that β_j is fixed, meaning that the only random term in (1) is e_{ij}, the model is called fixed. In that case all conclusions are confined to treatments and blocks used in the analysed experiment.
More common is to consider β_j as the random component of model (1). In this case the model is called mixed. Such a model can be set up using the principle of randomisation, see Caliński and Kageyama (2000).
In the mixed model the blocks are treated as a random sample of an infinite set of all possible blocks and conclusions are not confined to the blocks actually used in experiments. The conclusions are "valid" in the population of blocks from which the blocks can be considered as a random sample.
Analysis of variance of trial data is based on a division of the sum of squares of total variability (SS_c) into a component attributed to blocks (SS_b) a component attributed to treatments (SS_t) and to the error (SS_e) according to the equality

$$SS_c = SS_b + SS_t + SS_e \qquad (2)$$

Usually the main aim of the analysis of variance is to test the hypothesis, that there are no differences between treatments under comparison, namely the hypothesis

$$H_0: \tau_1 = \tau_2 = \ldots = \tau_t \quad \text{against} \quad H_1: \text{"}H_0 \text{ is not true"} \qquad (3)$$

This hypothesis is always tested by application of a Fisher F-test of the form

$$F_0 = MS_t / MS_e,$$

where MS_t and MS_e are the mean squares for treatment and error respectively. Usually the results of ANOVA are presented in an analysis of variance table as in table 2.

Table 2. Analysis of variance for a randomised complete block design (CBD)

Source of variation	Degrees of freedom	Sums of squares	Mean squares	F
Blocks	r-1	SS_b	MS_b	
Treatments	t-1	SS_t	MS_t	F_0
Error	(r-1)(t-1)	SS_e	MS_e	
Total	n-1	SS_c	-	-

If $F_0 > F^{\alpha}_{(t-1);((r-1)(t-1))}$, where $F^{\alpha}_{(t-1);((r-1)(t-1))}$ is the critical value of the F distribution for (t-1) and (r-1)(t-1) degrees of freedom at α significance level, the hypothesis (3) is rejected, meaning that not all treatments are the same (some treatments differ from the others). If hypothesis (3) is rejected, the researcher is usually interested to identify which pairs of treatment are different. To answer this question usually the so-called least significant difference (LSD) is calculated. If the researcher is interested in one particular comparison (that was chosen before establishing the experiment), the best way is to calculate the Fisher LSD^F, using formula

$$LSD^F = sqrt(2*MS_e/r)*t^{\alpha}_{\nu},$$

where MS_e is taken from the analysis of variance table and t^{α}_{ν} is the two-sided t-Student distribution critical value at α significance level for $\nu=(r-1)(t-1)$ degrees of freedom. If the absolute value of the difference between treatment-means calculated for e.g. treatment *1* and *2* is bigger than LSD^F, these two treatments are declared significantly different at α significance level. If more than one comparison with the use of LSD^F is made, the general significance level (for all comparisons) is larger than α.

If many comparisons between treatments are planned, it is recommended to use a method that minimises the risk of erroneously declaring pairs significant, such as the Tukey LSD^T which is of the form

$$LSD^T = sqrt(MS_e/r)*q^{\alpha}_{t,\nu},$$

where $q^{\alpha}_{t,\nu}$ is the critical value from studentised range distribution read at α significance level for t treatments involved in comparisons and ν degrees of freedom (degrees of freedom for error in the ANOVA table).

The rules of using LSD^T are the same as for LSD^F, but now all treatment comparisons can be made and still ensure that the risk of erroneous declaring any of these significant will be less than α.

3.2.1.2. Incomplete block design (alpha design).

A slightly more complicated situation appears in the case of incomplete block design (which includes the alpha designs). Because blocks and treatments are not orthogonal to each other (which it was in CBD), the division of the total sum of squares into parts attributed to blocks and treatments is not unique. Usually the ANOVA table instead of single sum of squares for blocks (as in CBD), will mention two sums, the first attributed to complete replicates (superblocks), the second attributed to blocks (within superblocks) – ignoring treatments.

The linear model of observations in alpha design is of the form

$$y_{ijk} = \mu + \tau_i + \rho_j + \beta_{jk} + e_{ijk} \quad (4)$$

where y_{ijk} denotes the value of the observed trait for i-th treatment received in the k-th block within j-th replicate (superblock), t_i is the fixed effect of the i-th treatment ($i = 1,2,\ldots,t$); ρ_j is the effect of the j-th replicate (superblock) ($j = 1,2,\ldots,r$); β_{jk} is the effect of the k-th incomplete block within the j-th replicate ($k = 1,2,\ldots s$) and e_{ijk} is an experimental error associated with the observation of the i-th treatment in the k-th incomplete block within the j-th complete replicate. There are $n = rt$ observations in total. The whole experiment consists of rs incomplete blocks forming r complete replicates. The whole discussion concerning randomness of blocks in randomised complete block design also applies to incomplete blocks and complete replicates in alpha design. In accordance with the linear model of observations (4), the analysis of variance is usually presented in the form given in table 3.

Table 3. Analysis of variance for alpha design

Source of variation	Degrees of freedom	Sums of squares	Mean squares	F
Replicates	r-1	SS_r	MS_r	
Blocks (within replicates, ignoring treatments)	rs-r	SS_b	MS_b	
Treatments(adjusted for blocks)	t-1	SS_t	MS_t	F_0
Error	rt-rs-t+1	SS_e	MS_e	
Total	n-1	SS_c	-	-

The term “ignoring treatments” means that the sum of squares for blocks is not free of treatment effects. Instead of the sum of squares for treatments (as for CBD), the sum of squares for treatments adjusted for block effects appear. It means that this sum of squares is free from block effects. The hypothesis tested is the same as in CBD (see (3)) and it is verified in exactly the same manner using a Fisher F-test. The value of $F_0 = MS_t / MS_e$ is now compared with the critical $F^{\alpha}_{t-1,\ rt-rs-t+1}$ value with t-1 and rt-rs-t+1 degrees of freedom. Treatment means are now not just simple averages over replicates as in CBD but are “adjusted”. This adjustment is different for a fixed model of observation (in so-called intra-block analyses) and for a mixed model (in analyses with recovery of inter-block information). Additional difficulties arise when LSD is applied for treatment comparisons. Due to the lack of orthogonality, the variances of treatment comparisons (treatment contrasts) will often be different for different pairs of treatments. So in an extreme case for every pair of treatments a specific LSD (Fisher or Turkey) should be applied. However for moderate variations it may be acceptable to average the variance of treatment-comparisons and then use the average LSD value. But in this situation comparisons must be made with special caution. Usually the design is chosen so that the difference between the largest and the smallest variance of treatment comparisons is as small as possible. This means that balanced designs are preferable

3.2.1.3. Split-plot design

As described in 2.1, the split-plot design is applicable for two-factorial trials. The mathematical model of observations reflects the situation that experimental units (plots) of two different sizes appear. This implies that two different errors related to these plot sizes are present in a model of the form

$$y_{ijk} = \mu + r_i + a_j + \eta_{ij} + b_k + (ab)_{jk} + e_{ijk} , \quad (5)$$

where y_{ijk} denotes the observations from experimental unit from i-th block (i=1,2,…r), concerning j-th level of main plot factor A (j = 1,2,…,a) and k-th level of sub-plot factor B (k = 1,2,…,b). r_i is the random effect of i-th block; a_j is the fixed effect of j-th level of factor A; b_k is the fixed effect of k-th level of factor B; $(ab)_{jk}$ is the fixed effect of interaction of j-th level of factor A with k-th level of factor B, and, finally, η_{ij} and e_{ijk} are the errors connected with main plots and sub-plots respectively.

There are n = rab observations in total. The analysis of variance of split-plot data is based on the division of sum of squares of total variability SS_t into the following components

$$SS_t = SS_b + SS_A + SS_\eta + SS_B + SS_{AB} + SS_e, \quad (6)$$

where SS_b, SS_A, SS_η, SS_B, SS_{AB} and SS_e denote sums of squares attributed to blocks, factor A, main-plot error, factor B, interaction of factor A and B, and sub-plot error, respectively.

The traditional form of the related analysis of variance table is as follow:

Table 4. Analysis of variance for split-plot

Source of variation	Degrees of freedom	Sums of squares	Mean squares	F
Blocks	r-1	SS_b	MS_b	
Levels of factor A	a-1	SS_A	MS_A	F_A
Error η	(r-1)(a-1)	SS_η	MS_η	
Levels of factor B	b-1	SS_B	MS_B	F_B
Interaction A*B	(a-1)(b-1)	SS_{AB}	MS_{AB}	F_{AB}
Error e	a(r-1)(b-1)	SS_e	MS_e	
Total	n-1	SS_c	-	-

Usually, in a split-plot design three hypotheses are tested. First, the hypothesis of no differences among effects of factor A is tested, then the hypothesis of differences among effects of factor B is tested and finally the hypothesis of no interaction between levels of factor A and B is tested. Formally, the hypotheses tested are:

no effects for factor A: H_{0A}: "$a_1 = a_2 = \ldots = a_a$" against H_{1A}: "H_A is not true"

(the appropriate F statistics $F_A = MS_A/MS_\eta$ is compared with the F-distribution critical value at chosen α significance level with (a-1) and (r-1)(a-1) degrees of freedom),

no effects for factor B: H_{0B}**:** "$b_1 = b_2 = \ldots = b_b$" against H_{1B}: "H_B is not true"

(the F statistic to verify it is $F_B = MS_B/MS_e$, with (b-1) and a(r-1)(b-1) degrees of freedom), and finally the hypothesis that there is no interaction between Factor A and factor B, namely:

no interaction between factor A and factor B: H_{0AB}: "$(ab)_{11} = (ab)_{12} = \ldots = (ab)_{ab}$" against H_{1AB}: "H_{AB} is not true"

(the F statistic to verify it is $F_{AB} = MS_{AB}/MS_e$, with (a-1)(b-1) and a(r-1)(b-1) degrees of freedom).

After rejecting these hypotheses the researcher is "entitled" to make comparisons between levels of appropriate factors. The researcher can use either LSD^F or LSD^T values as a threshold of significance between levels. Here only formulas for LSD^F are given but they can easily be modified to LSD^T. So, to compare two levels of factor A, the appropriate LSD^F value is calculated as

$$LSD^F = sqrt(2*MS_\eta/rb)t^\alpha_\nu,$$

where t^α_ν is the two sided t-Student distribution critical value read at α significance level for $\nu=(r-1)(a-1)$ degrees of freedom. The rules to use this LSD are exactly the same as in one-factorial designs (e.g. in CBD design).
To compare two levels of factor B, the LSD is calculated using the formula

$$LSD^F = sqrt(2*MS_e/ra)t^\alpha_\nu,$$

where ν are degrees of freedom associated with MS_e, i.e. $\nu = a(r-1)(b-1)$.
If the hypothesis H_{AB} is rejected some additional comparisons are possible. One can compare two levels of factor A within the particular level of B, or two levels of factor B within the particular level of factor A, or any combination of levels A and B. The appropriate LSD value for comparing two levels of factor B within a chosen level of factor A is calculated using the formula

$$LSD^F = sqrt(2*MS_e/r)t^\alpha_\nu,$$

where ν are degrees of freedom associated with MS_e. The formulas for other comparisons can be found in the literature (see e.g. Gomez and Gomez, 1984). To apply LSD^T instead of LSD^F, the presented formulas can be easily modified in a similar way to the description for CBD design.

Additional remarks on application of the split-plot design.
When there are only two levels of factor A (or B), there is no need to calculate LSD values to make a comparison of these two levels as rejecting the hypothesis H_{0A} (or H_{0B}) means that the two levels differ significantly. Another way of analysing the data from split plot trial is to split the overall (described here) analysis into independent analyses made within each level of factor A. It means that if there are "a" levels of factor A, then "a" separate CBD analyses are performed. But such an approach has two disadvantages: (1) the number of degrees of freedom for error for each partial analysis is smaller than the number of degrees of freedom for error "e" in a full split plot analysis and (2) the separate analyses do not allow to test the presence of interactions between levels of factor A and B, which are often the most interesting.

3.2.2. Analysis of variance including covariates (ANCOVA).

One of the aims of the researcher in the choice of an experimental design, the choice of plot size and shape, the choice of a mathematical model of observation etc. is to decrease the variance of the experimental error. The estimate of this variance is the mean square for error MS_e (appearing in the ANOVA table and in the denominator of the F-test) and the smaller the value is, the higher the probability of rejecting the null hypothesis of no difference between treatments and the higher the chance of declaring significant differences between chosen pairs of treatments.
Analysis of co-variance (ANCOVA) is one method that may be used to reduce the size of the error MS_e. Analysis of co-variance is a method of analysis that can be used to eliminate effects resulting from variables in which there is no interest. An example of such a variable in field experiments is the number of plants in a plot. Different numbers of plants for different treatments can influence the final results and decisions. Assuming that there are two variables observed in an experiment, the main variable *Y* and the additional variable *X* where *X* can influence *Y* but *X* is not influenced by the treatments (e.g. measured before the treatments are applied), then ANCOVA may be used to remove (at least partly) the influence of *X* on *Y*.

Analyses of co-variance consist of three parts: analysis of variance for main variable *Y*, analysis of variance for additional (also called concomitant) variable *X* and regression analysis of variable *Y* on *X*. The mathematical model of observations in ANCOVA is the same as for ANOVA but is extended by a term related to regression. So, for an experiment performed in CBD this model is of the form:

$$y_{ij} = \mu + \tau_i + \beta_j + \gamma x_{ij} + e_{ij} \qquad (7)$$

where the meaning of used symbols is the same as in formula (1) with the additional symbol γ used for denoting the common (for all treatments) coefficient of regression of the main variable on the concomitant variable and x_{ij} denotes the value of the concomitant variable observed for the *i*-th treatment in the *j*-th block. The x_{ij} is assumed to be fixed and not to be influenced by the treatments. Usually in ANCOVA three hypotheses are tested in turn:
– the hypothesis of no differences between treatments for the concomitant variable. If such differences exist, it usually means that values of the concomitant variable are influenced by treatments and ANCOVA should not be applied;

– the hypothesis that there is a significant linear relationship between variable *Y* and *X*. If there is no such relationship (regression is not significant), ANCOVA can formally be applied but is ineffective in decreasing the experimental error;
– the hypothesis that there are no differences between treatment-means for the main variable adjusted for values of the concomitant variable.

In a similar way the model for analysis of variance for alpha designs can be extended to the ANCOVA model. It is possible to include more concomitant variables.

When comparing treatments after an analysis of covariance, the variance on treatment comparisons is additionally influenced by different values of the concomitant variable for each treatment. The average influence of the concomitant variable on the variance of comparisons can often be applied. One of the possibilities is to apply the approximation proposed by Finney (1946). More information on the interpretation of ANCOVA analysis can be found in Little and Hills (1978).

3.2.3. Regression analysis

Regression is a statistical method to describe the association between two or more observed variables (traits) or between one observed variable and a design parameter (such as the amount of applied nitrogen or the year in which the observation is recorded). In the situation of two observed variables it can be used to estimate the effect of one of them (the assumed predictor variable) on the other (the assumed response variable), by expressing the response variable as a function of the predictor variable. Which variable is taken as predictor and which as response is a matter of biological knowledge; the basic regression methods do not check these assumptions. The simplest choice of the function linking the two variables is a linear function. It is equivalent to assuming a constant change of the value of the response variable for each unit change of the predictor variable in the whole range of observations. If we denote the observations of the predictor variable by X_i, $i = 1,2...n$, and the observations of the response variable by Y_i, the linear regression means that

$$Y_i = a + bX_i + e_i,$$

where *a* and *b* are regression coefficients, and e_i is a random deviation of the *i*-th observation of Y from the exact linear relationship. The values of *a* and *b* are calculated using the principle of "least squares". The process of calculation is sometimes called "fitting". Statistical significance of the regression coefficients can be tested by a *t* test. The equation implies that expectation (mean value) of Y_i is equal to $(a + bX_i)$. In mathematical statistics expectation expressed in terms of a variable (in this case - X) is called conditional expectation. Thus, the fitted regression function informs us about the expected (mean) value of the response variable for a chosen value of the predictor variable. The values of *a* and *b* can only be interpreted when X_i is measured without error, as the values of *a* and *b* are biased if the variable X_i is influenced by random variation – although the formula can be used for prediction in both situations.

Although regression analysis is a computational estimation method, it has several important connections with less formal graphical exploratory procedures. This is not strange knowing that any consideration concerning two observed variables can be conveniently illustrated by simple two-dimensional Y-X scatter plots. The role of graphical data exploration is two-fold:
– before computation, the scatter plot can indicate if a linear relationship between variables is plausible,

– after computation, the plot supplemented by the fitted regression line can tell which of the data points (units) are very close to the line, and which deviate considerably.

Moreover, a scatter plot can show many data set properties that affect the quality of the estimated coefficients and consequently the quality of the conclusions. The analysis may, as an example, strongly depend on some data points, which are particularly influential in the sense that the result will be quite different without these points.
Or, the data points can form clusters which, when considered separately, would show no significant linear relationship between variables. Thus, it is strongly advised to use the graphs as an aid and a presentation tool whenever a regression function is fitted.

The regression line fitted by the computational procedure should be used with caution. In addition to simple graphical procedures described above, there are several diagnostic methods which can be utilised to check if the assumptions of the regression model are met and whether the obtained regression equation can be used to describe a biological process. Weinsberg (1985) describes several techniques designed to find problems with the assumptions and influential data points.

The general rules for simple linear regression can also be applied with several extensions to more complex situations. The most important generalisations are:
– nonlinear regression, used when the relation between X and Y cannot be assumed to be linear,
– multiple (linear or nonlinear) regression, used when one wants to study the influence of several predictor variables on the response,
– multivariate regression, used in case of more than one response variable.

The regression equations in each of these three cases are straightforward generalisations of the linear equation. The fitting method is in most cases the same, based on the least-squares algorithm, and the conclusions about parameters are very similar. However, the simple scatter-plots cannot be used for critical assessment for a multiple or multivariate regression because a (2-dimensional) plot of the response variable against one predictor variable may be masked by a second predictor variable. This makes the more advanced diagnostic tools like the analysis of deviations or partial leverage plots more relevant.

Due to its simplicity, regression analysis is broadly used in all types of experimental studies. Unfortunately, it is also misused in several manners (see, e.g., Gomez and Gomez 1984, p. 416). Let us mention just the most common cases.
Firstly, the user must realise, that a necessary condition for regression analysis is some variability of the observations, both in X and Y. This variability must be caused by well understood or controlled sources (factors) if the regression equation is to be interpreted in a sound way.
Secondly, the fitted equation can be considered as valid only within the range of observed values of the variables; generalisations outside of this range are not justified.
Thirdly, in designed (replicated) experiments, the regression equation should be fitted to treatment (variety) means instead of plot observations in order to remove experimental error from consideration and because the interpretation may be difficult/wrong if the regression is calculated across several levels of variations.
Finally it should be noted that a significant regression coefficient (*b* significantly different from zero) does not prove that the predictor variable causes the variation found in the response variable unless the predictor variable is controlled by the investigator.

3.2.4. Generalised linear mixed models (GLMM)

Analyses of linear regression and analyses of variance, as described above, rely on models that express the response variable (e.g., yield) as a sum of:

- the so-called linear predictor, which is a linear function of parameters (that are fixed but unknown, such as regression coefficients) and random variables (such as block or sub-block effects in a mixed model of results of replicated experiments)
- the residuals, which are assumed to have a normal distribution.

Such a formulation implies that the expectation of the observed variable itself is a linear function of the parameters and variables included in the predictor. Although linearity and normality are often acceptable approximations for many continuous variables, and many significance testing methods are quite robust against violation of these assumptions, there are situations in which it is better to do the analysis using a more general model. The formulation of the generalised linear mixed model (GLMM), as described e.g. by Engel and Keen (1994), allows for this, because it assumes that:

- the expectation of the response variable is related to the linear predictor through the so-called link function (e.g., logarithm),
- the residual variability follows one of the distributions belonging to the exponential family, e.g. a binomial, Poisson or gamma distribution.

Initially, the classical linear model of observations, involving only fixed effects, was extended to a generalised linear model (GLM) (see e.g. McCullagh and Nelder, 1989). After realising that inclusion of random effects in GLM can be equally helpful as in linear models, the generalised linear mixed model (GLMM) was described (see Breslow and Clayton, 1993; Engel and Keen, 1994). Appropriate statistical procedures of estimation and hypotheses testing were developed, and GLMMs can now be fitted and analysed using statistical systems such as SAS or Genstat.

Several examples of GLMM applications to biological problems can be found in literature. One of them is the analysis of disease incidence data described by Piepho (1999). The author considers a situation, in which an experiment is designed with three replications to compare the effect of six treatments against downy mildew of grape. In each plot, five randomly chosen shoots from each of three vines were scored for mildew by counting m, the number of leaves with at least one mildew lesion and the total number of leaves per shoot, n. Two ways of modelling the data are considered:

a) a linear mixed model for the observed disease incidence m/n,

b) a generalised linear mixed model, in which the logit function of unknown probability of disease incidence, $log[\pi/(1-\pi)]$, is assumed to depend linearly on the fixed treatment and block effects, random effects of plots, random effects of vines within plots, and random errors caused by sampling of shoots within vines.

Under b), two sub models are discussed, which differ in the definition of "shoots within vines" effect. One of them involves the so-called over dispersion parameter, which here describes the extent to which the variances on the recorded values exceed those expected in the binomial distribution. The other sub model assumes a random effect of each shoot.

According to the author's final remarks, the analysis using GLMM is not much different in interpretation from the one using a linear model. In order to set up the analysis, a basic knowledge of similar rules is necessary. The advantage is that the parameters of GLMM may have a better interpretation; a disadvantage is that some statistical tests are valid only asymptotically (for large samples).

A similar problem is considered by Madden (2002), who gives some general rules on superiority of different GLMMs in the situation of an experiment conducted over five years to study the effects of different fungicide treatments on the control of Phomosis leaf blight of strawberry. The recorded variable was the number of diseased leaves in a sample of 15 leaves, representing a given plot and treatment. In the formulation of GLMM the logit function was used.

Another example of an interesting application of GLMM is given by Candy (2000). The author describes a study of incidence of some insects in tree leaves. The experiment consists of a multi-level sampling of plots within compartments of the plantation (trees within plots, branches within trees and shoots within branches), to count the number of leaves on the shoot occupied by insects. As the total number of leaves per shoot is very large and counting them is impractical, the response variable here is assumed to have a Poisson distribution. The logarithm of the expected number of affected leaves per shoot is modelled by a linear function of fixed compartment effects and random plot effects. Apart from the estimation of model parameters, the analysis described in the paper is meant to give hints on a better design of the experiment, in particular about an optimal relation between the number of sampled plots and sampled trees within plots.

Finally, GLMM can be applied to predict weed intensity in the field based on soil properties and counts of weeds observed over years, in the context of development of site-specific farming techniques, described by Christensen and Waagepetersen (2002). Here, the model is a spatial one. A GLMM with the Poisson distribution and the log link function is used to account for a non-normal distribution of the response variable.

3.3. Multi-environment trials (MET)

Multi-environment data originates from replicated experiments carried out in several years, at a number of sites, or in different environments defined by e.g. agricultural practice. Although in each case the observations are classified by environments, treatments and replications, the required analysis may be different for different meanings of the word "environment". Usually, full analysis of MET data with estimation and significance testing is completed for traits that are continuous and normally distributed, such as yield. Linear mixed models provide the most general analysis framework for such traits (Searle, Casella and McCulloch, 1992; Denis et al. 1997). Utilisation of linear models with only fixed effects may not be satisfactory due to the random nature of environmental sources of variation.

Most of the MET data are collected to study behaviour of plant genotypes (varieties, lines) in different environments. The analysis of such data can be done using two different approaches:

(a) a two-stage analysis, in which the data from all environments (experiments) are first analysed separately, and the estimated mean values are collected for the second stage devised to answer questions about the treatment-environment interaction,
(b) a one-stage analysis, in which plot-level data is modelled and analysed to give answers about the main effects and interaction.
Appropriate instructions for (a) are given e.g. by Patterson (1997). The methods for type (b) are described by Smith et al. (2001) and Caliński et al. (2005). The estimation method used extensively in mixed models for MET data is the REML algorithm (Patterson, Thompson 1971). The advantage of the recently developed approach (b) over the more traditional one (a) is that all observations are analysed within one model in order to estimate the parameters of interest and to test the corresponding hypotheses. A disadvantage of the one stage analysis (b) is that it is computationally more intensive and that it may require special algorithms.

Independently of the actual method of analysis, the MET data are collected to give answers about the variety x environment interaction. This interaction is defined as a differential response of genotypes to conditions in different environments. The presence of an interaction of a particular genotype with environment can also be understood as a situation, in which the genotype's reaction to the environments is different from the mean reaction of a set of reference (standard) genotypes or the mean reaction of all genotypes included in the trials. This definition implies a practical requirement for the trials: the set of genotypes used in different trials should be as uniform as possible. Although the REML algorithm can treat data even with a very incomplete (non-orthogonal) structure, caution should be taken when the variety x environment table contains many missing values.

The estimated genotype x environment interaction parameters, if statistically significant, can be submitted to some additional analyses aimed at explaining the nature of interaction. Very often the joint regression analysis (JRA, Eberhart and Russel 1966; Shukla 1972) is used for that purpose. JRA tries to explain genotype x environment interaction by an environmental index, usually calculated from the mean values for the environments. However, it should be noticed that a good determination of interaction variability by regression on such a simple index is seldom satisfactory. Therefore, more complicated indices are formed; for this task, the knowledge of weather and soil characteristics of the trial locations is extremely helpful.

Finally, one should acknowledge the importance of several explorative or analytic methods in the analysis of MET data. An initial component analysis of genotype x environment interaction deviations, and its graphical representation in the form of a bi-plot (Kempton 1984), can be very helpful in discovering advantage or disadvantage of genotypes for particular environments. Experience with using other geometrical methods is reported by Westcott (1986).

3.4. Analysis of data recorded on a discrete scale

Several traits important for the behaviour and quality of genotypes are expressed on a discrete scale, usually from 1 to 5 or from 1 to 9. As an example we can take disease severity, which is visually assessed as percentage of the area (of plant or leaf) affected and recorded as a number from 1 to 9, according to a rating scale (see chapter Disease assessment page D 11). Another example is disease incidence measured as the percentage of the affected plants in the plot (see chapter Disease assessment page D 10). Statistical analysis of such data is not always

straightforward, because the measurement scale may cause problems with the assumption of normality underlying several procedures. Therefore, some researchers do not carry out formal significance tests for disease or quality traits. This practice is acceptable, because in most of the experiments the trait of primary interest is the yield, and ranking of the treatments for additional traits can provide sufficient basis for the breeder's decision. However, if the statistical analysis of the discreet scale traits is interesting, the following solutions are possible.

In an analysis of replicated experiments, assuming that the unit of measurement is plant or leaf as it is the case for disease severity, there are several measurements per plot, which can be averaged to provide the plot observation. Such means may be assumed to behave as a variable measured on a continuous scale, and can be subjected to analysis of variance, possibly after a transformation. If the unit of measurement is a plot, there is only one observation, and for such data, analysis of variance should not be used. A possible solution is to create replicated observations by sampling within experimental plots (Gomez and Gomez, 1984, p. 532).

In a regression analysis, estimation and testing of regression coefficients should not be done on plot data, but on treatment (variety) means estimated from the analysis of variance model. Distribution of such values can be approximately normal.

In multi-environment data analysis, mean values for treatments over replications within environments can also be considered as approximately normally distributed. Therefore, the analysis of main effects of treatments can be completed using the analysis of variance if the significance of these effects is tested by comparison with the treatment x environment interaction.

For some of the traits there is a possibility to keep the observations in the form of counts (of units affected out of total number of units investigated). Very often, it is found that such observations have a binomial or Poisson distribution and can be modelled by GLMM as described in Section 3.5.

Aknowledgement:
The authors wish to thank John Law, Haidee Philpott, Andy Horwell and Jane Garner for their advice on the English style and grammar of this chapter.

4. References

Bliss CI, 1934, The methods of probits, Science 79, pp. 38-39.

Breslow NE, Clayton DG (1993). Approximate inference in generalised linear mixed models. Journal of the American Statistical Association 88: 9-25.

Caliński T, Czajka S, Kaczmarek Z, Krajewski P and Pilarczyk W (2005). Analyzing multienvironment variety trials using randomisation-derived mixed models. Biometrics 61: 448-455.

Caliński T, Kageyama S (2000). Block Designs: A Randomisation Approach, Volume I: Analysis, Lecture Notes in Statistics, Volume 150, Springer, New York 2000.

Candy SG (2000). The application of generalized linear mixed models to multi-level sampling for insect population monitoring. Environmental and Ecological Statistics 7: 217-238.

Christensen OF, Waagepetersen R (2002). Bayesian prediction of spatial count data using generalized linear mixed models. Biometrics 58: 280-286.

Clarke FR, Baker RJ & DePauw RM (1998): Interplot Interference Distorts Yield Estimate in Spring Wheat. Crop Sci. 38, 62-66.

Cochran WG and Cox GM (1966): Experimental Designs. New York. Wiley

David O and Kempton RA (1996): Designs for Interference. Biometrics 52, 597-606.

Denis J-B, Piepho H-P and van Eeuwijk FA (1997). Modelling expectation and variance for genotype by environment data. Heredity 79: 162-171.

Eberhart SA, Russel WA (1966). Stability parameters for comparing varieties. Crop Science 6: 36-40.

Engel B and Keen A (1994). A simple approach for the analysis of generalized linear mixed models. Statistica Neerlandica 48: 1-22.

Finney DJ (1946). Standard errors of yields adjusted for regression on an independent measurement. Biometrics 2: 53-55.

Finney DJ (1953). An introduction to statistical science in agriculture. Ejnar Munksgaard, Copenhagen.

Goldringer I, Brabant P & Kempton RA (1994): Adjustment for Competition between Genotypes in Single-Row-Plot Trials of Winter Wheat (*Triticum aestivum*). Plant Breeding 112, 294-300.

Gomez and Gomez (1984). Statistical procedures for agricultural research. John Wiley and Sons, New York.

John JA and Williams ER (1995): Cyclic and computer generated designs. 2nd ed. Chapman & Hall

Kempton RA (1984). The use of biplots in interpreting variety by environment interactions. Journal of Agricultural Science, Cambridge 103: 123-135.

Kempton RA and Fox PN (1997): Statistical Methods for Plant Variety Evaluation. Chapman & Hall, London.

Little TM and Hills FJ (1978). Agricultural experimentation. John Wiley and Sons.

Madden L.V. (2002). Evaluation of generalized linear mixed models for analysing disease incidence data obtained in designed experiements. Plant Disease 86: 316-325.

McCullagh PM and Nelder JA (1989). Generalized linear models. Chapman and Hall, London.

Mead R (1988): The design of experiments: Statistical principles for practical applications. Cambridge university press.

Mead R, Curnow RN and Hasted AM (1983). Statistical methods in agriculture and experimental biology. Chapman and Hall, London.

Patterson HD (1997). Analysis of series of variety trials. *In* R.A. Kempton & P.N. Fox, ed. *Statistical methods for plant variety evaluation*, p. 139-161. London, Chapman & Hall.

Patterson HD (1997). Analysis of series of variety trials. In Statistical Methods for Plant Variety Evaluation (R.A. Kempton and P.N. Fox, eds.), Chapman & Hall, London, 19-25.

Patterson HD and Thompson R (1971). Recovery of inter-block information when block sizes are unequal. Biometrika 58: 545-554.

Patterson HD and Willams ER (1976): A new class of resolvable incomplete block designs. Biometrika 63, 83-92

Patterson HD, Williams ER & Hunter EA (1978): Block designs for variety trails. Journal agric. Sci, Camb. 90, 395-400.

Payne RW (editor) (2006) The Guide to GenStat® Release 9 - Part 2: Statistics, by the GenStat. Published by VSN International, ISBN 1-904375-26-X.

Petersen RG (1994): Agricultural field experiments. Decker, New York

Piepho HP (1999). Analysing disease incidence data from designed experiments by generalized linear models. Plant Pathology 48: 668-674.

Piepho HP, Büchse A & Emrich K (2003): A hitchhiker's guide to the mixed model analysis of randomised experiments. Journal of Agronomy and Crop Science, 189, 310-322.

SAS Institute (2006) SAS Online Doc® 9.1.3. SAS Institute Inc. Cary, NC., USA (http://support.sas.com/onlinedoc/913/docMainpage.jsp)

Schabenberger O and Pierce FJ (2002): Contemporary statistical models for the plant and soil sciences. CRC Press LLC

Searle SR, Casella G and McCulloch CE (1992). Variance components. John Wiley & Sons, New York.

Shapiro SS, Wilk MB (1965). An analysis of variance test for normality (complete samples). Biometrika 52: 591-611.

Shukla GK (1972). Some statistical aspects of partitioning genotype x environment components of variability. Heredity 29: 237-245.

Smith A, Cullis B, Thompson R (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57: 1138-1147.

Talbot M, Milner AD, Nutkins MAE & Law JR (1995): Effect of interference between plots on yield performance in crop variety trials. J. agric. Sci., Camb. 124, 335-342.

Tukey JW (1949). Comparing individual means in the analysis of variance. Biometrics 5: 99-114.

Weinsberg (1985). Applied linear regression. John Wiley and Sons, New York.

Westcott B (1986). Some methods of analysing genotype-environment interaction. Heredity 56: 243-253.