

David Makowski
François Piraux
François Brun

From Experimental Network to Meta-analysis

Methods and Applications with R
for Agronomic and Environmental Sciences

éditions
Quæ

Éditions Cirad, Ifremer, Inra, Irstea
www.quae.com

 Springer

From Experimental Network to Meta-analysis

David Makowski • François Piraux
François Brun

From Experimental Network to Meta-analysis

Methods and Applications with R for
Agronomic and Environmental Sciences

 Springer

David Makowski
INRA, CIRED, CIRAD, AgroParisTech
University Paris-Saclay
Thiverval-Grignon, Nogent-sur-Marne
France

François Piraux
Arvalis Institut du végétal
Boigneville, France

François Brun
ACTA
Castanet Tolosan, France

ISBN 978-94-024-1695-4 ISBN 978-94-024-1696-1 (eBook)
<https://doi.org/10.1007/978-94-024-1696-1>

Éditions Quæ, R10,78026 Versailles cedex, France

© Éditions Quæ 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature B.V.
The registered company address is: Van Godewijkstraat 30, 3311 GX Dordrecht, The Netherlands

Acknowledgments

The authors are grateful to their home institutions:

- INRA, French National Institute for Agricultural Research (<http://www.inra.fr/en>)
- ARVALIS, the French arable crops R&D institute (<https://www.english.arvalisinstitutduvegetal.fr/index.html>)
- ACTA, head of the network of French Technical Institutes for Agriculture (<http://www.acta.asso.fr/en.html>)

The authors would like to thank Jean Boiffin who has supported us since the beginning of this project. This edition was partly funded by:

- Network Data Science and Modeling for Agriculture (www.modelia.org), funded by a grant from the French Ministry for Agriculture and Fisheries
- #DigitAg – Digital Agriculture Convergence Lab (<https://www.hdigitag.fr/en>), funded by a grant of French National Research Agency (ANR-16-CONV-0004)

Contents

1	Introduction and Examples	1
1.1	Objectives of Experimental Network Analysis and Meta-analysis	1
1.2	Data	3
1.3	Analysis	4
1.4	Principal Steps	5
1.5	Objective of This Book	6
1.6	A Simple Example of a Mixed Model	6
	1.6.1 Definition	6
	1.6.2 Data	7
	1.6.3 Definition of the Model	7
	1.6.4 Estimation	9
	1.6.5 Comparison with a Fixed-Effect Model	11
	References	14
 Part I Analysis of Experimental Networks		
2	Basic Notions	17
2.1	Experimentation	17
2.2	Experimental Networks	17
	2.2.1 Definition	17
	2.2.2 Example	18
	2.2.3 The Notion of “Environment”	18
	2.2.4 Objectives of an Experimental Network	20
2.3	Notion of a “Population of Environments”	20
2.4	Notion of “Interaction”	21
	References	23

3	Analysis of a Network of Randomized Complete Block Design Experiments with One Factor	25
3.1	Objective of the Chapter	25
3.2	Example “Wheat”	26
3.3	Modeling	27
3.3.1	Model with Experiment as a Random Effect	27
3.3.2	Model with Experiment as a Fixed Effect	28
3.3.3	Example	29
3.3.4	Should Experiment Effect be Considered Fixed or Random?	32
3.4	Evaluation of the Model	33
3.4.1	Normality	33
3.4.2	Homoscedasticity	34
3.4.3	Independence	34
3.4.4	Suspect Data	35
3.5	Comparisons of Means	35
3.5.1	Hypothesis Tests: Tests of Equality	35
3.5.2	Confidence Intervals	37
3.5.3	Hypothesis Tests: Tests of Equivalence	38
3.5.4	Example	40
3.6	Analysis of the “Wheat” Example and R Script	44
	References	57
4	Advanced Methods for Network Analysis	59
4.1	Analysis of Mean Data	59
4.1.1	Step 1: Analysis of Individual Experiments to Estimate Treatment Means	60
4.1.2	Step 2: Analysis of Mean Data	60
4.1.3	Example	61
4.1.4	A Variant: Analysis of Mean Data with a Fixed-Effects Model	65
4.1.5	Estimation of Variance for the Treatment \times Experiment Interaction	65
4.1.6	R Script	67
4.2	Experiments with Heterogeneous Variances	69
4.2.1	Introduction	69
4.2.2	Example “Wheat”	70
4.2.3	Going Further	73
4.3	Missing Data	77
4.3.1	Origin of Missing Data	77
4.3.2	Adjusted Means	78
4.4	The Factors “Site” and “Year”	84
4.4.1	Objective	84
4.4.2	The “Wheat_Multi” Example	84

4.4.3 Model for the Analysis of Mean Data 85

4.4.4 Estimation of the Variance of the Treatment
 × Year × Site Interaction 86

4.4.5 Variance of the Difference Between Two
 Treatments 87

4.4.6 Analysis of the “Wheat_Multi” Example,
 with R Script 87

References 94

5 Planning an Experimental Network 95

5.1 Objective 95

5.2 Comparison of Two Treatments 96

5.2.1 Case of a Multisite Network 96

5.2.2 Case of a Multisite and Multiyear Network 98

5.3 Other Contrasts 100

5.3.1 Comparison to the Mean of Several Controls 100

5.3.2 Comparison with the General Mean 101

References 102

Part II Meta-analysis

6 Basic Concepts in Meta-analysis 105

6.1 Definition, Origin and Principal Steps of Meta-analysis 105

6.2 Estimation of Mean Effect Size 107

6.2.1 Objective 107

6.2.2 Systematic Search for Studies, Selection of References and
 Data Extraction 108

6.2.3 Estimation of Mean Effect Size with a Fixed-Effects Model 109

6.2.4 Estimation of the Mean Effect Size with a Random-Effects
 Model 113

6.3 Meta-regression 116

6.3.1 Objective 116

6.3.2 Example 118

6.3.3 Regression Models with and Without Random Effects . . . 118

6.3.4 Example (Continued) 121

6.4 Critical Analysis of the Results 123

References 125

7 Statistical Problems Specific to Meta-analysis 127

7.1 Definition of Effect Size 127

7.1.1 Correction of Biases Linked to the Use of Ratios 127

7.1.2 Difference Between Observation Means 128

7.1.3 Effect Sizes for Binary and Count Data 128

7.1.4 Correlation Coefficient 129

7.1.5 Effect Sizes Based on Variance 130

- 7.2 Generalized Linear Models for the Analysis of Discrete Data 131
 - 7.2.1 Binomial Logit Random-Effects Model for Analyzing the Effect of a Treatment 131
 - 7.2.2 Example 132
- 7.3 Non-linear Mixed Models 135
 - 7.3.1 Utility and Definition 135
 - 7.3.2 Example 136
- 7.4 Bayesian Models 138
 - 7.4.1 Definition 138
 - 7.4.2 Example: Meta-analysis with MCMCglmm 140
- References 145

- Appendix: R Resources to Implement Network Analysis and Meta-analysis Methods 147**

Chapter 1

Introduction and Examples



1.1 Objectives of Experimental Network Analysis and Meta-analysis

Many experiments are performed each year to analyze the effects of agricultural practices on production and the environment. An experiment consists of experimental treatments applied to experimental plots for one or several years. Depending on the objective of the experiment, these experimental treatments may correspond to different types of agricultural practices, such as pest control methods, fertilization practices, different sowing dates, crop rotations or varieties. In these experiments, one or several variables are measured for each treatment, with the aim of characterizing the plant cover (yield, product quality, leaf area, etc.) or environment (soil characteristics, greenhouse gas emissions, water pollution, impact on biodiversity, etc.) of the crop. These experiments generate valuable results, providing insight into the effects of agricultural practices on crops and environmental risks. They are performed by various organizations: research institutes, technical institutes, chambers of agriculture and companies.

The results of agronomic experiments are likely to differ strongly between sites and between years. The effects of agricultural practices depend on local factors linked, in particular, to climate and soil characteristics. Agronomic experiments are often performed in networks, to make it possible to evaluate the variability of the effects of agricultural practices. In such networks, the same set of experimental treatments is tested at several sites, in several years, to assess the variability of effects and to estimate mean effect size. Variety testing is a classic example of an experimental network. In this particular case, different varieties are cultivated at several sites, in several years, and their mean performances and variability are analyzed. In

an experimental network, observations may be obtained in very different conditions (soil type, climate, etc.). However, the global analysis of experimental networks testing common hypotheses is very useful, because such an analysis makes it possible:

- To consolidate the results of the experiment. The results of individual experiments are sometimes obtained with only a small number of replicates, and may therefore be imprecise. The analysis of several experiments increases the number of replicates used to estimate the quantities of interest.
- To generalize conclusions. As the results of agronomic experiments vary between sites and years, a synthesis of the results of several experiments can be used to assess the robustness of conclusions and the extent to which conclusions can be generalized to a set of environments.
- To understand variability. The analysis of experimental networks also provides information about the origin of the variability for certain results. If each individual experiment is characterized by one or several covariables (e.g. precipitation, temperature, soil type, pest incidence), then these covariables can be used to identify the sources of variability of the effects of the experimental treatments studied. Such an analysis sheds light on the effects of the treatments studied, making it possible to refine recommendations concerning agronomic practices according to the particular features of the local context.

The objective of meta-analysis is not dissimilar to that of experimental network analysis. Meta-analysis makes it possible to analyze the results of a set of experiments performed in different conditions but dealing with a common subject. Unlike experimental networks, the experiments considered in meta-analyses were not initially performed with the intention of synthesis and global analysis in mind. A meta-analysis synthesizes the findings of experiments conducted independently, generally by different institutions, and sometimes in very distant geographic regions. These experiments are combined *a posteriori*, often well after their completion.

Unlike experimental network analysis, meta-analysis is not just a statistical analysis of a set of experimental data obtained in different conditions; it includes a preliminary step, in which the objective is to recover the largest possible number of experiments performed to tackle a given question. This data collection step is very important and makes use of specific approaches that may be time-consuming to implement. For certain strategic subjects, many studies have been performed and published independently, by different teams, from diverse institutions. This is the case, for example, for studies aiming to compare “organic agriculture” and “conventional agriculture” systems, and for those studying the impact of no-till techniques or evaluating the performance of genetically modified crops or greenhouse gas emissions due to nitrogen fertilization. For such subjects, tens, or even hundreds of studies have been performed, over time periods of various lengths (generally 10–20 years) and the results have been published independently in reports and scientific journals. The recovery of published data for these subjects is not simple and requires the use of rigorous search and data extraction procedures.

1.2 Data

The objectives of experimental network analysis and meta-analysis may be similar, but the types of data used, and the collection and validation of these data may be very different.

Type of Data

In the case of experimental networks, the elementary data are generally available. It is therefore possible to use these data to analyze individual experiments or the entire network.

In meta-analyses, the elementary data are often unavailable. Generally, only aggregate data (e.g. means, frequencies and standard deviations) are available for each individual experiment.

Data Collection

In experimental networks, the organization of data collection is described in an experimental protocol that specifies the objective of the network, the treatments studied, the number of experiments to be set up, the choice of experimental sites, the conditions in which the experiments are performed (cropping practices, etc.), the number of replicates and the experimental design, the type of measurements to be performed, and so on.

In meta-analyses, the data are often extracted from scientific publications or reports obtained through a systematic literature review of articles and/or reports published on the study theme. The experimental protocols are thus described in each individual reference and may differ considerably between references. The data collection procedure requires a specific formalized approach, which may be very time-consuming. In some meta-analyses, the data are obtained by directly contacting the principal investigators of the experiments concerned and requesting the crude data.

Data Validation

Data validation is designed to ensure that the data collected really do address the question posed.

In experimental networks, all the experiments have the same experimental protocol and address the same objective. Data validation therefore consists of checking that the experiments were performed in accordance with the protocol. The investigators should identify all accidents and deviations from the protocol (heterogeneous emergence due to soil aggregation, late nitrogen fertilization because the plot was not accessible on the scheduled date, etc.) occurring during the experiment. At the end of the experiment, the investigator determines whether the experiment meets the objectives of the protocol.

In meta-analyses, the available information concerning the protocols of the various experiments is generally less detailed. The person performing the meta-analysis must use the available information to estimate, for each experiment, whether the published results correspond to the objective of the meta-analysis.

In practice, the difference between experimental network analysis and meta-analysis may not always be so clear-cut. Indeed, the results for individual experiments may not always be available in experimental networks, necessitating analyses of aggregate data i.e., mean results. In such cases, the data for each experiment are first analyzed and validated separately. The global analysis is then performed with the mean values obtained, in the same way as for a meta-analysis.

In other cases, the protocol may differ between some of the experiments included in a given experimental network. These differences may be linked to the way in which experiments are performed, or the nature of the treatments studied. For example, an experimental network aiming to evaluate the yield of wheat varieties may consist of experiments testing the same varieties in different cropping conditions, with different crop protection methods, for example. Some of the experiments may thus be less well protected against diseases than others. Consequently, the differences observed between varieties in some experiments may essentially reflect true differences in yield potential, whereas, in other experiments, they may also reflect the susceptibility of the varieties to disease. The experimental networks evaluating organic waste products (OWPs; manure, slurry, etc.) constitute a particularly interesting example. OWPs have different origins in different experiments. For example, “cattle manure” may differ considerably in composition between experiments, according to the way in which it is produced. The interpretation of results in these two situations is clearly more difficult than in ideal conditions in which the same experimental protocol is applied rigorously in all the experiments.

In some cases, some of the experimental treatments may be absent from some sites. This situation may affect both experimental networks and meta-analyses, and it complicates the comparison of study treatments. Indeed, when two treatments are tested in the same experiment, it is possible to compare their effects directly. By contrast, the comparison of two treatments, each present in only one of two different experiments, is necessarily indirect and involves comparisons with the other treatments present in both experiments. In practice, comparisons between two treatments may involve a combination of direct and indirect comparisons. However, before the data are analyzed, attention should be paid to the reasons for the absence of certain treatments in certain experiments. Indeed, if a treatment is absent from certain experiments because it is unsuitable for the experimental conditions (pedoclimatic conditions unsuitable for the treatment concerned, for example), then the absence of the treatment cannot be considered to be independent of its expected value. Interpretation of the results is particularly difficult in this situation.

1.3 Analysis

Analyses of experimental network data and meta-analyses can make use of many different statistical methods, some descriptive and others inference-based.

In experimental network analysis, the analysis of interactions between the study treatments and the experimental conditions (treatment \times experiment interactions) is of particular importance. This interaction leads to variability in the effect of the

treatments between experiments, according to soil type or climate, for example. In meta-analysis, the principal objective of the statistical analysis is to estimate the mean effect size, but the analysis of interactions is also important in some cases.

Mixed models are often used as the basic statistical model for syntheses of series of experimental results. Models of this type can take treatment \times experiment interactions into account and extend the validity of conclusions to a collection of environments not restricted to those actually observed. Mixed models play a key role in both experimental network analysis and meta-analysis. A simple but illustrative example is presented in Sect. 1.6 of this chapter. Descriptive methods, many of them graphic in nature, are also used in both types of analyses. Examples of these methods include forest plots for meta-analysis and biplots for the analysis of experimental networks. Graphs also play an important role in analyses of treatment * experiment interactions.

1.4 Principal Steps

Both analyses of experimental networks and meta-analyses require the implementation of methodologies with several steps. The same steps are followed in both cases, but with important differences between the two types of analysis for some of these steps.

- **Presentation of the hypotheses tested**

The hypotheses tested in analyses of experimental networks or in meta-analyses must be described as precisely as possible.

- **Data collection**

In experimental networks, the data are acquired to address a precise objective. The protocol is developed with the goal of performing a global analysis of the experiments, and the elementary data are often available. In meta-analyses, the experiments are often heterogeneous and the data are less accessible. It is often necessary to extract the data from articles and reports.

- **Data validation**

In experimental networks, data validation is based largely on the experiment monitoring notes taken during field visits, and on an examination of the data for individual experiments. In meta-analyses, data validation involves a profound critical examination of the articles and reports collected, and it may be necessary, in some cases, to contact the investigators directly to obtain more information.

- **Data analysis**

Various analytical methods may be used, but mixed models are probably the most widespread, for both analyses of experimental networks and for meta-analyses.

- **Validation of the analysis**

This step involves validation of the hypotheses associated with the model, such as the normality of random effects, for example. In meta-analyses, the conclusions may be affected by publication bias if only certain results are

published, such as those showing a significant effect of the treatments studied. The validation of meta-analysis therefore involves an analysis of the risk of publication bias.

- **Communication of the results**

In any communication of the results, the way in which the steps listed here were performed should be described, and the results should be presented in a transparent manner, together with the uncertainty associated with them.

1.5 Objective of This Book

Our aim here is to present and illustrate the principal statistical methods that can be used in the quantitative synthesis of data from experimental networks and scientific publications. Each chapter presents one or more methods, illustrated by practical examples with the R software, which is freely available from the following site: <https://cran.r-project.org/>. The data and R codes are provided and discussed, to facilitate their adaptation to other practical situations. They can be used in the R KenSyn package (<https://cran.r-project.org/web/packages/KenSyn/index.html>) accompanying this book.

In the next part of this chapter, we present a first statistical model for the analysis of simple experimental networks and standard meta-analyses. The first part of this book (Chaps. 2, 3, 4 and 5) describes several extensions of this model and show how they can be used to analyze different types of experimental networks. The second part of the book (Chaps. 6 and 7) presents in detail the objectives and principal steps of meta-analysis. Finally, the appendix presents the principal R packages that can be used for statistical analysis. Each chapter can be read independently of the others.

1.6 A Simple Example of a Mixed Model

1.6.1 Definition

Random-effects models are a particular type of statistical model often used to take several sources of variability into account (Laird and Ware 1982). These models include a large number of variants generalizing classical linear and non-linear regression models. Some of these variants may be complex and include both random and fixed effects. Such models are often referred to as *mixed models*.

In this chapter, we present a very simple model with Gaussian random effects, describing two sources of variability in the data: intraindividual and interindividual variability. Each source of variability can be described by a random effect. This model is often used in situations in which several measurements are collected for each individual in the sample. Depending on the context, an “individual” may be an animal, a person, or a plant, for example. In the example presented here, an individual corresponds to a particular experimental study in which several

measurements have been collected. Despite its simplicity, this model can be used to address several practical questions, including:

- Estimation of the mean value of the variable studied,
- Description of the variability of this variable between studies and within each study,
- Obtaining “local” estimates (specific to the individual study concerned).

This model is often used for the analysis of experimental networks. It is also an indispensable tool for meta-analyses (Mengersen et al. 2013). Here, we illustrate its utility through an example in which the objective is to estimate the mean yield of a crop in an agricultural region.

1.6.2 Data

The dataset consists of 45 measurements of wheat yield obtained at 15 experimental sites (for a given year) in an agricultural region. In this dataset, a study corresponds to a site-year comprising two to four plots. The yield of the crop (expressed in t ha^{-1}) was measured for each plot, and we therefore have two to four yield measurements per study (Fig. 1.1).

In this example, the population corresponds to the set of wheat plots in the region considered, for the year studied (Fig. 1.2). The sample is defined as the 15 sites and 45 yield measurements. In the next section, we will show how this sample can be used:

- To estimate expected yield for the population (i.e. “mean” yield for the region)
- To estimate yield variability between sites and to compare this variability with intra-site variability
- To estimate yield at each of the 15 sites.

1.6.3 Definition of the Model

The model was defined by Laird and Ware (1982), as follows:

$$y_{ij} = \mu + b_i + \varepsilon_{ij} \quad (1.1)$$

where y_{ij} is the yield measured at site i for plot j , $i = 1, \dots, 15$, $j = 1, \dots, n_i$ ($2 \leq n_i \leq 4$), μ is the expected yield of the population, b_i is the random effect of “site”, $\mu + b_i$ is the mean yield of site i , ε_{ij} is the intra-site residual describing the difference between y_{ij} and $\mu + b_i$. The quantities b_i and ε_{ij} are defined as independent Gaussian random variables, such that $b_i \sim N(0, \sigma_b^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. The random variable b_i is the difference between the yield at site i and the general mean yield for the whole population.

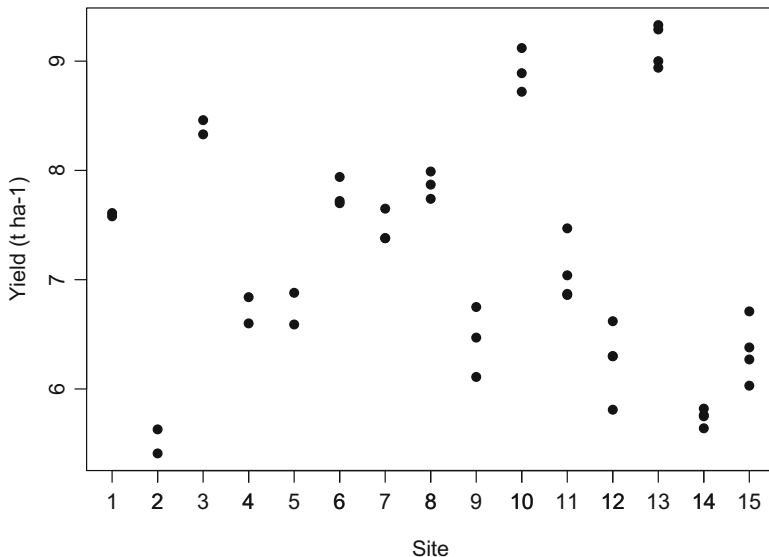


Fig. 1.1 Wheat yield measurements obtained at 15 sites in a given year. Two to four measurements were made on different plots, at each site

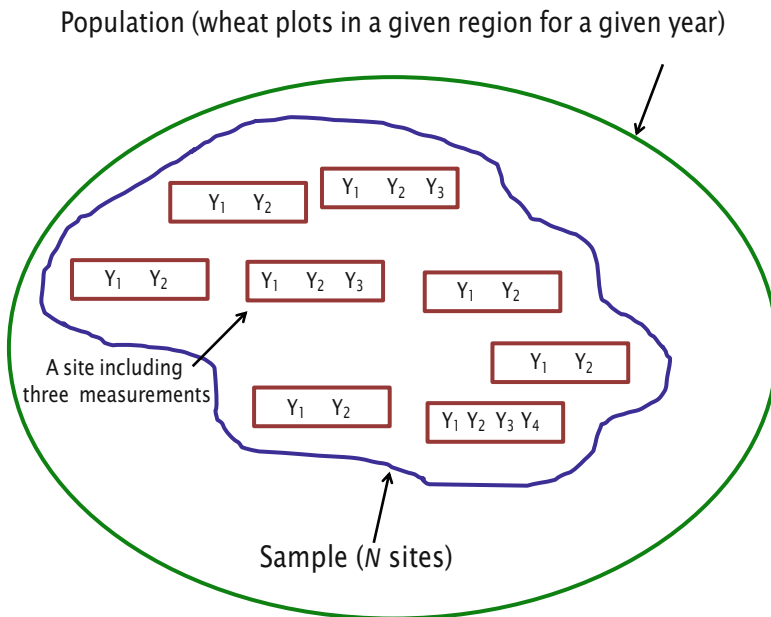


Fig. 1.2 Diagram of the population and the sample

This model includes several parameters of unknown value:

- Population mean yield (μ),
- Intersite variance (σ_b^2),
- Intra-site variance (σ_ε^2).

According to this model, two yield values obtained at the same site are not independent. Their covariance is σ_b^2 and their correlation is $\frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2}$. The hypothesis of the non-independence of measurements is consistent with the structure of our dataset. Indeed, Fig. 1.1 shows that data collected at the same site show some similarities; they group together such that each site is associated with a range of values different from that at other sites. The fixed-effects model requires only two parameters to describe correlations between measurements collected at the same site: σ_b^2 and σ_ε^2 .

We show below how to estimate these parameters of the model (1.1) with R software.

1.6.4 Estimation

If we assume that σ_b^2 and σ_ε^2 are known, the most precise (minimal variance) unbiased estimator of μ is defined as follows:

$$\hat{\mu} = \frac{\sum_{i=1}^p w_i \bar{y}_i}{\sum_{i=1}^p w_i} \quad (1.2)$$

where \bar{y}_i is the mean of the yields measured at site i , p is the number of sites in the sample (here, $p = 15$), and w_i is a weight defined by $w_i = \frac{1}{\sigma_b^2 + \sigma_\varepsilon^2/n_i}$. The estimator defined by (1.2) is a weighted mean of the mean yields observed at each site \bar{y}_i , $i = 1, \dots, p$. The weight w_i increases with the number n_i of available measurements for site i . Interestingly, weights are equal when exactly the same number of measurements is available at each site. They tend to be similar if the intersite variance σ_b^2 largely outweighs the intra-site variance. In this case, the estimator (1.2) is equal (or nearly equal) to the simple mean of p mean yields observed at the p sites. It should also be noted that, when the intersite variance tends to zero, the estimator (1.2) is equal to the mean of the values of \bar{y}_i , weighted by the number of observations collected at the various sites.

The value of b_i , the difference between the true yield at site i and the general mean yield for the whole population, can be estimated as follows:

$$\hat{b}_i = E(b_i | \bar{y}_i) = \frac{n_i \sigma_b^2}{n_i \sigma_b^2 + \sigma_\epsilon^2} (\bar{y}_i - \mu) \quad (1.3)$$

This quantity corresponds to the expected value of b_i based on the mean value observed at site i . This estimator is often considered to be an *empirical Bayesian* estimator because it corresponds to a posterior expectation. It is the estimator with the smallest variance of all the unbiased estimators (the best linear unbiased predictor or BLUP). In absolute terms, the value calculated with (1.3) is lower than $\bar{y}_i - \mu$. The estimator (1.3) is, therefore, closer to zero than $\bar{y}_i - \mu$.

For the calculation of (1.2) and (1.3), we need to estimate the values of the variances σ_b^2 and σ_ϵ^2 from data. These two variances can be estimated by restricted maximum likelihood (REML) methods, with the `nlme` and `lme4` packages of R. We present some examples of R code below. Many other examples were presented by Pinheiro and Bates (2000) and Bates et al. (2015).

The R code below can be used to estimate the parameters of the model (1.1) with the `nlme` library:

```
#Reading of the external file including the data
TAB<-read.table("dataMod_1.txt", header=T)
#Loading of the nlme library
library(nlme)
#Adjustment of the model with the function lme
Mod<-lme(Rdt~1, random=~1|Site, data=TAB)
#Display of the results
summary(Mod)
```

The parameters of the model are estimated with the function `lme`. The results of this estimation, visualized with the function `summary`, are presented in Fig. 1.3. The estimated values of the standard deviations σ_b and σ_ϵ are 1.09 and 0.22, respectively. Variance is the square of standard deviation. The estimated value $\hat{\mu}$ of μ is 7.2 (Fig. 1.3). The standard deviation of $\hat{\mu}$ is 0.28 (Fig. 1.3). This standard deviation describes the uncertainty on estimates of $\hat{\mu}$. It should not be confused with σ_b , which describes the variability between sites.

The parameters of the model (1.1) can also be estimated with the `lmer` function of the `lme4` library, with the help of the following code:

```
#Loading of the lme4 library
library(lme4)
#Adjustment of the model with the lmer function
Mod_bis<-lmer(Rdt~1+(1|Site), data=TAB)
#Display of the results
summary(Mod_bis)
```

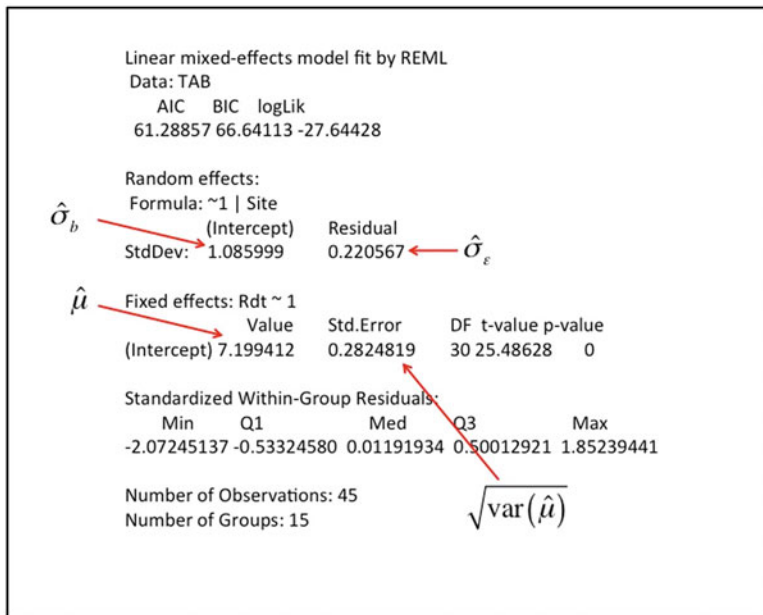


Fig. 1.3 Results obtained with the `lme` function of the `nlme` library. The symbol “^” indicates that the values reported are estimates rather than real measured values

The results are identical to those obtained with `lme` but are presented in a slightly different format (Fig. 1.4).

Yield estimates for each of the 15 sites are presented in Fig. 1.5. These estimates are equal to $\hat{\mu} + \hat{b}_i$, $i = 1, \dots, 15$, where $\hat{\mu}$ is calculated from Eq. (1.2) and \hat{b}_i is calculated from Eq. (1.3) by replacing σ_b^2 and σ_ϵ^2 with their estimated values. The estimated yield values for each site are obtained with the `predict` function of the `nlme` package.

1.6.5 Comparison with a Fixed-Effect Model

It is interesting to compare the results obtained with model (1.1) with those obtained without a random effect of “site”. This fixed-effect model is defined as follows:

$$y_{ij} = \mu + \epsilon_{ij} \tag{1.4}$$

In (1.4), the residues ϵ_{ij} are assumed to be independent, normally distributed, and of equal variance, $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. This model does not distinguish between intra-site and intersite variability and assumes that yield measurements obtained at the same

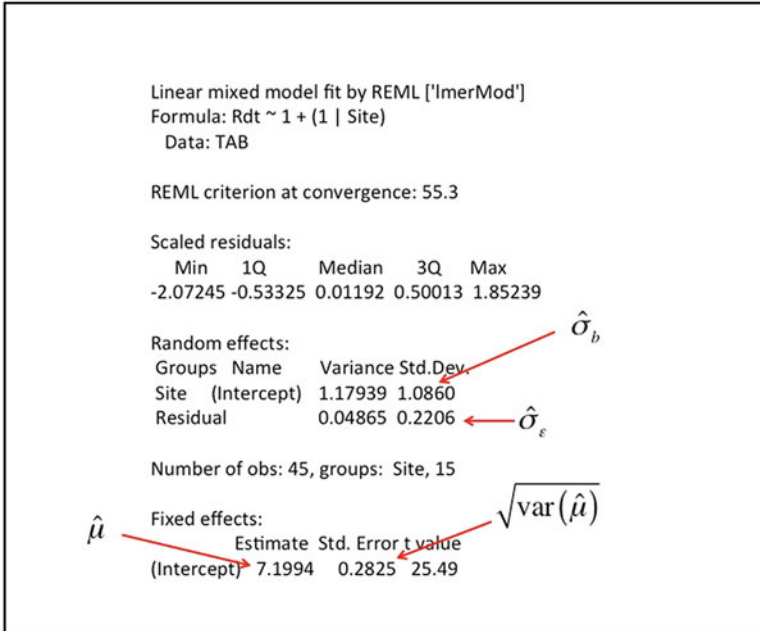


Fig. 1.4 Results obtained with the lmer function of the lme4 library. The symbol “^” indicates that the values reported are estimates rather than real measured values

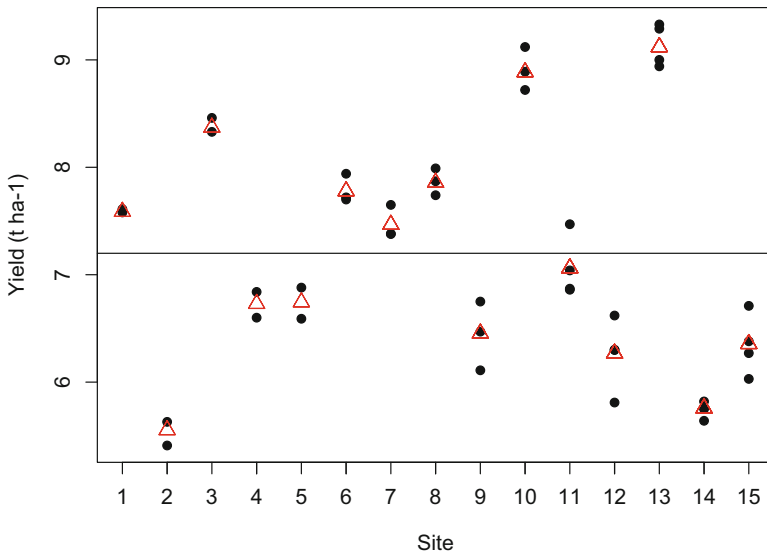


Fig. 1.5 Estimation of expected yield (horizontal line) in the population of sites and estimations of yields at individual sites (triangles). The black dots are data points

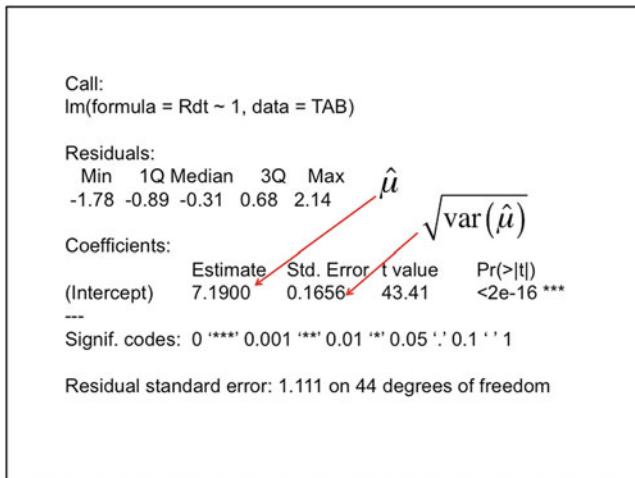


Fig. 1.6 Results obtained with the lm function

site are independent. This assumption is clearly unrealistic according to Fig. 1.5. Observations collected at the same site tend to be either higher or lower than the mean regional yield. If measurements at the same site were truly independent, they would be randomly distributed on either side of the regional mean.

The model (1.4) can be adjusted to the data with the assistance of the lm function of R:

```

#Adjustment of the model with the lm function
Mod0<-lm(Rdt~1,data=TAB)
#Display of the results
summary(Mod0)
    
```

With this model, the estimated value of mean regional yield is simply the mean of the 45 measurements. Here, this estimate is close to that obtained with model (1.1) (7.19 vs. 7.2), but the standard deviation of the estimator (which estimates the imprecision of the estimate) is lower for the lm function than for the lme function (0.17 vs. 0.28) (Fig. 1.6). Model (1.4) provides too optimistic a result. It overestimates the precision of the estimate, because it assumes that all the measurements are independent, whereas this is not the case in reality. The lack of realism of the hypothesis of residual independence is confirmed in Fig. 1.7. The residuals of model (1.4) are all positive at some sites and all negative at others. The residuals of model (1.4) are markedly larger than those of model (1.1) because they include two sources of variability (inter- and intra-site).

In this example, the principal advantage of a random-effects model (1.1) is that it provides more realistic information about the precision of the estimate.

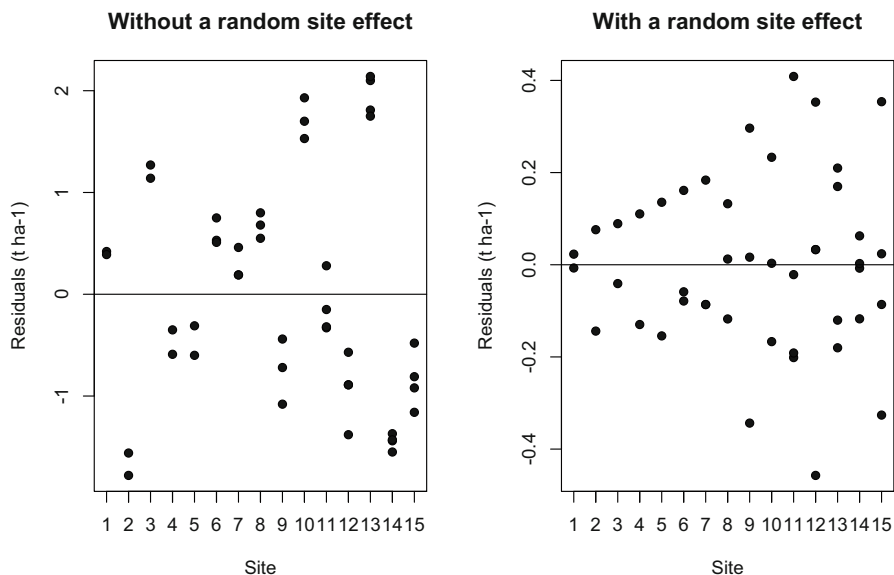


Fig. 1.7 Distribution of the residues of the model without a site effect (1.4) and of the model with a site effect (1.1)

In the next chapter, we show how the random-effects model presented here can be modified for the analysis of more complex experimental networks.

References

- Bates D, Maechler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67:1–48
- Laird NM, Ware JH (1982) Random-effects models for longitudinal data. *Biometrics* 38:963–974
- Mengersen K, Schmid CH, Jennions MD, Gurevitch J (2013) Statistical models and approaches to inference. In: Koricheva J, Gurevitch J, Mengersen K (eds) *Handbook of meta-analysis in ecology and evolution*. Princeton University Press, Princeton, pp 89–107
- Pinheiro JC, Bates DM (2000) *Mixed effects models in S and S-Plus*. Springer, New York

Part I

Analysis of Experimental Networks

This part deals with the planning and analysis of data from experimental networks.

Chapter 2 deals with the various basic notions that can help to explain the role of experimental networks and their analysis.

Chapter 3 presents networks of randomized complete block experiments with one factor. This experimental design is classically used in agronomy. This chapter introduces general notions relating to the analysis of experimental networks that users must master if they are to make use of the more complex methods.

Chapter 4 extends the analysis of experimental networks presented in Chap. 3, to take more complex situations into account. The first part of this chapter presents a two-step analysis of an experimental network, the method most widely used in practice, particularly if the individual experiments of the network are performed with different experimental designs. The second part of Chap. 4 presents the analysis of a network in which precision differs between experiments. The third part shows how missing data can be taken into account in network analyses. Finally, the fourth part of Chap. 4 explains how to take the factors “site” and “year” into account in a multisite, multiyear experimental network.

The planning of experimental networks involves deciding how many replicates there should be per experiment, and the number of experiments to be included in the network. However, it also concerns the way in which choices are made concerning the environments in which the experiments are performed. Chapter 5 deals with these questions and proposes a method for dimensioning a network.

Chapter 2

Basic Notions



2.1 Experimentation

In agricultural and environmental sciences, experiments aim to study the effect of factors on a quantity of interest, in controlled conditions, in a given environment.

Each experiment is the subject of formalized planning based on an experimental protocol that must include at least the following elements (Dagnelie 1981):

- Definition of the objective and conditions of the experiment,
- The factor(s) studied:
 - Definition
 - Choice of treatments for qualitative or quantitative factors (each different level of a factor is considered to be a treatment)
 - Implementation of a control or reference treatment
- Definition of the variable(s) of interest and of the secondary variables
- Definition of the experimental units: size and shape of the experimental units, borders, etc.
- The experimental design: type of design and number of replicates

In agricultural research, the conclusions of a study are mostly destined to be applied to the whole of a given region in the next cropping season or seasons. The results obtained in a particular experiment performed in a given pedoclimatic context must therefore be confirmed at other sites and in other years, in the framework of an experimental network covering diverse pedoclimatic conditions.

2.2 Experimental Networks

2.2.1 Definition

An experimental network is a set of experiments with the same experimental protocol, performed in a set of different environments.

Ideally, in an experimental network:

- The definition of factors and treatments is identical for all experiments,
- All the treatments should be present in all the experiments,
- The experimental design and the number of replicates should be identical for all experiments.

2.2.2 Example

Figure 2.1 provides an illustration of an experimental network.

2.2.3 The Notion of “Environment”

The notion of “environment” may depend on the experimental context.

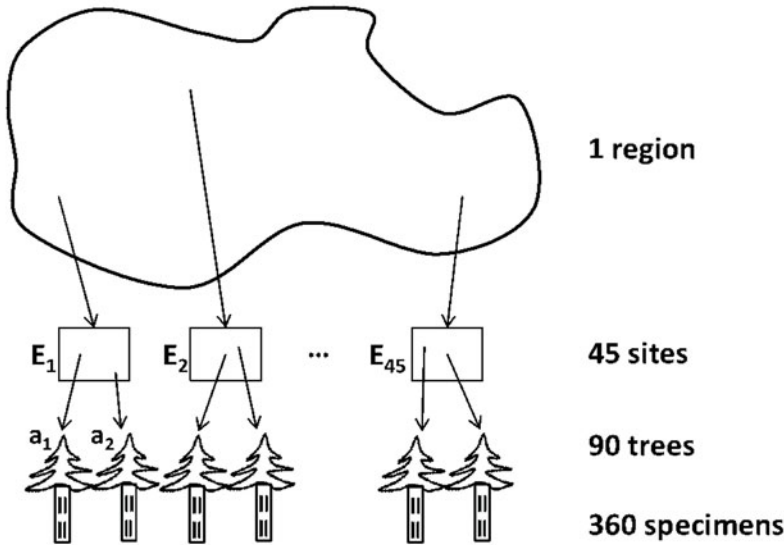
Thus, for example, in the case of evaluations of new wheat varieties, an environment is defined as the combination of a particular site and a particular climatic year. In this case, the experimental network can take into account the experimental variability linked to soil type, cropping history, or climate, for example.

Evaluations of anti-slug products are sometimes performed in growth chambers. In these conditions, there is no climatic variability between two experiments. However, if the slugs used for the experiments come from natural populations present on agricultural fields, two experiments performed on different dates may be conducted with different slug populations, potentially leading to different results for anti-slug product efficacy. In this case, the use of an experimental network makes it possible to take into account the experimental variability linked to the origin of the slugs.

In the case of experimental networks for studies of the fattening of farm animals, many factors are likely to vary between experiments:

- Even if performed in agricultural buildings, animal experiments are subjects to a certain degree of climatic variability,
- It is possible to ensure that the basic feed supplied to the animals has the same composition in different experiments and to check various criteria, including energy and protein contents. Nevertheless, despite such precautions, two rations of feed prepared from maize grown in two different climatic years can never be absolutely identical,

The following example is taken from Dagnelie (1998). It concerns a study of the technological quality of common spruce wood. A network of experiments at 45 sites representative of the region concerned was set up, with two trees selected and chopped down at each of these sites. For each of these trees, four specimens of 1 m in length, with a cross-sectional area of 5 cm x 5 cm, were obtained. Two of these specimens were taken from the lower part of the tree, and two were taken at mid-height. Various observations were made for each of the 360 specimens, including assessments of flexion resistance. The sampling scheme is illustrated below.



The 45 sites were considered to be representative of the region in which the study was performed. The two trees at each site were assumed to be representative of the site concerned, and the two specimens collected from each part of the tree were assumed to be representative of the part of the tree concerned. The various specimens, trees and sites observed therefore constituted a sample of the possible specimens, trees and sites. The questions tackled by this experimental protocol concerned the difference in wood quality between the upper and lower parts of the tree and the aim was to reach conclusions that could be generalized to all possible specimens, trees and sites within the study region.

Fig. 2.1 Example of a network of experiments with the same protocol

- It is possible to ensure that the animals used in two different experiments have the same genetic background (e.g. Landrace × Duroc Blanc pigs), but there may still be genetic variability between the animals used in different experiments, because the parents of the animals used are not necessarily the same.

2.2.4 *Objectives of an Experimental Network*

Experimental networks may be set up to address three types of objectives:

- Consolidation of results

The results of experiments are sometimes obtained with only a small number of replicates. These results may, therefore, be imprecise if based on a single experiment. The consideration of results from several experiments makes it possible to increase the number of replicates used to estimate the quantities of interest.

- Generalization of conclusions

In a network, experiments performed in different environments are subject to different experimental conditions. If the results of an experiment vary with the experimental conditions used, then a synthesis of results from several experiments can be used to analyze the robustness of the conclusions and to extend the validity of the results to a broader set of environments.

- Description of variability

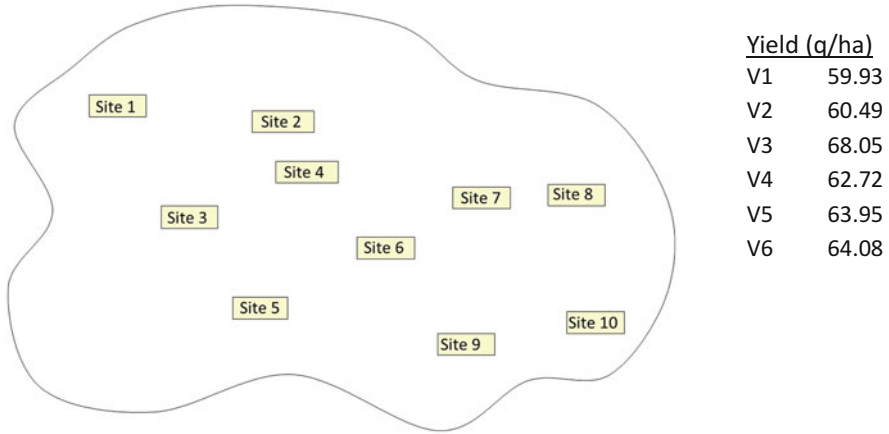
Experimental network analysis can also improve our understanding of the variability of certain results. If each individual experiment is characterized by one or several covariables (e.g. precipitation, temperature, soil type, pest incidence), then it may be possible to link these covariables to the effects of the experimental treatments studied. Such an analysis can improve our understanding of the effects of these treatments, making it possible to refine recommendations concerning agronomic practices according to the particular features of the local context.

2.3 Notion of a “Population of Environments”

A farmer wishing to make use of the results of experiments evaluating new varieties may be located in an area in which no experiments took place. Even for farmers located in the areas of experimentation, the varieties used will always be planted in a climatic year other than that in which testing took place.

The principal aim of experimental networks is to make it possible to generalize the conclusions of a study to a larger set of environments. The collection of possible environments is called the population of environments. In practice, it is not possible to perform an experiment in each of the environments of the population. A subset of environments must therefore be chosen in which to conduct the experiments. This subset constitutes a sample that must be as representative as possible of the population. It is generally recommended, where possible, to choose the environments at random, to ensure such representativeness (Dagnelie 1981). Figure 2.2 illustrates this notion of the generalization of the results of an experimental network to a population of environments.

Just as each individual experiment has its experimental protocol, so an experimental network should have, in principle, a planned and governed experimental



The left part of Figure 2.2 represents a zone of pea production in which a network of 10 experiments has been set-up, with each experiment conducted at a different site. The objective of the network is to evaluate the yields of six varieties, to make it possible to recommend the best variety to cultivate in any part of the production zone, i.e. at one of the sites of the experimental network or any other site in the pea production zone.

The right part of Figure 2.2 shows the mean yields of the six varieties estimated from the experimental network data. These results can be used to choose a variety for cultivation at any site in the production zone if they constitute an estimate of mean yield for the variety concerned **over the production zone**. For this to be the case, the experimental sites used must be representative of the production zone.

Fig. 2.2 Generalization of the results of an experimental network to a population of environments

protocol. In this case, in addition to the elements classically included in the protocol for a single experiment, it is also necessary to describe the objective of the network and to define the population of environments and the criteria used to select the environments to be observed in the network. It is also necessary to specify the number of environments making up the network.

2.4 Notion of “Interaction”

The treatment \times environment interaction reflects the differences between treatments as a function of the environment in which the experiment was performed.

The aim of an experimental network is generally to compare the studied treatments with each other. We are therefore interested in the difference between treatments, rather than the absolute values obtained. In the absence of an interaction between treatment and environment, the differences between treatments are independent of the environment. In these conditions, only one experiment is required, with a large enough number of replicates to take the experimental error into account, making it possible to draw firm conclusions about the differences between treatments at the scale of the population of environments.

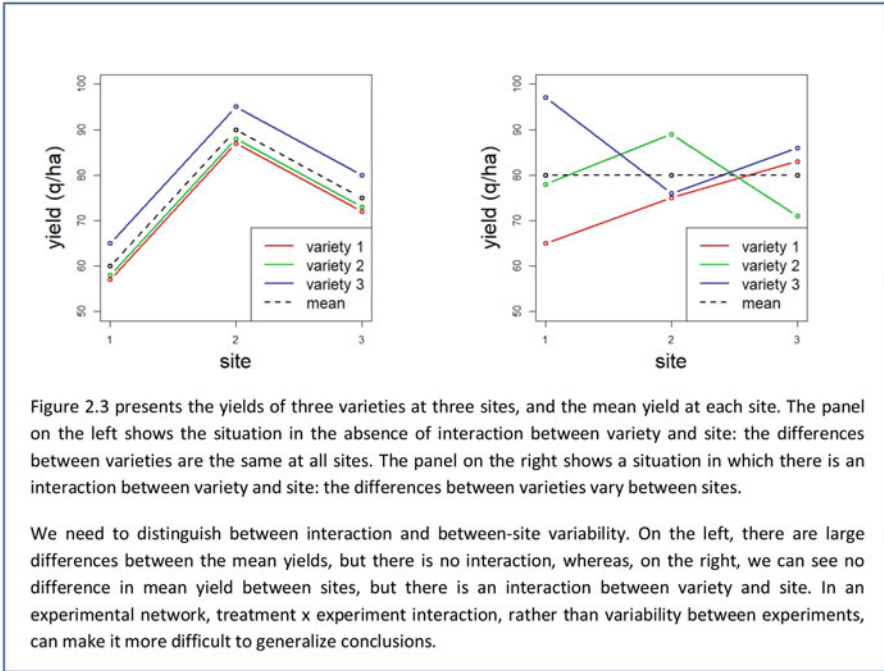


Fig. 2.3 Illustration of the notion of “interaction”

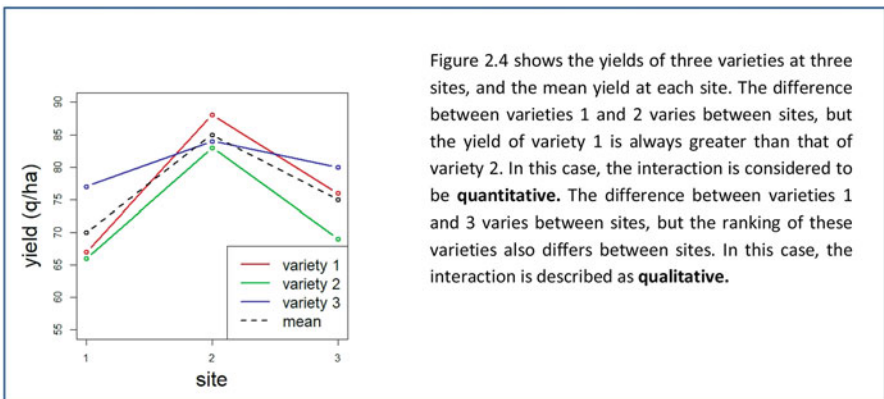


Fig. 2.4 Quantitative vs. qualitative interactions

In the presence of an interaction between treatment and environment, it is more difficult to draw firm conclusions about the differences between treatments at the scale of the population of environments. If the interaction is large, and if the number of environments included in the network is small, the observed results are likely to vary considerably depending on the sample of environments selected.

Figures 2.3 and 2.4 illustrate the notion of “interaction”.

References

- Dagnelie P (1981) Principes d'expérimentation. Presses Agronomiques, Gembloux, 182 p
- Dagnelie P (1998) Statistique Théorique et Appliquée. Tome 2: inférence statistique à une et à deux dimensions. De Boeck & Larcier, Paris et Bruxelles, 659 p

Chapter 3

Analysis of a Network of Randomized Complete Block Design Experiments with One Factor



3.1 Objective of the Chapter

This chapter presents in detail the principal methods for analyzing networks of experiments with a randomized complete block design including one factor (Dagnelie 1981). In networks of this type, each experiment has a randomized complete block design, within which the various experimental units are grouped together in homogeneous blocks. Each block contains as many experimental units as there are study treatments, and each treatment is tested in each block. Thus, in any given experiment, each treatment is present only once in each block. Treatments are distributed at random between the various experimental units, independently in each block. In a randomized complete block design, the uncertainty on the difference between the means of two treatments stems solely from the variability between experimental units within blocks, and the precision of the experiment can be increased by setting up the most homogeneous blocks possible (Cochran and Cox 1957). In agronomic experiments, two neighboring plots are often more similar than two more distant plots, and it is usual practice to group neighboring plots together within the same block (Cochran and Cox 1957; Zimmerman 1991; Kempton 1981). The randomized complete block design is a popular experimental design in agronomic experiments, and experimental networks based on this design are very common.

In this chapter, we consider a complete network including several experiments with a randomized complete block design, each with the same number of blocks. The expression “complete network” indicates that all the treatments are present in each experiment.

The second part of this chapter presents the results of an experimental network on wheat varieties. These data will be used to illustrate the various theoretical notions developed in this chapter.

Experimental networks can be analyzed by considering the factor “experiment” to have fixed or random effects. The two types of analysis are presented in the third part of this chapter.

The interpretation of data from a network of experiments is based on a statistical model. For the interpretation to be valid, the model used must best describe the phenomenon studied. The fourth part of this chapter presents several model validation tools.

The principal objective of an experimental network is generally to compare treatments with each other. The fifth part of this chapter presents different methods for comparing means between treatments.

Finally, the sixth part of this chapter presents a complete analysis of the “wheat” example and the R script used for this analysis.

3.2 Example “Wheat”

The “wheat” example concerns the evaluation of wheat varieties in field experiments including microplots. The grain yield of 10 wheat varieties was evaluated in five experiments at five different sites. The sites were chosen so as to be representative of the environmental variability observed in the study region. All the experiments were performed in the same year, and all the experiments had a randomized complete block design. Each experiment included three replicates corresponding to three blocks.

The dataset includes experiment number, variety name, block number in the experiment and observed yield (q/ha). Table 3.1 shows the first few lines of the dataset.

Table 3.1 Excerpt from the “wheat” dataset

Experiment	Variety	Block	Yield
1	V1	1	71.5
1	V2	1	62.7
1	V3	1	74.7
1	V4	1	72.4
1	V5	1	76.7
1	V6	1	72.5
1	V7	1	68.8
1	V8	1	79.3
1	V9	1	80.6
1	V10	1	73.1
1	V1	2	71.9
1	V2	2	71.8
1	V3	2	81.0
...

3.3 Modeling

Two types of model can be used to analyze data from experimental networks: models considering the effect of experiment to be random and models considering this effect to be fixed. We present the application of these two types of models to the analysis of individual data from a network, and illustrate the differences between them for the “wheat” example. Recommendations are made concerning the choice of model.

In this book, we have generally chosen to consider the factor “treatment” to be fixed, but the question as to whether the effects of this factor should be considered fixed or random should be raised for each study. In the case of networks for the evaluation of varieties, it is frequently considered to be a random factor. Interested readers may wish to consult Piepho et al. (2008) to find out more.

3.3.1 Model with Experiment as a Random Effect

The following mixed model can be used for the analysis of individual data from a network of experiments with a randomized block design including one factor (Dagnelie 1998):

$$y_{ijk} = \mu + t_i + E_j + tE_{ij} + B|E_{jk} + \varepsilon_{ijk} \quad (3.1)$$

where y_{ijk} is the observed value of the quantity of interest for treatment i ($i = 1, \dots, I$) in block k ($k = 1, \dots, K$) of experiment j ($j = 1, \dots, J$) and μ , t_i , E_j , tE_{ij} , $B|E_{jk}$, and ε_{ijk} are the general mean, the effect of treatment i , the effect of experiment j , the effect of the interaction between treatment i and experiment j , the effect of block k in experiment j , and the residual associated with y_{ijk} , respectively. The residual ε_{ijk} , sometimes known as the “residual error”, is the sum of two components: a measurement error and an intra-block “experimental unit” error.

The blocks are assumed to be representative of the experiment from which they are taken and their effects are considered random in the model (3.1).

The environments in which the experiments take place are assumed to represent a random sample of the population of possible environments. This implies that E_j is random and, consequently, that the effects of treatment x experiment interaction are also random.

The random effects E_j , tE_{ij} , $B|E_{jk}$ and ε_{ijk} are assumed to be normally distributed, with a mean of zero and a variance equal to σ_E^2 , σ_{tE}^2 , σ_B^2 and σ_ε^2 , respectively. In addition, the various random effects are assumed to be independent.

Variance of the Difference Between Two Treatments

As the objective of experimental networks is generally to compare treatment means, it is interesting to examine the formula of the variance of the estimator used to estimate the difference between the means of two treatments. In model (3.1), the variance of the difference between the means of two treatments is (Littell et al. 1996):

$$\text{Var}(\hat{\mu}_i - \hat{\mu}_{i'}) = 2 \left(\frac{\sigma_{tE}^2}{J} + \frac{\sigma_\varepsilon^2}{JK} \right) \quad (3.2)$$

where the true mean (unknown) of treatment i is defined by $\mu_i = \mu + t_i$ and the estimate of this mean is denoted $\hat{\mu}_i$. $\text{Var}(\hat{\mu}_i - \hat{\mu}_{i'})$ is the variance of the estimator of the difference between treatment means i and i' , and σ_{tE}^2 , σ_ε^2 , J and K are the variance of the treatment x experiment interaction, the variance of the residuals, the number of experiments and the number of blocks per experiment.

The variance of the difference between two treatment means depends on the variance of the treatment x experiment interaction, σ_{tE}^2 , and the residual variance, σ_ε^2 , but not on the variance between experiments, σ_E^2 . The number of blocks per experiment, K , affects only the part of the variance due to residual errors, whereas the number of experiments, J , affects both components of the variance of the difference between means.

3.3.2 Model with Experiment as a Fixed Effect

The following model can also be used for the analysis of individual data from a network of experiments with a randomized complete block design including one factor:

$$y_{ijk} = \mu + t_i + e_j + te_{ij} + B|e_{jk} + \varepsilon_{ijk} \quad (3.3)$$

where y_{ijk} is the observed value of the quantity of interest for treatment i ($i = 1, \dots, I$) in block k ($k = 1, \dots, K$) of experiment j ($j = 1, \dots, J$) and μ , t_i , e_j , te_{ij} , $B|e_{jk}$, and ε_{ijk} are the general mean, the effect of treatment i , the effect of experiment j , the effect of the interaction between treatment i and experiment j , the effect of block k in experiment j , and the residual associated with y_{ijk} .

The difference between model (3.3) and model (3.1) lies in the terms relating to the effect of experiment and of the interaction between treatment and experiment, which are considered to be fixed in model (3.3) and random in model (3.1). The blocks are assumed to be representative of the experiment from which they come, and are thus considered to be random as with model (3.1). Model (3.3) is thus a mixed model, because it considers the block effect to be random, whereas the effects of treatment and treatment \times experiment interaction are considered to be fixed.

The random effects $B|e_{jk}$ and ε_{ijk} are assumed to follow a normal distribution with a mean of 0 and a variance of σ_B^2 and σ_ε^2 , respectively. Furthermore, the various random effects are assumed to be independent.

Variance of the Difference Between Two Treatments

In model (3.3), the variance of the difference between the means of two treatments is (Littell et al. 1996):

$$\text{Var}(\hat{\mu}_i - \hat{\mu}_{i'}) = 2 \left(\frac{\sigma_\varepsilon^2}{JK} \right) \quad (3.4)$$

where the mean for treatment i is $\mu_i = \mu + t_i$ and the estimate of this mean is denoted $\hat{\mu}_i$. $\text{Var}(\hat{\mu}_i - \hat{\mu}_{i'})$ is the variance of the estimator of the difference between the means of treatments i and i' , and σ_ε^2 , J and K are the variance of the residuals, the number of experiments and the number of blocks per experiment, respectively.

The variance of the difference between the means of two treatments is dependent solely on the variance of the residual errors, σ_ε^2 . It is not dependent on the variance of the treatment x experiment interaction, σ_{TE}^2 , unlike the variance obtained with model (3.1).

3.3.3 Example

We illustrate the difference between a model with experiment as a random effect and a model with experiment as a fixed effect, in Figs. 3.1 and 3.2, in which we present some results for the analysis of the data for the “wheat” example obtained with models (3.1) and (3.3).

The first part of Fig. 3.1 shows the variance components table for model (3.1), as estimated from the “wheat” dataset. The variance components shown are the variances of the random effects of the model. A high variance indicates that there are large differences between the levels of the factor considered, whereas a low variance indicates that there is little difference between these levels. A variance of zero indicates that there is no difference between the levels of the factor, or equivalently, that all the effects of the factor are zero. The “group” column in the table identifies the effects and the “variance” column shows the estimated variances of these effects. The terms “Experiment:variety”, and “Experiment:block”, “Experiment” and “Residual” correspond to the effects of the experiment x variety interaction, the effects of the experiment x block interaction, the effects of experiments and the residuals of the model. As is frequently the case in experimental networks for variety evaluation, the variance of the effects of experiments is markedly higher than the other variance components.

The second part of Fig. 3.1 shows the results of a contrast, which is a particular type of comparison of means. Here, we compare the mean value for variety V1 with

<i>Variance component</i>					
Group	Variance				
Experiment:variety	6.5261				
Experiment:block	5.0578				
Experiment	77.3305				
Residual	6.1034				
<i>Comparison of two treatments</i>					
contrast	estimate	SE	df	lower.CL	upper.CL
V1 - V10	2.3832329	1.85047	36	-1.3696950	6.13616077

Fig. 3.1 Analysis of the “wheat” example with the mixed model (3.1) considering experiment as a random effect

<i>Variance components</i>					
Group	Variance				
Experiment:block	5.0577				
Residual	6.1034				
<i>Comparison of two treatments</i>					
contrast	estimate	SE	df	lower.CL	upper.CL
V1 - V10	2.3832329	0.9021032	90	0.59104735	4.1754184

Fig. 3.2 Analysis of the “wheat” example with mixed model (3.3) considering experiment as a fixed effect

that of variety V10. The “estimate”, “SE”, “df”, “lower.CL” and “upper.CL” columns show the estimated contrast, the standard error of the estimator, the number of degrees of freedom associated with the variance of the estimator and the lower and upper limits of the confidence interval for the estimate with an alpha risk of 0.05. The variance of the estimator can be calculated with equation (3.2). By replacing the

theoretical variances with the estimates in Fig. 3.1, bearing in mind that there are $J = 5$ experiments in the network, each with $K = 3$ blocks, we obtain: $\widehat{Var}(\hat{\mu}_i - \hat{\mu}_j) = 2 \left(\frac{6.5261}{5} + \frac{6.1034}{5*3} \right) = 3.42$. The square root of this variance is 1.85, corresponding to the standard error value indicated in Fig. 3.1. The lower and upper limits of the confidence interval for the estimate are given by the following formula:

$$estimate \pm t_{0.975;36} SE$$

where SE is the standard error of the estimator and $t_{0.975; 36}$ is the 97.5th percentile of a Student's t distribution with 36 degrees of freedom. The value of $t_{0.975; 36}$ is 2.02809, and, using the values given in Fig. 3.1, we obtain: $2.3832329 \pm 2.02809 * 1.85047 = -1.3697$ and 6.136 , corresponding to the confidence interval shown in Fig. 3.1.

The first part of Fig. 3.2 shows the table of variance components for model (3.3), estimated with the “wheat” dataset. It does not include the “Experiment:variety” or “Experiment” components, because these effects are fixed in model (3.3). By contrast, the variances for the “Experiment:block” effects and “Residual” are included in model (3.3), as in model (3.1). The variances of these effects are identical for the two models.

The second part of Fig. 3.2 shows the results for the comparison of the mean values of varieties V1 and V10. The variance of the estimator can be calculated with Eq. (3.4), by replacing the theoretical variances with the estimates from Fig. 3.2, bearing in mind that there are $J = 5$ experiments in the network, each with $K = 3$ blocks. We obtain: $\widehat{Var}(\hat{\mu}_i - \hat{\mu}_j) = 2 \left(\frac{6.1034}{5*3} \right) = 0.814$.

The square root of this variance is 0.902, corresponding to the value of the standard error given in Fig. 3.2. The lower and upper limits of the confidence interval for this contrast estimate are provided by the same equation as for model (3.1), but using the 97.5th percentile of a Student's t distribution with 90 degrees of freedom: $estimate \pm t_{0.975; 90} SE$.

If we compare the results for models (3.1) and (3.2) presented in Figs. 3.1 and 3.2, we can see that the estimated contrast is the same for the two models. By contrast, the standard error of the estimator, the number of degrees of freedom associated with the variance of the estimator and the limits of the confidence interval for the estimate are different:

- The estimate provided by model (3.3) is more precise than that provided by model (3.1), which has a standard error about twice that of model (3.3) for this example,
- The number of degrees of freedom associated with the variance of the estimator is higher for model (3.3) than for model (3.1). For model (3.3), the variance of the estimator is dependent solely on the residual variance. The number of degrees of freedom of this variance can be calculated as follows: $(I-1)*(K-1)*J = (10-1)*(3-1)*5 = 90$. For model (3.1), the variance of the estimator is dependent on the residual variance, but also on the variance of the interaction between varieties and experiments, and the number of degrees of freedom of the variance of the estimator is that associated with the variance of the interaction between varieties and experiments. It can be calculated as follows: $(I-1)*(J-1) = (10-1)*(5-1) = 36$.

- The differences in standard error and degrees of freedom between models (3.1) and (3.3) result in the calculation of a confidence interval for the difference in mean values between varieties V1 and V10 that is narrower for model (3.3) than for model (3.1). In this example, the confidence interval obtained with model (3.3) does not include the value 0, and we can, therefore, conclude that there is a significant difference between the two varieties, at an alpha risk of 0.05. By contrast, for model (3.1), the confidence interval includes the value 0 and the difference between the two varieties therefore appears to be non-significant.

3.3.4 *Should Experiment Effect be Considered Fixed or Random?*

As shown above with the “wheat” example, very different conclusions may be drawn from analyses based on a model with experiment effect considered fixed and a model with experiment effect considered random. In general, it is easier to demonstrate the existence of differences between treatments with the fixed-effects model than with the random-effects model. However, the choice between the two types of model should not be based on the conclusions drawn, but on the objectives of the experimental network.

The choice between a model with experiment effect as fixed and a model with experiment effect as random depends primarily on the way in which the environments of the experiments are chosen, which itself depends on the objectives of the experimental network. The choice between the two types of model may also depend on the number of experiments included in the network.

The choice between model (3.1) and model (3.3) should be based on the way in which the environments for the experiments are selected. If the environments are chosen at random from a population of environments, then model (3.1) should be used. In this case, the conclusions of the analysis can be extended to the population of environments. If the environments are chosen deliberately on the basis of particular features, then model (3.3) should be used. In this case, the inference space (i.e. the scope of the conclusions) is restricted to the selected environments. For model (3.1), the objective of the network is to draw conclusions that apply at the scale of the population of environments. The analysis of an experimental network with model (3.1) provides estimates of theoretical means of treatments which are the means over the entire population of environments. For model (3.3), the objective of the network is to draw conclusions restricted to the environments observed in the network. Experimental network analyses with model (3.3) generate estimates of theoretical means of treatments which are the means over the environments observed only. The results obtained with this model therefore have a much narrower inference space. Model (3.3) provides more accurate estimates than model (3.1), as illustrated by the results for the “wheat” example presented in Figs. 3.1 and 3.2, but it should be borne in mind that the two models are not designed to estimate the same quantities.

In some situations, we wish to draw conclusions valid at the level of the population of environments and we therefore choose environments at random to constitute an experimental network, but the number of environments observed in the network is small. In this case, an analysis of the data with model (3.1) may be problematic. Indeed, in a mixed model, the smaller the number of levels of a random factor available, the greater the uncertainty of the estimated variance of the effects of this factor. Furthermore, it is also highly likely that the variance of the effects of this factor will be estimated at zero, not because it really is zero, but because the data contain too little information for the correct assessment of their variance. In experimental networks, estimates of the variance between experiments are very imprecise if the number of experiments is small. If, in addition, the number of treatments studied is also small, estimation of the variance of the effects of the interaction between treatments and experiments is also highly imprecise, with a high probability of being estimated at zero, not because there is no interaction, but because the data contain too little information for correct estimation. Thus, if the network includes fewer than five experiments, we recommend the use of model (3.3) for analysis of the data, bearing in mind that this decreases the scope of the conclusions.

3.4 Evaluation of the Model

Statistical models are used to interpret the data generated in experimental networks. For the interpretation to be valid, the plausibility of the model used must be evaluated.

Analyses of experimental networks based on model (3.1) require checks of the hypotheses of normality, independence and homoscedasticity for the random effects of the model. It is also important to check for the absence of aberrant data. These checks are often limited to the residuals of the model, but for full rigor, they should also be applied to the other random effects, such as the “experiment”, “block” and “experiment x treatment” interaction effects for model (3.1).

3.4.1 Normality

Several different tests of normality can be used to test this hypothesis, including the widely used Shapiro-Wilk test (Dagnelie 1998). It is also possible to calculate the values of the symmetry and kurtosis coefficients and to compare them with the theoretical values for a normal distribution (Dagnelie 1992; Dagnelie 1998). However, a visual validation, based on a histogram or a boxplot, for example, is often considered sufficient.

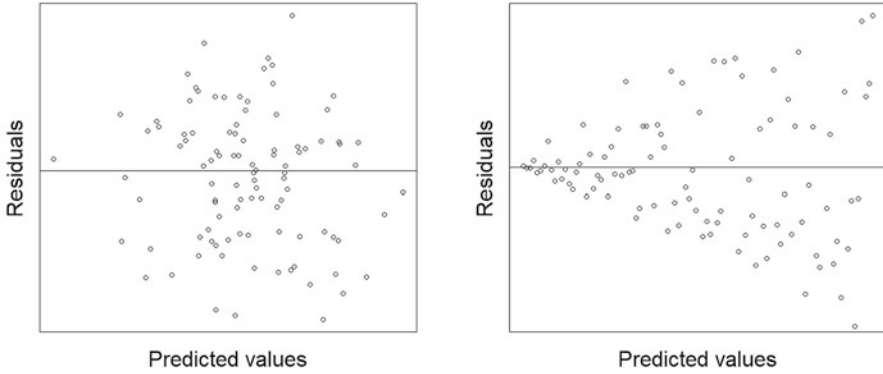


Fig. 3.3 Plot of residuals against predicted values. On the left, the result expected in cases of homoscedasticity. On the right, a graph revealing the existence of a relationship between the variance and the mean

3.4.2 *Homoscedasticity*

Validation of the hypothesis of homoscedasticity involves checking that the residual variance remains constant. Heteroscedasticity is the term used to describe situations in which this is not the case. Heteroscedasticity occurs principally in two types of situation. In the first, it arises because some treatments or experiments are more heterogeneous than others, due to their specific characteristics. For example, an experiment performed on a shallow soil is likely to display greater experimental variability than an experiment performed on a deep soil. The other type of situation in which heteroscedasticity is observed is that in which there is a link between the variance of a variable and its mean, as for count data, for example. This situation frequently arises in agricultural studies.

Several different homoscedasticity tests, including that of Bartlett (Dagnelie 1998), can be used to test for homoscedasticity, but a visual validation based on a graph of the relationship between the residuals of the model and their predicted values is often considered sufficient. The predicted values correspond to the expected values of the variable of interest conditional on the random effects, expressed as follows for model (3.1): $\mu + t_i + E_j + tE_{ij} + B|E_{jk}$. A scatter plot can be used to determine whether there is a link between the variance and the mean, as shown in Fig. 3.3.

3.4.3 *Independence*

The hypothesis of independence can be tested by checking that the random effects and residuals of the model are not dependent on the mode of data acquisition. For example, in the case of a multiyear network, we might expect the experiments performed in the same year to yield similar results, reflecting the year effect, and

the effects of these experiments would therefore not be considered independent. Similarly, the results of experiments performed at the same or nearby sites would be expected to more similar than those of experiments performed at sites farther apart. Chapter 4 deals with the question of the factors “site” and “year”.

Generally, it is important to check that no external factor not taken into account by the model used to describe the data is linked to the phenomenon studied: chronology of data acquisition, researcher, etc.

3.4.4 *Suspect Data*

The problem of aberrant data is generally dealt with by checking for the absence of suspect residuals, but this problem may also apply to other random effects. Here again, a visual examination of the residuals is generally practiced, by checking on a plot of the residuals against predicted values that none of the residuals is much further away than the others. In addition to this visual examination, a test for suspect data, such as the Grubbs test (Dagnelie 1998), can be used.

The detection of a suspect residual invariably raises questions about how best to deal with it. Should it be retained or eliminated? In cases of doubt, it may be useful to perform two analyses: one in which the suspect residual is retained, and another in which it is eliminated.

3.5 Comparisons of Means

The principal objective of experimental networks is to compare the study treatments with each other. In practice, this comparison involves an estimation of the means of the treatments, hypothesis testing and the calculation of confidence intervals (Dagnelie 1992).

3.5.1 *Hypothesis Tests: Tests of Equality*

At the end of an experiment, tests are classically performed to determine whether the means of the study treatments are equal. In an experimental network, the number of treatments studied is generally greater than two, and may even be very high, with several tens of treatments studied. In situations in which there are more than two treatments, the first hypothesis tested is generally that of the equality of all the treatment means:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_i = \dots = \mu_I$$

where μ_i is the mean for treatment i and I is the total number of treatments.

Fisher's F test can be used to test whether all the treatment means are equal (Dagnelie 1998).

The F test is a global test. When the hypothesis of equal treatment means is rejected in this test, the next step is identifying which treatments are different from the others. Many methods for comparing means exist and can be used for this purpose, and the choice of method depends on the question addressed. The objective of comparisons of means is often either to compare the various treatment means with a reference treatment or to establish homogeneous groups of treatments with means that are not significantly different. In both cases, the methods to be used are part of the multiple comparison methods (Dagnelie 1998).

The common feature of multiple mean comparison methods is that they have to manage the inflation of the type I error risk associated with multiple hypothesis tests (Hothorn et al. 2008). Type I error or alpha risk is the probability of concluding that there is a significant difference between treatments when, in reality, they are identical. This risk is often set to 5% ($\alpha = 0.05$). If the probability of wrongly concluding that there is a significant difference between two treatments is 0.05, then the probability of not concluding that there is a significant difference between two treatments that really are identical is $1 - 0.05 = 0.95$. Let us consider three treatments, A, B, and C, for which we wish to perform all the possible pairwise comparisons. We therefore need to test three hypotheses: $\mu_A = \mu_B$, $\mu_A = \mu_C$, and $\mu_C = \mu_B$. Each hypothesis is tested with an alpha risk of 0.05, or, put another way, with a probability of not concluding there is a significant difference between two treatments that are, in reality, identical, of 0.95. The risk of not concluding that there is a significant difference between treatments that are, in reality, identical, for the entire set of three tests, is equal to the risk of not concluding that there is a significant difference in the first test, the second test or the third test. The overall probability is, therefore, $0.95 * 0.95 * 0.95 = 0.95^3 = 0.857$, if the three tests are independent. The global type I error, the risk of concluding that there is a significant difference between two treatments that are, in reality, identical, in at least one of the comparisons performed, is thus, $1 - 0.857 = 0.143$. This global alpha risk is much higher than the nominal alpha risk of 0.05 chosen for each of the tests. If we compare 10 treatments, the number of pairwise comparisons to be performed is $(10^2 - 10) / 2 = 45$, and the global alpha risk is $1 - 0.95^{45} = 0.90$! In other words, if we compare 10 treatments that are actually identical, there is a nine in ten chance of identifying at least one significant difference between two treatments. In reality, pairwise comparison tests are not independent, and the global alpha risk lies somewhere between the upper limit calculated as shown above and the nominal alpha risk selected for each test performed. Tests for the comparison of multiple means aim to provide the highest possible level of power, whilst controlling the global alpha risk and restricting it to the value chosen.

In this book, we use Dunnett's method (Dagnelie 1998) to compare various treatment means with a reference treatment, Tukey's method for the establishment of homogeneous groups and Sidak's method to compare each treatment mean with the general mean (SAS Institute Inc. 1989). Other methods are also available.

3.5.2 *Confidence Intervals*

Many criticisms have been made of tests of equality between means. The p values associated with such hypothesis tests are the tool most widely used for data interpretation, but they are often misinterpreted (Littell et al. 2006; Madden et al. 2015; Lew 2013).

One of the principal criticisms of tests of equality between means is that, in real applications, the null hypothesis is never exactly true and will, therefore, always be rejected if the number of observations is sufficiently large. It is therefore not relevant to test whether there is a difference between two treatments, A and B, because there always is one, even if it is extremely small (Tukey 1991).

Most experiments aim to estimate the difference between treatment means and to quantify the uncertainty associated with the estimate, rather than to prove that there is a difference between treatments. In this case, the information provided by the p value obtained in a hypothesis test can be usefully complemented by the calculation of a confidence interval. The confidence interval provides information about the precision of the estimate of the effect of a treatment or of the difference between two treatments (SAS Institute Inc. 1989), even if this property is not always verified (Morey et al. 2015). Figure 3.4 illustrates the value of calculating a confidence interval when performing a test of equal means to compare two treatments. It displays the difference between two treatments in five separate experiments. The interval around each point corresponds to the 95% confidence interval for the difference between treatments. The p values for tests of equality between two treatments are not represented, but we can affirm that the difference is significant with an alpha risk of 5% for the situations in which the confidence intervals do not include the value zero (Dagnelie 1992). In situations 1 and 4, the difference between the two treatments is not significant. We could come to this conclusion based on the test of equality between means alone, but we can take the interpretation further based on the confidence intervals obtained. In situation 1, we can affirm that if there is a difference between treatments, this difference is probably small and would possibly be of little interest in practice. By contrast, for situation 4, the difference is not significant either, but if there is a real difference between treatments, it could be large. We cannot, therefore draw any operational conclusions from experiment 4. The only valid conclusion for this experiment is that we do not have sufficient information to determine the real size of the difference between the two treatments. For experiments 2, 3 and 5, the confidence intervals do not contain zero, providing a statistical demonstration of the existence of a significant difference between the two treatments. However, the practical conclusions to be drawn from these three experiments are very different. In experiment 2, the confidence interval is very narrow, indicating that the real difference between treatments is probably small and unlikely to be of interest in practice. In experiment 3, the confidence interval is very narrow and indicates that the real difference between treatments is probably large and likely to be of considerable interest. Finally, in experiment 5, the very large confidence interval indicates that the real difference between the treatments could turn out to be

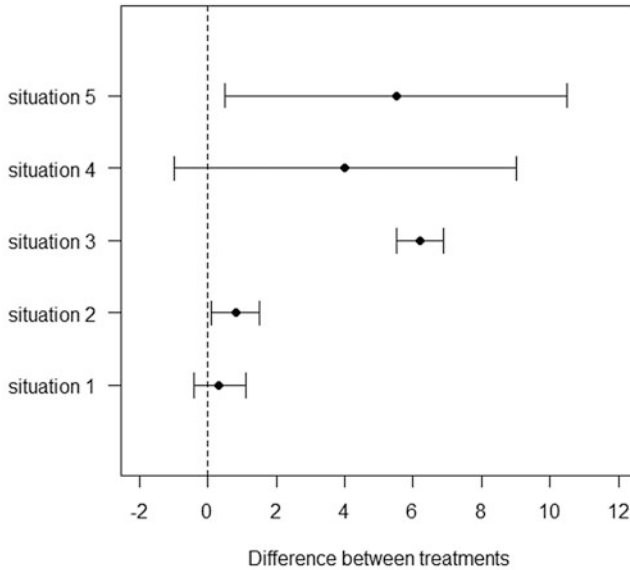


Fig. 3.4 Estimated difference between two treatments and 95% confidence interval for this difference in five experiments (situations)

very small, very large, or anything in between. Thus, as for experiment 4, we can draw no firm conclusions given that we lack the necessary information to determine the real difference between the two treatments.

3.5.3 Hypothesis Tests: Tests of Equivalence

Tests of equality are often used in an abusive manner to demonstrate equality between the treatments studied. Tests of equality between means are performed with the aim of proving that there are differences between treatments, but a non-significant difference in such tests is not sufficient to affirm that two treatments are identical. Such a result indicates only that we have not managed to prove that they are different (Jaykaran et al. 2011). Indeed, a non-significant difference between treatments may be due to a real absence of difference between the treatments concerned, but it may also result from the experimental variability being too high or from the use of too few replicates to be able to demonstrate a real difference between treatments.

The objective of an experiment can be to prove that the effect of a treatment is not different from that of another treatment. This is the case, for example, when trying to show that a genetically modified crop has characteristics equivalent to those of a

conventional crop. Another example is provided by tests of the efficacy of biological control products, in which the aim may be to demonstrate an efficacy similar to that for a classic fungicide treatment. In such situations, tests of equality are not appropriate to address the question posed. It is therefore preferable to use a test of equivalence, with the aim of demonstrating the equality of treatments.

The principle of a test of equivalence is to define an equivalence bound, within which the difference between two treatments is considered negligible. An equivalence threshold Δ is selected by defining the largest difference, in absolute terms, that can be accepted for two treatments to be considered equivalent. A difference greater than Δ is considered to be large in practice.

Tests of equivalence are based on the following hypotheses:

$$H_0 : \mu_1 - \mu_2 \leq -\Delta \text{ or } \mu_1 - \mu_2 \geq \Delta$$

$$H_1 : -\Delta < \mu_1 - \mu_2 < \Delta$$

where μ_1 and μ_2 are the theoretical means of the two treatments.

According to the null hypothesis H_0 , the two treatments are not equivalent. According to the alternative hypothesis H_1 , the two treatments are equivalent. The rejection of the null hypothesis in favor of the alternative hypothesis implies an acceptance of the hypothesis of equivalence between the two treatments. The risk of concluding that the two treatments are equivalent, when, in reality, the difference between them, in absolute values, is greater than Δ , is equal to the alpha risk or type I error used to perform the test.

The most widely used procedure for equivalence tests is TOST procedure (two one-sided t tests), which involves performing two one-tailed hypothesis tests (Schuirmann 1987). In this procedure, equivalence is considered demonstrated if it is possible to conclude that both $\mu_1 - \mu_2 > -\Delta$ and $\mu_1 - \mu_2 < \Delta$.

In practice, the calculation of confidence intervals is recommended when performing equivalence tests (Committee for Proprietary Medicinal Products 2001). Figure 3.5 illustrates the interpretation of an equivalence test performed with the calculation of confidence intervals. If the confidence interval of the difference between treatments lies entirely within the equivalence bound, then the two treatments can be considered equivalent. If not, the null hypothesis of non-equivalence cannot be rejected (Robinson and Froese 2004). The conclusions on the interpretation of a confidence interval are identical to those of the TOST test performed with a type I error risk alpha in the case of an equal-tailed confidence interval at the $100(1-2\alpha)\%$ confidence level (Berger and Hsu 1996).

Certain similarities in interpretation can be noted between Figs. 3.4 and 3.5. One of the advantages of the equivalence testing approach is that it obliges the user to specify an equivalence threshold Δ beyond which a difference between two treatments is considered practically relevant. An equivalence threshold is often used in the interpretation of experimental results, whatever the approach used, but not necessarily explicitly. Equivalence testing has the advantage of rendering the interpretation of results more transparent.

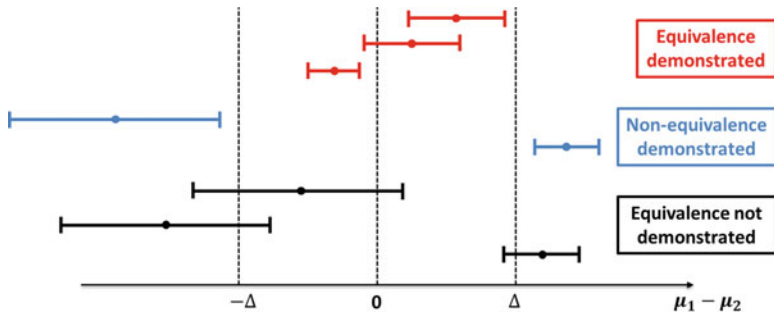


Fig. 3.5 Interpretation of an equivalence test based on confidence intervals

```

Response: yield

          F Df Df.res    Pr(>F)
variety   4.3791  9    36 0.0006507 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    
```

Fig. 3.6 Analysis of the “wheat” example: *F* test for the factor “variety”

3.5.4 Example

We illustrate the different methods of comparing means here, by presenting, in Figs. 3.6, 3.7, 3.8, 3.9, 3.10 and 3.11, several results of data analyses for the “wheat” example performed with model (3.1).

Figure 3.6 shows the results of an *F* test of the equality of the means for the 10 varieties. The “*F*” column shows the statistic generated by the test, which is assumed to follow a Fisher distribution. The columns “*Df*” and “*Df.res*” show the number of degrees of freedom of the numerator and denominator, respectively, of the Fisher distribution, and the “*Pr(>F)*” column show the probability associated with the test statistic. The row “*Variety*” tests the hypothesis that the mean yield of the 10 varieties is identical. The probability associated with the test result, which is well below 0.05, indicates that this hypothesis should be rejected: at least one of the varieties clearly has a yield different from the others.

Figure 3.7 shows the results of a Dunnett test comparing the means of the varieties with the mean of variety V2, used as a control. The “*contrast*” column indicates the comparison performed, the “*estimate*” column shows the estimates of the differences between the varieties and variety V2, and the “*SE*” and “*df*” columns show the standard error of the estimator and the degrees of freedom associated with the variance of the estimator. The “*t.ratio*” column show the value of the test statistic,

contrast	estimate	SE	df	t.ratio	p.value
V1 - V2	3.595430	1.85047	36	1.943	0.3095
V10 - V2	1.212197	1.85047	36	0.655	0.9612
V3 - V2	5.859157	1.85047	36	3.166	0.0230
V4 - V2	7.828792	1.85047	36	4.231	0.0012
V5 - V2	5.978395	1.85047	36	3.231	0.0196
V6 - V2	4.879941	1.85047	36	2.637	0.0811
V7 - V2	6.348242	1.85047	36	3.431	0.0116
V8 - V2	8.812733	1.85047	36	4.762	0.0003
V9 - V2	4.584436	1.85047	36	2.477	0.1143

P value adjustment: dunnett method for 9 tests

Fig. 3.7 Analysis of the “wheat” example: comparison with a control. Dunnett’s test

which is assumed to follow a Student’s t distribution, and the “p.value” column is the probability associated with the test statistic. This probability is adjusted by Dunnett’s method, to take multiple testing into account and to guarantee a global risk of error equal to the risk chosen by the user. If the chosen global alpha risk is 0.05, then differences with a probability of 0.05 or less are considered significant. In this example, varieties V3, V4, V5, V7 and V8 are significantly different from variety V2 at an alpha risk of 0.05.

Figure 3.8 shows the confidence intervals for the difference between each variety and the control variety, V2. The “contrast” column indicates the comparison performed, the “estimate” column shows the estimated differences between the varieties and variety V2, the “SE” and “df” columns show the standard error of the estimator and the degrees of freedom associated with the variance of the estimator, respectively. The “lower.CL” and “upper.CL” columns show the lower and upper limits, respectively, of the confidence interval for the difference between each variety and the control variety, V2. The confidence intervals are adjusted by Dunnett’s method to take into account the multiplicity of calculated intervals and to guarantee a global level of confidence equal to that selected by the user.

Figure 3.9 shows the results of a Tukey test comparing the means of varieties two by two. The “contrast” column indicates the comparison performed, the “estimate” column shows the estimated difference between the varieties and the “SE” and “df” columns show the standard error of the estimator and the degrees of freedom associated with the variance of the estimator, respectively. The “t.ratio” column

contrast	estimate	SE	df	lower.CL	upper.CL
V1 - V2	3.595430	1.85047	36	-1.6763394	8.867199
V10 - V2	1.212197	1.85047	36	-4.0595723	6.483966
V3 - V2	5.859157	1.85047	36	0.5873882	11.130927
V4 - V2	7.828792	1.85047	36	2.5570226	13.100561
V5 - V2	5.978395	1.85047	36	0.7066256	11.250164
V6 - V2	4.879941	1.85047	36	-0.3918278	10.151711
V7 - V2	6.348242	1.85047	36	1.0764732	11.620012
V8 - V2	8.812733	1.85047	36	3.5409640	14.084502
V9 - V2	4.584436	1.85047	36	-0.6873328	9.856206

Confidence level used: 0.95

Conf-level adjustment: dunnett method for 9 estimates

Fig. 3.8 Analysis of the “wheat” example: comparison with a control. Calculation of confidence intervals

shows the value of the test statistic, which is assumed to follow a Student’s t distribution, and the “p-value” column is the probability associated with the test statistic. This probability is adjusted by the Tukey method, to take multiple testing into account and to guarantee a global risk of error equal to that chosen by the user. If the global alpha risk selected is 0.05, then all differences with a corresponding probability value of 0.05 or below should be considered significant. In this example, none of the differences shown is significant with an alpha risk of 0.05. The results of these pairwise comparisons can be used to generate a compact representation in the form of homogeneous groups identified by letters. A compact representation of this type is shown in Fig. 3.10.

Figure 3.11 shows the confidence intervals for the effect of variety: the difference between each variety and the general mean. The “contrast” column indicates the comparison performed, the “estimate” column shows the estimated effect of variety,

contrast	estimate	SE	df	t.ratio	p.value
V1 - V10	2.3832329	1.85047	36	1.288	0.9501
V1 - V2	3.5954297	1.85047	36	1.943	0.6415
V1 - V3	-2.2637277	1.85047	36	-1.223	0.9635
V1 - V4	-4.2333621	1.85047	36	-2.288	0.4207
V1 - V5	-2.3829651	1.85047	36	-1.288	0.9501
V1 - V6	-1.2845117	1.85047	36	-0.694	0.9994
V1 - V7	-2.7528127	1.85047	36	-1.488	0.8883
V1 - V8	-5.2173035	1.85047	36	-2.819	0.1681
V1 - V9	-0.9890067	1.85047	36	-0.534	0.9999
V10 - V2	1.2121969	1.85047	36	0.655	0.9996
...					
V7 - V9	1.7638060	1.85047	36	0.953	0.9932
V8 - V9	4.2282968	1.85047	36	2.285	0.4223

P value adjustment: Tukey method for comparing a family of 10 estimates

Fig. 3.9 Analysis of the “wheat” example: Tukey’s test

and the “SE” and “df” columns show the standard error of the estimator and the degrees of freedom associated with the variance of the estimator, respectively. The “lower.CL” and “upper.CL” columns show the lower and upper limits, respectively, of the confidence interval for the variety effect. The confidence intervals were adjusted by Sidak’s method, to take the multiplicity of calculated intervals into account and to guarantee a global level of confidence identical to that chosen by the user.

variety	lsmean	SE	df	lower.CL	upper.CL	.group
V2	75.39573	4.185144	4.8	64.50405	86.28740	A
V10	76.60792	4.185144	4.8	65.71625	87.49959	AB
V1	78.99116	4.185144	4.8	68.09948	89.88283	ABC
V9	79.98016	4.185144	4.8	69.08849	90.87183	ABC
V6	80.27567	4.185144	4.8	69.38400	91.16734	ABC
V3	81.25488	4.185144	4.8	70.36321	92.14656	ABC
V5	81.37412	4.185144	4.8	70.48245	92.26579	ABC
V7	81.74397	4.185144	4.8	70.85230	92.63564	BC
V4	83.22452	4.185144	4.8	72.33285	94.11619	C
V8	84.20846	4.185144	4.8	73.31679	95.10013	C

Confidence level used: 0.95

P value adjustment: Tukey method for comparing a family of 10 estimates

significance level used: alpha = 0.05

Fig. 3.10 Analysis of the “wheat” example: representation of the results of Tukey’s test as homogeneous groups

3.6 Analysis of the “Wheat” Example and R Script

We propose here an R script for the analysis of individual data in a network of experiments with a randomized complete block design.

contrast	estimate	SE	df	lower.CL	upper.CL
V1 effect	-1.31450267	1.241333	36	-5.0157761	2.386770802
V10 effect	-3.69773553	1.241333	36	-7.3990090	0.003537938
V2 effect	-4.90993241	1.241333	36	-8.6112059	-1.208658942
V3 effect	0.94922501	1.241333	36	-2.7520485	4.650498481
V4 effect	2.91885941	1.241333	36	-0.7824141	6.620132884
V5 effect	1.06846240	1.241333	36	-2.6328111	4.769735868
V6 effect	-0.02999101	1.241333	36	-3.7312645	3.671282463
V7 effect	1.43830999	1.241333	36	-2.2629635	5.139583455
V8 effect	3.90280082	1.241333	36	0.2015274	7.604074292
V9 effect	-0.32549601	1.241333	36	-4.0267695	3.375777455

Confidence level used: 0.95

Conf-level adjustment: Sidak method for 10 estimates

Fig. 3.11 Analysis of the “wheat” example: comparison with the general mean. Calculation of confidence intervals

R Script and Commentary on the Analysis

```
# R version 3.4.3
# loading of packages
library(lme4) # version 1.1-14
library(emmeans) # version 1.0
library(car) # version 2.1-6
library(outliers) # Version: 0.14
```

The **lme4** library is used for adjustment of the mixed model, the **emmeans** library is used to calculate the mean for each variety and to compare these means, the **car** library is used to calculate the analysis of variance table and the **outliers** library is used to detect suspect residuals.

```
# Model adjustment
res.lmer <- lmer(rendement ~ variete + (1|experimentation) + (1|
experimentation:variete) + (1|experimentation:bloc), data=DF,
na.action=na.exclude)
res.lmer
```

The **lmer** function of the **lme4** library is used to adjust the mixed model. With the **lmer** function, the random factors are indicated in brackets, preceded by the number “1” and a vertical bar “|”. This syntax indicates that a random effect specific to each level of the factor is considered. The **res.lmer** command displays the result for model adjustment in the R console presented in Fig. 3.12. An estimate of the standard deviation is shown for the effects of “experiment” (8.794), “experiment: block” (2.249), “experiment:variety” (2.555), and the residuals of the model (2.471).

Estimated values of the fixed effects are also provided. With the contrasts used by default for model adjustment in R, the intercept corresponds to the mean for the first variety in alphabetical order, and the other estimates are the differences between the first variety and the other varieties. This constraint on the effect of first variety is an identifiability constraint essential for estimation of the model parameters (Venables and Ripley 1998). Other types of contrasts for model adjustment, implying other identifiability constraints, can also be specified (Galecki and Burzykowski 2013).

```
# Validation of the model
plot(fitted(res.lmer), residuals(res.lmer), abline(h=0))
hist(residuals(res.lmer))
```

The functions **plot** and **hist** generate graphs of the residuals of the model, for the validation of certain hypotheses concerning these residuals. These graphs are presented in Fig. 3.13. The residuals of the model are assumed to follow a normal distribution, with the same variance, and to be independent. These graphs provide no evidence to suggest that the hypotheses of normality and homoscedasticity do not apply. A similar validation should be performed for the other random effects.

```
DF$residus <- residuals(res.lmer)
grubbs.test(DF$residus)
DF[which.max(DF$residus),]
```

The “**residuals**” function stores the residuals of the mixed model as a new variable called “**residus**”. The **grubbs.test** function of the **outliers** library can be used to detect the presence of suspect residuals, and the “**which.max**” function displays the observation corresponding to the residual with the highest value.

```

Linear mixed model fit by REML ['lmerMod']

Formula:
rdt ~ variete + (1 | experimentation) + (1 | experimentation:variete) +
      (1 | experimentation:bloc)

Data: DF

REML criterion at convergence: 775.5289

Random effects:

Groups          Name          Std.Dev.
experimentation:variete (Intercept) 2.555
experimentation:bloc   (Intercept) 2.249
experimentation        (Intercept) 8.794
Residual              2.471

Number of obs: 150, groups:
experimentation:variete, 50; experimentation:bloc, 15; experimentation, 5

Fixed Effects:

(Intercept)  varieteV10  varieteV2  varieteV3  varieteV4  varieteV5
      78.991      -2.383      -3.595       2.264       4.233       2.383
varieteV6  varieteV7  varieteV8  varieteV9
      1.285       2.753       5.217       0.989

```

Fig. 3.12 Information about the mixed model

Grubbs’ test, for which results are presented in Fig. 3.14, detected no suspect residuals. Figure 3.15 shows the observation corresponding to the residual with the highest value.

```

# ANOVA table
anova(res.lmer, test.statistic="F")

```

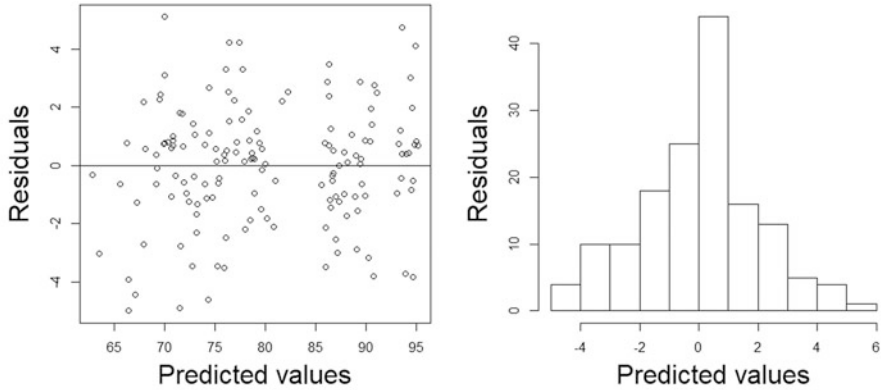


Fig. 3.13 Model validation graphs. On the left: scatterplot of residuals against predicted values. On the right: histogram of residuals

```
data: DF$residus

G = 2.5253, U = 0.9569, p-value = 0.8108

alternative hypothesis: highest value 5.10177381601873 is an outlier
```

Fig. 3.14 Grubbs' test

year	site	experiment	variety	block	yield	residuals
2003	L3	643	V4	1	75.14778	5.101774

Fig. 3.15 Observation corresponding to the residual with the highest value

The **Anova** function of the **car** library can be used to calculate the analysis of variance (ANOVA) table used for F tests. The results are shown in Fig. 3.16. In the framework of a mixed model with “experiment” as a random factor, the means tested correspond to the means of the varieties in the entire population of experiments, of which the 10 experiments observed are considered representative. The probability of the F statistic, 0.0006507, is very low, indicating that the differences between the varieties are highly significant.

```
# Adjusted means
moy_var <- lsmeans(res.lmer, ~variete)
moy_var
```

The **lsmeans** function of the **emmeans** library can be used to calculate the means of the varieties, as shown in Fig. 3.17. These means are adjusted, as indicated by the name “**lsmean**” (least squares means), although in this case, the adjusted

Analysis of Deviance Table (Type II Wald F tests with Kenward-Roger df)

Response: yield

```

      F Df Df.res    Pr(>F)
variety 4.3791  9    36 0.0006507 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Fig. 3.16 F test

```

Fig. 3.17 Adjusted means

```

variety	lsmean	SE	df	lower.CL	upper.CL
V1	78.99116	4.185144	4.8	68.09948	89.88283
V10	76.60792	4.185144	4.8	65.71625	87.49959
V2	75.39573	4.185144	4.8	64.50405	86.28740
V3	81.25488	4.185144	4.8	70.36321	92.14656
V4	83.22452	4.185144	4.8	72.33285	94.11619
V5	81.37412	4.185144	4.8	70.48245	92.26579
V6	80.27567	4.185144	4.8	69.38400	91.16734
V7	81.74397	4.185144	4.8	70.85230	92.63564
V8	84.20846	4.185144	4.8	73.31679	95.10013
V9	79.98016	4.185144	4.8	69.08849	90.87183

```

Confidence level used: 0.95

```

means are identical to the crude means because the data table is complete. The notion of an “adjusted mean” will be presented in detail in Chap. 4.

```

# Pairwise comparisons
pairs(moy_var, adjust="tukey")
cld(moy_var, Letters=c(LETTERS))

```


The `pairs` function of the `emmeans` library performs all possible pairwise comparisons between means, using the “`tukey`” method to adjust the probability to take multiple testing into account. Finally, the `cld` function of the `emmeans` library can be used to establish homogeneous groups.

Figure 3.18 shows pairwise comparisons between the means of the varieties. The estimated differences are indicated in the “`estimate`” column. A statistical test is performed to test the hypothesis of an absence of difference. The probability of the tests is adjusted to take multiple testing into account and to keep the global type I error at the level selected by the user. This means, for example, that, if we decide to consider differences with a p value of 0.05 or less significant, then there is a global type I risk of error of 5%.

Figure 3.19 presents pairwise comparisons of the means of varieties in the form of homogeneous groups indicated by letters, assuming an alpha risk of 5%. Two varieties that have no letter in common are considered to be significantly different, with an alpha risk of 5%. This is the case, for example, for varieties V2 and V7. By contrast, varieties V10 and V7 are not significantly different because they have the letter “B” in common.

```
# Comparisons with the general mean
contrast(moy_var, method="eff", adjust="sidak")
confint(contrast(moy_var, method="eff", adjust="sidak"))
```

The `contrast` function of the `emmeans` library, used with the “`eff`” method, calculates all the differences between each variety and the general mean. The “`sidak`” method adjusts the probability of tests to take multiple testing into account. The `confint` function of the `emmeans` library can be used to calculate confidence intervals for “variety” effects.

Figure 3.20 presents “variety” effects, the difference between the mean for each variety and the general mean. The “`contrast`” column indicates the comparison performed, the “`estimate`” column shows the estimated variety effects and the “`SE`” and “`df`” columns show the standard error of the estimator and the degrees of freedom associated with the variance of the estimator, respectively. The “`t.ratio`” column shows the test statistic, for testing whether the “variety” effect is equal to zero, and the “`p.value`” column indicates the probability associated with the test statistic. The test probabilities are adjusted by the Sidak method to take multiple testing into account and to guarantee a global alpha risk equal to that chosen by the user. At the alpha risk threshold of 0.05, variety V2 has a mean significantly lower than the general mean, whereas the mean of variety V8 is significantly higher than the general mean.

Figure 3.21 presents the “variety” effects, i.e. the differences between the mean of each variety and the general mean, and the confidence intervals of these effects, calculated with 95% confidence. The “`contrast`” column indicates the comparison performed, the “`estimate`” column shows the estimated variety effect and the

contrast	estimate	SE	df	t.ratio	p.value
V1 - V10	2.3832329	1.85047	36	1.288	0.9501
V1 - V2	3.5954297	1.85047	36	1.943	0.6415
V1 - V3	-2.2637277	1.85047	36	-1.223	0.9635
V1 - V4	-4.2333621	1.85047	36	-2.288	0.4207
V1 - V5	-2.3829651	1.85047	36	-1.288	0.9501
V1 - V6	-1.2845117	1.85047	36	-0.694	0.9994
V1 - V7	-2.7528127	1.85047	36	-1.488	0.8883
V1 - V8	-5.2173035	1.85047	36	-2.819	0.1681
V1 - V9	-0.9890067	1.85047	36	-0.534	0.9999
V10 - V2	1.2121969	1.85047	36	0.655	0.9996
V10 - V3	-4.6469605	1.85047	36	-2.511	0.2966
V10 - V4	-6.6165949	1.85047	36	-3.576	0.0301
V10 - V5	-4.7661979	1.85047	36	-2.576	0.2655
V10 - V6	-3.6677445	1.85047	36	-1.982	0.6162
V10 - V7	-5.1360455	1.85047	36	-2.776	0.1833
V10 - V8	-7.6005364	1.85047	36	-4.107	0.0074
V10 - V9	-3.3722395	1.85047	36	-1.822	0.7173
V2 - V3	-5.8591574	1.85047	36	-3.166	0.0802
V2 - V4	-7.8287918	1.85047	36	-4.231	0.0052
V2 - V5	-5.9783948	1.85047	36	-3.231	0.0692
V2 - V6	-4.8799414	1.85047	36	-2.637	0.2379
V2 - V7	-6.3482424	1.85047	36	-3.431	0.0431
V2 - V8	-8.8127332	1.85047	36	-4.762	0.0011
V2 - V9	-4.5844364	1.85047	36	-2.477	0.3138

Fig. 3.18 Pairwise comparisons of variety means

V3 - V4	-1.9696344	1.85047	36	-1.064	0.9852
V3 - V5	-0.1192374	1.85047	36	-0.064	1.0000
V3 - V6	0.9792160	1.85047	36	0.529	0.9999
V3 - V7	-0.4890850	1.85047	36	-0.264	1.0000
V3 - V8	-2.9535758	1.85047	36	-1.596	0.8413
V3 - V9	1.2747210	1.85047	36	0.689	0.9994
V4 - V5	1.8503970	1.85047	36	1.000	0.9904
V4 - V6	2.9488504	1.85047	36	1.594	0.8425
V4 - V7	1.4805494	1.85047	36	0.800	0.9981
V4 - V8	-0.9839414	1.85047	36	-0.532	0.9999
V4 - V9	3.2443554	1.85047	36	1.753	0.7583
V5 - V6	1.0984534	1.85047	36	0.594	0.9998
V5 - V7	-0.3698476	1.85047	36	-0.200	1.0000
V5 - V8	-2.8343384	1.85047	36	-1.532	0.8703
V5 - V9	1.3939584	1.85047	36	0.753	0.9988
V6 - V7	-1.4683010	1.85047	36	-0.793	0.9983
V6 - V8	-3.9327918	1.85047	36	-2.125	0.5228
V6 - V9	0.2955050	1.85047	36	0.160	1.0000
V7 - V8	-2.4644908	1.85047	36	-1.332	0.9392
V7 - V9	1.7638060	1.85047	36	0.953	0.9932
V8 - V9	4.2282968	1.85047	36	2.285	0.4223

P value adjustment: Tukey method for comparing a family of 10 estimates

Fig. 3.18 (continued)

```

variety  lsmean      SE  df lower.CL upper.CL .group
V2       75.39573  4.185144  4.8  64.50405  86.28740  A
V10      76.60792  4.185144  4.8  65.71625  87.49959  AB
V1       78.99116  4.185144  4.8  68.09948  89.88283  ABC
V9       79.98016  4.185144  4.8  69.08849  90.87183  ABC
V6       80.27567  4.185144  4.8  69.38400  91.16734  ABC
V3       81.25488  4.185144  4.8  70.36321  92.14656  ABC
V5       81.37412  4.185144  4.8  70.48245  92.26579  ABC
V7       81.74397  4.185144  4.8  70.85230  92.63564  BC
V4       83.22452  4.185144  4.8  72.33285  94.11619  C
V8       84.20846  4.185144  4.8  73.31679  95.10013  C

Confidence level used: 0.95

P value adjustment: Tukey method for comparing a family of 10 estimates

significance level used: alpha = 0.05

```

Fig. 3.19 Homogeneous groups

columns “SE” and “df” correspond to the standard error of the estimator and the degrees of freedom associated with the variance of the estimator. The “lower.CL” and “upper.CL” columns show the lower and upper limits, respectively, of the confidence interval for the variety effect. The confidence intervals are adjusted by the Sidak method, to take the multiplicity of calculated intervals into account and to guarantee a global confidence level equal to that chosen by the user.

```

# Comparison with a control
contrast(moy_var, method="trt.vs.ctrl", ref=2)
confint(contrast(moy_var, method="trt.vs.ctrl", ref=2))

```

The `contrast` function with the `method="trt.vs.ctrl"` argument and `ref=2` can be used to compare all the varieties to the second variety in alphabetical order, V10 in this case, which is treated as the control. The `confint`

Fig. 3.20 Comparisons of variety means with the general mean

contrast	estimate	SE	df	t.ratio	p.value
V1 effect	-1.31450267	1.241333	36	-1.059	0.9704
V10 effect	-3.69773553	1.241333	36	-2.979	0.0504
V2 effect	-4.90993241	1.241333	36	-3.955	0.0034
V3 effect	0.94922501	1.241333	36	0.765	0.9974
V4 effect	2.91885941	1.241333	36	2.351	0.2180
V5 effect	1.06846240	1.241333	36	0.861	0.9934
V6 effect	-0.02999101	1.241333	36	-0.024	1.0000
V7 effect	1.43830999	1.241333	36	1.159	0.9468
V8 effect	3.90280082	1.241333	36	3.144	0.0328
V9 effect	-0.32549601	1.241333	36	-0.262	1.0000

P value adjustment: Sidak method for 10 tests

function of the `emmeans` library can be used to calculate the confidence intervals for differences between the varieties and the control variety, V10.

In practice, the choice between pairwise comparisons to generate homogeneous groups and comparisons with a control depends on the objectives of the study. Only one test is generally used, but both tests are presented here as an illustration.

Figure 3.22 shows the comparisons between variety means and the mean of the control variety, V10. The estimated differences are shown in the “estimate” column. A statistical test is performed to test the hypothesis of an absence of difference. The probability of the test is adjusted to take multiple testing into account and to maintain the type I error at the level chosen by the user. That means, for example, that if we decide to consider differences with a *p* value of 0.05 or lower significant, then there is a global risk of error of 5%. With a global alpha risk of 5%, only varieties V4 and V8 are significantly different from variety V10.

Figure 3.23 reports the confidence intervals for the differences between each variety and the control variety, V10. The “contrast” column indicates the comparison performed, the “estimate” column shows the estimated difference between each variety and variety V10 and the “SE” and “df” columns show the standard error of the estimator and the degrees of freedom associated with the variance of the estimator, respectively. The “lower.CL” and “upper.CL” columns show the lower and upper limits, respectively, of the confidence interval for the difference between each variety and the control variety, V10. The confidence intervals are adjusted by Dunnett’s method, to take the multiplicity of calculated intervals into account and to guarantee a global level of confidence equal to that chosen by the user. The Dunnett

```

contrast      estimate      SE df  lower.CL  upper.CL

V1 effect    -1.31450267  1.241333  36 -5.0157761  2.386770802

V10 effect   -3.69773553  1.241333  36 -7.3990090  0.003537938

V2 effect    -4.90993241  1.241333  36 -8.6112059 -1.208658942

V3 effect     0.94922501  1.241333  36 -2.7520485  4.650498481

V4 effect     2.91885941  1.241333  36 -0.7824141  6.620132884

V5 effect     1.06846240  1.241333  36 -2.6328111  4.769735868

V6 effect    -0.02999101  1.241333  36 -3.7312645  3.671282463

V7 effect     1.43830999  1.241333  36 -2.2629635  5.139583455

V8 effect     3.90280082  1.241333  36  0.2015274  7.604074292

V9 effect    -0.32549601  1.241333  36 -4.0267695  3.375777455

Confidence level used: 0.95

Conf-level adjustment: Sidak method for 10 estimates

```

Fig. 3.21 Confidence intervals for variety effects

test shown in Fig. 3.22 indicates that varieties V4 and V8 are significantly different from variety V10. However, an examination of the confidence intervals associated with these two varieties reveals a high level of uncertainty on the true value of the difference between varieties V4 and V8 on the one hand, and variety V10 on the other.

```

# Random effects
ranef(res.lmer)

```

The `ranef` function of the `lme4` library can be used to display random effects. Figure 3.24 shows the estimated random effects of the experiments. These effects are assumed to follow a normal distribution, to have the same variance and to be independent. For example, from Fig. 3.24, we can see that in experiment e1, mean yield was 4.81 q/ha lower than the mean for the population of environments, or that there was a difference in yield of 18.22 q/ha between experiments e2 and e3.

In summary, we can consider the model used to be appropriate for describing the results for the “wheat” example, because there is no evidence to call its validity into question. We can, therefore, use this model to interpret the data. The objective of the analysis is, above all, to compare the mean values of the varieties, and the results of

Fig. 3.22 Comparisons with the control variety V10: Dunnett's test

contrast	estimate	SE	df	t.ratio	p.value
V1 - V10	2.383233	1.85047	36	1.288	0.6974
V2 - V10	-1.212197	1.85047	36	-0.655	0.9612
V3 - V10	4.646961	1.85047	36	2.511	0.1065
V4 - V10	6.616595	1.85047	36	3.576	0.0079
V5 - V10	4.766198	1.85047	36	2.576	0.0927
V6 - V10	3.667745	1.85047	36	1.982	0.2904
V7 - V10	5.136046	1.85047	36	2.776	0.0593
V8 - V10	7.600536	1.85047	36	4.107	0.0018
V9 - V10	3.372240	1.85047	36	1.822	0.3730

P value adjustment: dunnettx method for 9 tests

contrast	estimate	SE	df	lower.CL	upper.CL
V1 - V10	2.383233	1.85047	36	-2.8885363	7.655002
V2 - V10	-1.212197	1.85047	36	-6.4839661	4.059572
V3 - V10	4.646961	1.85047	36	-0.6248086	9.918730
V4 - V10	6.616595	1.85047	36	1.3448258	11.888364
V5 - V10	4.766198	1.85047	36	-0.5055713	10.037967
V6 - V10	3.667745	1.85047	36	-1.6040247	8.939514
V7 - V10	5.136046	1.85047	36	-0.1357237	10.407815
V8 - V10	7.600536	1.85047	36	2.3287672	12.872306
V9 - V10	3.372240	1.85047	36	-1.8995297	8.644009

Confidence level used: 0.95

Conf-level adjustment: dunnettx method for 9 estimates

Fig. 3.23 Comparisons with the control variety V10: confidence intervals

Fig. 3.24 Random effects of “experiment”

\$experimentation	
(Intercept)	
1	-4.818378
2	9.372484
3	-8.848200
4	-4.956037
5	9.250131

the F test indicate that, in general, the variety means are not identical. Multiple comparison tests can then be used to refine this overall conclusion. For example, comparisons with the control variety V10 show that the means of varieties V4 and V8 are significantly higher than the mean for variety V10. None of the varieties has a mean significantly lower than that of variety V10. The confidence intervals of the differences between the varieties and the control variety V10 are, however, relatively large, indicating that the precision of these comparisons is relatively low. Thus, the real difference between varieties V4 and V10, or, to a lesser extent, between varieties V8 and V10, might not be very relevant in practice. Similarly, we cannot rule out the possibility that there is, in reality, a large difference between variety V10 and the other varieties. A larger network, with a larger number of experiments, would be required to draw more precise conclusions. The problem of the dimensions of experimental networks is dealt with in detail in Chap. 5.

References

- Berger RL, Hsu JC (1996) Bioequivalence trials, intersection-union tests and equivalence confidence sets (with comments). *Stat Sci* 11:283–319
- Cochran WG, Cox GM (1957) *Experimental designs*. Wiley, New York, 611 p
- Committee for Proprietary Medicinal Products (2001) Points to consider on switching between superiority and non-inferiority. *Br J Clin Pharmacol* 52(3):223–228. <https://doi.org/10.1046/j.0306-5251.2001.01397-3.x>
- Dagnelie P (1981) *Principes d’expérimentation*. Presses Agronomiques, Gembloux, 182 p
- Dagnelie P (1992) *Statistique Théorique et Appliquée. Tome 1: les bases théoriques*. Les Presses Agronomiques de Gembloux, Belgique, 492 p
- Dagnelie P (1998) *Statistique Théorique et Appliquée. Tome 2 : inférence statistique à une et à deux dimensions*. De Boeck & Larcier, Paris et Bruxelles, 659 p
- Galecki A, Burzykowski T (2013) *Linear mixed-effects models using R*. Springer, New York, 542 p
- Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. *Biom J* 50(3):346–363

- Jaykaran DS, Yadav P, Kantharia ND (2011) Nonsignificant P values cannot prove null hypothesis: absence of evidence is not evidence of absence. *J Pharm Bioallied Sci* 3(3):465–466. <https://doi.org/10.4103/0975-7406.84470>
- Kempton RA (1981) The use of neighbouring plot values in the analysis of variety trials. *J R Stat Soc Ser C Appl Stat* 30:59–70
- Lew MJ (2013) To P or not to P : on the evidential nature of P -values and their place in scientific inference, 31 p. arXiv:1311.0081v1 [stat.ME]
- Littell RC, Milliken GA, Stroup WW, Wolfinger RD (1996) SAS system for mixed models. SAS Institute Inc., Cary, 633 p
- Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O (2006) SAS system for mixed models, 2nd edn. SAS Institute Inc., Cary, 814 p
- Madden LV, Shah DA, Esker PD (2015) Does the P value have a future in plant pathology? *Phytopathology* 105(11):1400–1407. <https://doi.org/10.1094/PHYTO-07-15-0165-LE>
- Morey R, Hoekstra R, Rouder J, Lee M, Wagenmakers E-J (2015) The fallacy of placing confidence in confidence intervals. *Psychon Bull Rev*:1–21. <https://doi.org/10.3758/s13423-015-0947-8>
- Piepho H-P, Möhring J, Melchinger AE (2008) BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161:209–228
- Robinson AP, Froese RE (2004) Model validation using equivalence tests. *Ecol Model* 176:349–358
- SAS Institute Inc (1989) SAS/STAT user's guide, version 6, vol 2, 4th edn. SAS Institute Inc., Cary, 846 p
- Schuirmann D (1987) A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm* 15 (6):657–680
- Tukey JW (1991) The philosophy of multiple comparisons. *Stat Sci* 6(1):100–116
- Venables WN, Ripley BD (1998) Modern applied statistics with S-PLUS. Springer, New York, 548 p
- Zimmerman DL (1991) A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics* 47:223–239

Chapter 4

Advanced Methods for Network Analysis



4.1 Analysis of Mean Data

In analyses of experimental networks, it is always advisable to begin by analyzing the various individual experiments separately. In an experimental network, different experimental designs may be used in different experiments. In such cases, the model for the analysis of individual experiments should take into account the specific experimental design for each experiment.

The models used to analyze individual experiments may also differ between experiments, even in cases in which the experimental designs used are identical (e.g. for a randomized complete block design, such as that for the “wheat” example in Chap. 3). Indeed, in some experiments, the initial experimental design chosen to control for the experimental variability may prove unsuitable to describe the experimental variability actually observed. For example, in experiments with a randomized complete block design, a fertility gradient is often observed within blocks during analyses of the results (Monod 2001). Another example would be an experiment in which damage occurs due to the passage of game animals or a hailstorm. These events may occur in a highly sporadic manner during an experiment and may thus increase the uncontrolled experimental variability. It is sometimes possible to take these unpredicted variability factors into account by adapting the statistical model used. Spatial analysis and analysis of covariance are frequently used for this purpose (Gilmour 2000; Kempton and Fox 1997). These unpredicted variability factors, which may occur in some experiments but not in others, and which may differ between experiments when they do occur, may necessitate the use of different analysis models for different individual experiments.

In the two situations described above (mixture of experimental designs and/or mixture of analysis models for individual experiments), the analysis of experimental networks based on individual data becomes highly complicated. In such situations, it is therefore simpler to perform an experimental network analysis based on the mean treatment values calculated in each experiment. This analysis of mean data is

performed in two steps: in the first, the treatment means are calculated for each experiment, and in the second, these mean data are analyzed. This two-step analysis is generally simpler, but it may perform less well than a single-step analysis based on individual values (Smith et al. 2001a).

4.1.1 Step 1: Analysis of Individual Experiments to Estimate Treatment Means

The objective of this first step is to validate each experiment from an agronomic and statistical viewpoint and to calculate, for each experiment, the means of the treatments studied and the precision of these means.

The agronomic validation involves checking that the experimental protocol has been followed correctly and that the experiment can be used to address the questions posed in the protocol.

The statistical analysis of individual experiments can be used for statistical validation of the experiment and calculation of the means of the treatments studied and the precision of these means. The statistical validation of the experiment is based, in particular, on an examination of the residuals of the model. In particular, suspect residuals potentially indicating an incident occurring in one or several experimental units are sought, and the precision of the experiments is assessed, as this is an important element for validation.

In certain cases, a network may be constituted from experiments performed without repetitions across experiments. In this case, the statistical analysis of data by experiment is not possible. As no analysis of the individual experiments is feasible, it is not possible to validate the data statistically or to calculate their precision.

4.1.2 Step 2: Analysis of Mean Data

Step 2 is the analysis of the mean data obtained in step 1. In networks of experiments without repetition across experiments, the elementary data are used directly. In individual experiments leading to the calculation of adjusted means, it is these means that are used in this second step.

The following mixed model is widely used for the analysis of mean data from experimental networks (Piepho 1996):

$$y_{ij} = \mu + t_i + E_j^* + \varepsilon_{ij}^* \quad (4.1)$$

where y_{ij} is the mean value of the variable of interest for treatment i ($i = 1, \dots, I$) in experiment j ($j = 1, \dots, J$) and $\mu, t_i, E_j^*, \varepsilon_{ij}^*$ are the general mean, the effect of treatment i , the effect of experiment j and the residual associated with y_{ij} .

The observed experiments are assumed to constitute a random sample from a population of possible experiments. This implies that E_j^* is a random effect.

The random effects E_j^* and ε_{ij}^* are assumed to follow a normal distribution with a mean of zero and a variance of $\sigma_{E^*}^2$ and $\sigma_{\varepsilon^*}^2$, respectively. Furthermore, the different random effects are assumed to be independent of each other.

4.1.3 Example

We will now come back to the “wheat” example presented in Chap. 3, which we will analyze in two steps. Just to recap, this example concerns a network of experiments with a randomized complete block design. There are three blocks per experiment and the dataset is complete.

4.1.3.1 Step 1: Analysis of the Individual Experiments

The first step is the calculation of treatment means for each experiment and estimation of the precision of these means.

In Chap. 3, we presented an analysis of a network of experiments with a randomized complete block design in which the “block” effect was considered to be a random effect. This approach can be justified by considering the observed blocks to be representative of the plot or site at which the experiment is performed. However, in individual experiments, analyses are generally performed assuming the “block” effect to be fixed. This approach is recommended when the number of blocks per experiment is small (fewer than five or 10), which is often the case in practice. Indeed, in such cases, estimations of variance between blocks are highly imprecise if the “block” effect is considered to be random, with estimated between-block variances of zero frequently obtained. In these conditions, mixed-model estimation by the REML approach is not optimal, and it is preferable to use a model in which the “block” effect is considered to be fixed (Piepho et al. 2003). This does not modify the variance of the differences between treatments and, thus, has no effect on comparisons between treatments, which is generally the objective of the experiment (Spike et al. 2005). The model classically used for randomized complete block experiments is thus:

$$y_{ik} = \mu + t_i + b_k + \varepsilon_{ik} \quad (4.2)$$

where y_{ik} is the observed value for the variable of interest for treatment i ($i = 1, \dots, I$) in block k ($k = 1, \dots, K$) and $\mu, t_i, b_k,$ and ε_{ik} are the general mean, the effect of treatment i , the effect of block k , and the residual associated with y_{ik} , respectively. The random effects ε_{ik} are assumed to be independent and to follow a normal

Table 4.1 Mean data for the “wheat” example

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
E1	72.3	68.5	77.5	77.6	78.6	74.6	72.4	80.7	79.2	71.9
E2	89.7	87.4	89.7	93.7	86.3	90.9	95.7	94.0	85.8	86.6
E3	72.2	66.4	72.9	73.4	76.2	70.7	72.8	70.4	69.7	66.9
E4	71.8	67.5	77.4	77.7	78.7	74.3	72.1	80.7	80.4	71.2
E5	88.9	87.1	88.9	93.7	87.1	91.0	95.7	95.2	84.7	86.3

The rows correspond to experiments and the columns to varieties, and individual cells contain the mean yield, in q/ha, of the varieties in the experiments. Each mean is calculated from three observations

distribution with a mean of zero and a variance of σ_e^2 . In model (4.2), the estimators of treatment means have the same variance and are independent. The variance of the treatment means is given by $\frac{\sigma_e^2}{K}$. $\frac{\sigma_e^2}{K}$ is the effective variance of the treatment means, i.e. the variance used for comparisons between treatments.

In the case of the “wheat” example, the dataset is complete and the treatment means for each experiment are simply the arithmetic means of the three replicates. These mean data are summarized in Table 4.1. The residual variance estimates obtained following the adjustment of model (4.2) are 7.095, 5.658, 4.758, 7.273, and 5.733, for experiments 1 to 5, respectively. If we consider the theoretical residual variances of the different experiments to be identical, we can easily estimate this residual variance for the “wheat” example, as the dataset is complete and the number of blocks is identical for all experiments. Indeed, in this case, the estimate of the common residual variance is given by $\hat{\sigma}_e^2 = \frac{\sum_{j=1}^J MSE_j}{J}$, where MSE_j is the estimated residual variance of experiment j and J is the number of experiments. In the case of the “wheat” example, we have: $\frac{\sum_{j=1}^J MSE_j}{J} = \frac{7.095+5.658+4.758+7.273+5.733}{5} = 6.1034$. This value is identical to that obtained in the single-step analysis of individual data (Fig. 3.1). Still assuming that all the experiments have the same theoretical residual variance, the variance of the treatment means of experiment j can be estimated as: $\frac{\hat{\sigma}_e^2}{K_j}$, where K_j is the number of blocks in experiment j . In the case of the “wheat” example, the number of blocks per experiment K is 3 and the variance of the treatment means is identical in all experiments and is estimated by: $\frac{\hat{\sigma}_e^2}{K} = \frac{\sum_{j=1}^J MSE_j}{K} = \frac{6.1034}{3} = 2.0345$. This variance represents the uncertainty on the values presented in Table 4.1 resulting from these data being estimates rather than the true treatment means of the experiments.

4.1.3.2 Step 2: Analysis of Mean Data

We illustrate here the difference between the one-step analysis of individual data and the two-step analysis by presenting a few results of the analysis of mean data for the “wheat” example.

<i>Variance component</i>					
Group	Variance				
Experiment	79.0165				
Residual	8.5606				
<i>Comparison of two treatments</i>					
contrast	estimate	SE	df	lower.CL	upper.CL
V1 - V10	2.3832329	1.85047	36	-1.3696951	6.13616080

Fig. 4.1 Analysis of the “wheat” example based on mean data, with mixed model (4.1)

The first part of Fig. 4.1 shows the table of variance components for model (4.1) for the “wheat” dataset.

The between-experiment variance estimated from mean data is slightly higher than that obtained for the analysis of individual data presented in Fig. 3.1, because it also includes between-block variability. Indeed, in the analysis of mean data with model (4.1), the random effect E_j^* includes the effect of experiment, but also, as the analysis is based on the mean data for K blocks, the (mean) block effects. The between-experiment variance of the model based on mean data is, thus, equal to the sum of the between-experiment variance of the model for individual data plus the between-block variance, divided by the number of blocks per experiment:

$$\sigma_{E^*}^2 = \sigma_E^2 + \frac{\sigma_B^2}{K} \tag{4.3}$$

where $\sigma_{E^*}^2$, σ_E^2 , σ_B^2 and K are the between-experiment variance for the model based on mean data, the between-experiment variance for the model based on individual data, the between-block variance for the model based on individual data and the number of blocks per experiment. We can obtain, within the range of rounding errors, the value of the between-experiment variance indicated in Fig. 4.1, by replacing the variances σ_E^2 and σ_B^2 in formula 4.3 with the estimated values shown in Fig. 3.1: $79.0164 = 77.3305 + 5.0578/3$.

The residual variance estimated from the mean data is different from that obtained from the analysis of individual data presented in Fig. 3.1. Indeed, ε_{ij}^* , the residual associated with the mean y_{ij} , is the sum of two combined effects, an effect of the interaction between treatment i and experiment j , and an estimation error on the mean of treatment i in experiment j due to experimental errors (Frensham et al. 1997; Smith et al. 2001b). The effects of the interaction between treatments and experiments are assumed to be independent of the estimation errors on the mean, and the variance of the residuals.

ε_{ij}^* is thus equal to the sum of the variance of the interaction between treatments and experiments and the variance of the estimation errors for the mean data. The variance of the estimation errors on the mean data is given by $\frac{\sigma_\varepsilon^2}{K}$, where σ_ε^2 is the variance of the experimental errors for individual data and K is the number of blocks per experiment. The variance of the residues ε_{ij}^* is therefore:

$$\sigma_{\varepsilon^*}^2 = \sigma_{tE}^2 + \frac{\sigma_\varepsilon^2}{K} \quad (4.4)$$

We can obtain the value for residual variance indicated in Fig. 4.1, by replacing the variances σ_{tE}^2 and σ_ε^2 in Eq. 4.4 with the estimated values from Fig. 3.1: $8.5606 = 6.5261 + 6.1034/3$.

The variance of the means in Eq. (4.4) is the effective variance, that used for comparisons between means within an experiment. There is no between-block variance in Eq. (4.4) because, in the analysis of mean data, the block means are confounded with the effect of experiment.

The second part of Fig. 4.1 shows the result of the comparison of the mean of variety V1 with that of variety V10. This result is identical to that obtained with mixed model (3.1) applied to the individual data. Indeed, in model (4.1), the variance of the difference between the two treatment means is given by:

$$\text{Var}(\hat{\mu}_i - \hat{\mu}_{i'}) = 2 \frac{\sigma_{\varepsilon^*}^2}{J} \quad (4.5)$$

where $\text{Var}(\hat{\mu}_i - \hat{\mu}_{i'})$ is the variance of the estimator of the difference between treatment means i and i' , and $\sigma_{\varepsilon^*}^2$ and J are the residual variance of model (4.1) and the number of experiments, respectively.

By replacing the residual variance $\sigma_{\varepsilon^*}^2$ in Eq. (4.5) with its expression in Eq. (4.4), we obtain:

$$\text{Var}(\hat{\mu}_i - \hat{\mu}_{i'}) = 2 \left(\frac{\sigma_{tE}^2}{J} + \frac{\sigma_\varepsilon^2}{JK} \right)$$

corresponding to Eq. (3.2), which gives the variance of the difference between two treatment means obtained by the analysis of individual data with model (3.1).

In the case of a network of experiments with a complete block design and identical numbers of blocks in each experiment, the two-step analysis with model (4.1) is equivalent to a one-step analysis on the individual data with model (3.1). Comparisons between treatments are performed in the same manner whether the individual or mean data are used.

In the case of experimental networks with a complete block design in which the individual experiments are analyzed with model (4.2), the two-step analysis with model (4.1) is not really justified and it is simpler to analyze the network based on the individual data. This analysis of individual data is even preferable in cases in which

the number of blocks per experiment varies between experiments. The two-step analysis with model (4.1) is used mostly in cases in which the network includes a mixture of experimental designs and/or a mixture of analysis models for individual experiments. However, in such cases, the results of an analysis of mean data with model (4.1) are generally different from those obtained in analyses of individual data.

4.1.4 A Variant: Analysis of Mean Data with a Fixed-Effects Model

The analysis of mean data for experimental networks is usually based on mixed model (4.1), but the following fixed-effects model can also be used:

$$y_{ij} = \mu + t_i + e_j + \varepsilon_{ij}^* \quad (4.6)$$

where y_{ij} is the mean value for the variable of interest for treatment i ($i = 1, \dots, I$) in experiment j ($j = 1, \dots, J$) and μ , t_i , e_j , and ε_{ij}^* are the general mean, the effect of treatment i , the effect of experiment j and the residual associated with y_{ij} , respectively.

The random effect ε_{ij}^* is assumed to follow a normal distribution with a mean of zero and a variance of $\sigma_{\varepsilon^*}^2$. These residual random effects are also assumed to be independent of each other.

In a complete network, the estimated treatment means obtained with model (4.6) are identical to those obtained with mixed model (4.1). Similarly, the variance of the differences between treatments, which is not dependent on the differences between experiments, as shown by Eq. (4.5), is identical for the two models. In an experimental network, the principal objective is generally to estimate the differences between treatment means. The conclusions obtained are, thus, the same with a mixed model or a fixed-effects model. It should be noted that, for analyses of mean data, the analysis based on a fixed-effects model does not change the scope of the conclusions, which still relate to a population of environments.

Figure 4.2 summarizes the results of the analysis of mean data with a fixed-effects model. The only difference with respect to the results of the analysis of mean data with a mixed model (Fig. 4.1) is the absence of a variance component for the variable “experiment”.

4.1.5 Estimation of Variance for the Treatment \times Experiment Interaction

In analyses of mean data with model (4.1), ε_{ij}^* , the residual associated with the mean y_{ij} , is the sum of two combined effects, an effect of the interaction between treatment

<i>Variance component</i>					
Group	Variance				
Residual	8.5606				
<i>Comparison of two treatments</i>					
contrast	estimate	SE	df	lower.CL	upper.CL
V1 - V10	2.3832329	1.85047	36	-1.3696951	6.13616080

Fig. 4.2 Analysis of the “wheat” example, for mean data and a fixed-effects model

i and experiment j , and the estimation error for the mean of treatment i in experiment j . Estimation of the variance of the treatment \times experiment interaction requires knowledge of the variance of estimation errors for the mean data.

4.1.5.1 Calculation of the Variance of Mean Data

The effective variance of the mean data for each experiment can be estimated with the following formula:

$$Vmean_j = 0.5 VMC_j \quad (4.7)$$

where VMC_j is an estimate of the mean variance for pairwise comparisons of treatments for experiment j .

An estimate of the common variance of mean data for the experimental network can be obtained by calculating the mean of the variances of mean data by experiment, as follows:

$$Vmean = \frac{\sum_{j=1}^J Vmean_j}{J} \quad (4.8)$$

4.1.5.2 Estimation of the Variance of the Treatment \times Experiment Interaction

The variance of the treatment \times experiment interaction can be obtained by calculating the following difference:

$$\hat{\sigma}_{IE}^2 = \hat{\sigma}_{e^*}^2 - Vmean \quad (4.9)$$

where $\hat{\sigma}_{e^*}^2$ is the estimated residual variance of model (4.1).

4.1.5.3 Case of a Network of Randomized Complete Block Experiments

In the case of an experiment with a randomized complete block design analyzed with model (4.2), we have:

$VMC_j = 2\frac{MSE_j}{K_j}$, where MSE_j and K_j are the estimated residual variance and the number of blocks in experiment j , respectively. Equation (4.7) thus becomes: $Vmean_j = \frac{MSE_j}{K_j}$, which corresponds to an estimate of the effective variance of treatment means, i.e. the variance used in comparisons between treatments (see Sect. 4.1.3.1).

In the cases of networks of experiments with a randomized complete block design, Eq. (4.8) thus becomes: $Vmean = \frac{\sum_{j=1}^J \frac{MSE_j}{K_j}}{J}$, where MSE_j and K_j are the estimated residual variance and the number of blocks in experiment j , respectively. For the “wheat” example presented in Chap. 3, the estimated residual variances are 7.095, 5.658, 4.758, 7.273, and 5.733 for experiments 1 to 5, respectively. For a constant number of three blocks per experiment we therefore obtain: $Vmean = \frac{(7.095+5.658+4.758+7.273+5.733)/3}{5} = 2.0345$. This value is identical to that obtained in Sect. 4.1.3.1. The estimated residual variance of the model for mean data, $\hat{\sigma}_{e^*}^2$, is 8.5606 (Fig. 4.1). The estimated variance of the treatment \times experiment interaction is, therefore, $8.5606 - 2.0345 = 6.5261$. This value is identical to that obtained in the analysis of individual data (Fig. 3.1).

For a network of experiments with a randomized complete block design and identical numbers of blocks in each experiment, the Eqs. (4.7, 4.8 and 4.9) yield an estimate of the variance of the treatment \times experiment interaction identical to that obtained in a one-step analysis of individual data with model (3.1). This is a particular case and, generally, the values obtained with Eqs. (4.7, 4.8 and 4.9) are approximations of the value obtained in a one-step analysis of individual data.

4.1.6 R Script

We present here a script for the analysis of an experimental network in two steps. The script presented concerns only the second step of this analysis, the analysis of mean data. Most of this script is identical to that presented in Chap. 3 for the analysis of individual data. The only differences concern the rubrics relating to model adjustment and calculation of the variance of the treatment \times experiment interaction. The outputs of this script are described in detail in Chaps. 3 and 4 and are therefore not reported here.

```

# R version 3.4.3
# loading of the packages
library(lme4) # version 1.1-14
library(emmeans) # version 1.0
library(car) # version 2.1-6
library(outliers) # Version: 0.14

# adjustment of the model
res.lmer <- lmer(rendement ~ variete + (1|experimentation),
data=DFmoy, na.action=na.exclude)
res.lmer

# validation of the model
plot(fitted(res.lmer), residuals(res.lmer), abline(h=0))
hist(residuals(res.lmer))

DFmoy$residus <- residuals(res.lmer)
grubbs.test(DFmoy$residus)
DFmoy[which.max(DFmoy$residus),]

# ANOVA table
Anova(res.lmer, test.statistic="F")

# Adjusted means
moy_var <- lsmeans(res.lmer, ~variete)
moy_var

# Pairwise comparisons
pairs(moy_var, adjust="tukey")
cld(moy_var, Letters=c(LETTERS))

# Comparisons with the general mean
contrast(moy_var, method="eff", adjust="sidak")
confint(contrast(moy_var, method="eff", adjust="sidak"))

# Comparison with a control
contrast(moy_var, method="trt.vs.ctrl", ref=2)
confint(contrast(moy_var, method="trt.vs.ctrl", ref=2))

# Random effects
ranef(res.lmer)

# Variance of the means and variance of the interaction
Vmoy <- 0.5*mean(VMCj) # VMCj = VMC vector per experiment
Vmoy

var_intera <- summary(res.lmer)$sigma^2 - Vmoy
var_intera

```

4.2 Experiments with Heterogeneous Variances

4.2.1 Introduction

The model (4.1) used to analyze mean data considers the means of all experiments to have the same precision. Indeed, in the two-step analysis, one of the assumptions of model (4.1) is that the residuals of the model have the same variance. The residuals of this model are the sum of an effect of the interaction between treatment i and experiment j and the estimation error on the mean of treatment i in experiment j due to experimental errors. If we assume that the effects of the treatment \times experimentation interaction are independent of the estimation errors on the means, this implies that we also assume that the variance of the errors of estimation on the means is constant.

In a network of experiments with a randomized complete block design, the treatment means estimated in each experiment have the same precision between experiments if the variance of the experimental errors on individual data σ_ϵ^2 is identical for all experiments and if the number of blocks K is also identical in all experiments (actually, for the variance of the means to remain constant, all that is required is that $\frac{\sigma_\epsilon^2}{K}$ remains constant).

The assumption that the variance of the estimation errors on mean data is identical for all the experiments in a network is not always realistic. Indeed, we can envisage situations in which the variance of the experimental errors on individual data σ_ϵ^2 varies between experiments. For example, a field experiment in a deep soil might be expected to yield more precise results than an experiment in a superficial soil, which would probably be more heterogeneous. In networks of experiments with a randomized complete block design, the number of blocks may also differ between experiments. Model (4.1) assumes that the means of all experiments have the same precision. This model thus accords equal weightings to all experiments. If one experiment is less precise than the others, or includes a smaller number of replicates, it should have a lower weighting than the other experiments in the calculation of treatment means. In this section, we consider the analysis of an experimental network based on mean data for which precision varies between experiments.

If, in a network of experiments with a randomized complete block design, the number of blocks K varies between experiments but the variance of the experimental errors on individual data σ_ϵ^2 is approximately constant, then model (3.1) can be used for the analysis of individual data. Indeed, differences in the number of blocks between experiments pose no problem in the analysis of individual data with the mixed model (3.1), because the effect of “block”, and thus of the number of blocks per experiment, is taken into account explicitly in the model.

4.2.2 Example “Wheat”

We will now come back to the “wheat” example presented in Chap. 3, for which we will perform a two-step analysis, by considering the variance of the means known in the second step.

4.2.2.1 Analysis of the “Wheat” Example, with Uniform Variances Between Experiments

For comparison of the results with those obtained in Chap. 3, we will first consider the variances of the means to be identical between experiments.

The variance of the means can be obtained with the following formula, initially presented in Sect. 4.1.5.3: $V_{mean} = \frac{\sum_{j=1}^J \frac{MSE_j}{K_j}}{J}$. In the “wheat” example from chapter three, the estimated residual variances are 7.095, 5.658, 4.758, 7.273, 5.733 for experiments 1 to 5, respectively. For a constant number of three blocks per experiment, we therefore obtain: $V_{mean} = \frac{(7.095+5.658+4.758+7.273+5.733)/3}{5} = 2.03447$.

R Script and Annotated Analysis

```
# R version 3.4.3
# loading of the packages
library(metafor) # Version : 2.0-0
```

The `metafor` library can be used to adjust a mixed model for mean data by assuming a known variance-covariance matrix for the residuals of the model, i.e., for an experimental network, the variance-covariance matrix of the treatment means by experiment.

```
V1 <- diag(2.03447, nrow(DFmoy))
```

`V1` is the variance-covariance matrix of the means. Here, we consider `V1` to be a scalar matrix such that $V1 = 2.03447 \mathbf{I}$, where \mathbf{I} is an identity matrix with a dimension equal to the number of means analyzed during the second step of the two-step analysis. For the “wheat” example, this number is 50.

```
res_metafor <- rma.mv(rendement ~ -1 + variete, V1, random=list(~
1|experimentation, ~1|variete:experimentation), rho=0,
data=DFmoy)
```

The `rma.mv` function can be used to adjust the mixed model. The first argument of the function describes the fixed part of the model. The “-1” eliminates the intercept of the model such that when there is a single fixed factor, as in this case, the estimates of the effects of each level of the factor correspond to the means for these levels. The second argument specifies the variance-covariance matrix of the means. The argument `random` describes the random part of the model. This random part of the model differs from model (4.1) in that it includes the variety \times experiment interaction. It is possible to estimate the variance of the variety \times treatment interaction directly because the variance-covariance matrix of the means is assumed to be known.

```
res_metafor$sigma2
```

The object `res_metafor` contains the results of the adjustment of the mixed model. The command `res_metafor$sigma2` can be used to recover the between-experiment variance, 79.0165. This value is identical to that in Fig. 4.1 obtained with model (4.1). This variance includes the between-experiment and between-block variances.

```
res_metafor$tau2
```

The command `res_metafor$tau2` can be used to recover the variance of the variety \times experiment interaction, 6.5261. This value is identical to that in Fig. 3.1 obtained with model (3.1).

```
res_metafor$b
```

The command `res_metafor$b` can be used to recover the fixed effects. In our case, the fixed effects correspond to the variety means, which are shown in Fig. 4.3. These means are identical to those shown in Fig. 3.17 obtained with model (3.1).

```
# contrast
var1 <- c(1,0,0,0,0,0,0,0,0,0)
var10 <- c(0,1,0,0,0,0,0,0,0,0)
var1moinsvar10 <- var1 - var10
covb <- vcov(res_metafor)
theta <- res_metafor$b
estimate <- var1moinsvar10%*%theta
stderr.est <- sqrt(t(var1moinsvar10)%*%covb%*%var1moinsvar10)
estimate
stderr.est
```

Fig. 4.3 Variety means obtained with homogeneous variances for mean data

```
V1 78.99116
V10 76.60792
V2 75.39573
V3 81.25488
V4 83.22452
V5 81.37412
V6 80.27567
V7 81.74397
V8 84.20846
varieteV9 79.98016
```

Comparisons between particular varieties can be performed with the commands shown in the example above for the comparison of varieties V1 and V10. The estimated difference between these two varieties is 2.383233 and the standard error of this estimate is 1.85047. These values are identical to those in Fig. 4.1 obtained with model (4.1).

4.2.2.2 Analysis of the “Wheat” Example with Heterogeneous Variances Between Experiments

We will now consider the variances of the means to be different between experiments. These variances are calculated with Eqn. (4.7): $V_{mean_j} = 0.5 VMC_j$. For a network of experiments with a randomized complete block design, $VMC_j = 2 \frac{MSE_j}{K_j}$, where MSE_j and K_j are the estimated residual variance and the number of blocks in experiment j , respectively. In the “wheat” example from Chap. 3, the estimated residual variances are 7.095, 5.658, 4.758, 7.273, and 5.733 for experiments 1 to 5, respectively. Thus, for a constant number of three blocks per experiment, we obtain the following variances for experiments 1 to 5, respectively: 2.365, 1.886, 1.586, 2.424, 1.911.

R Script and Annotated Analysis

```
# R version 3.4.3
# loading of packages
```

(continued)

```
library(metafor) # Version : 2.0-0
Vmoyj <- c(2.365, 1.886, 1.586, 2.424, 1.911)
V2 <- diag(rep(Vmoyj, each=10)) # 10 = number of treatments
```

V2 is the variance-covariance matrix of the means. Here, we consider V2 to be a diagonal matrix with a dimension equal to the number of means analyzed during the second step of the two-step analysis. For the “wheat” example, this number is 50. The values of this diagonal matrix are given by the vector `Vmoyj`, based on the estimated variances of the means in each experiment.

```
res_metafor2 <- rma.mv(rendement ~ -1 + variete, V2, random=list(~
1|experimentation, ~ 1|variete:experimentation), rho=0,
data=DFmoy)
```

The function `rma.mv` can be used to adjust the mixed model.

```
res_metafor2$b
```

The command `res_metafor2$b` can be used to recover the fixed effects, which, in this example, correspond to the variety means. Figure 4.4 shows these means and Fig. 4.5 compares them with the means obtained with the model assuming the variance of the mean data to be identical between experiments. For this example, there is little difference between the two sets of means. The use of heterogeneous variances does not modify the conclusions of the analysis.

4.2.3 Going Further...

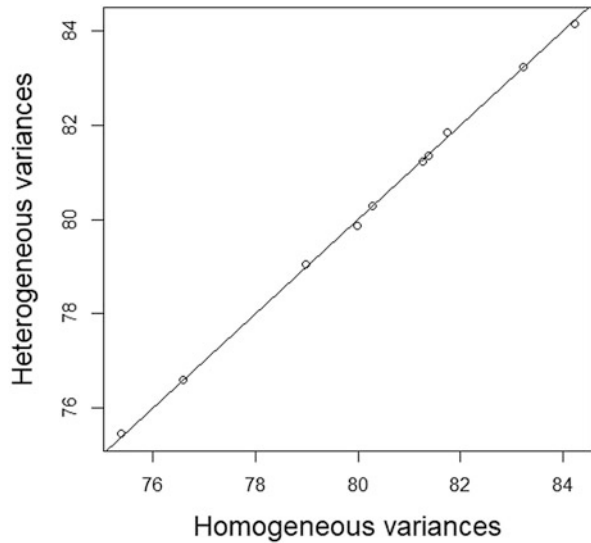
4.2.3.1 Variance-Covariance Matrix for Treatment Means

Generally, in experimental networks, the variance-covariance matrix V for the treatment means analyzed in the second step of a two-step analysis is a block diagonal matrix (Möhring and Piepho 2009). The diagonal of this matrix consists of as many blocks as there are experiments in the network, each block being the variance-covariance matrix of the treatment means for the corresponding experiment. Outside of the diagonal are zeros, the experimental errors of the different experiments being considered independent. Different variance-covariance matrices for mean data can be envisaged for the second step in two-step analyses of mean data. We present here three variance-covariance models, but others could be used (Möhring and Piepho 2009).

Fig. 4.4 Variety means obtained with heterogeneous variances for mean data

V1	79.05309
V10	76.60438
V2	75.45154
V3	81.23428
V4	83.23115
V5	81.35333
V6	80.29065
V7	81.84678
V8	84.13694
V9	79.85983

Fig. 4.5 Comparison of the variety means estimated with a model assuming homogeneous variances and those obtained with a model assuming heterogeneous variances between experiments. Each point corresponds to a variety and the straight line corresponds to the first bisector



$R_{us,j}$ Matrix

The most general form for the variance-covariance matrix for mean data is a block diagonal matrix $V = R_{us,j}$, in which each block is an R_{us} matrix in which R_{us} is any type of variance-covariance matrix. An example of this type of variance-covariance matrix is shown in Fig. 4.6 for a network of four experiments with three treatments.

Variance-covariance matrices of this type are suitable, in particular, for experimental networks in which the individual experiments are analyzed with a spatial

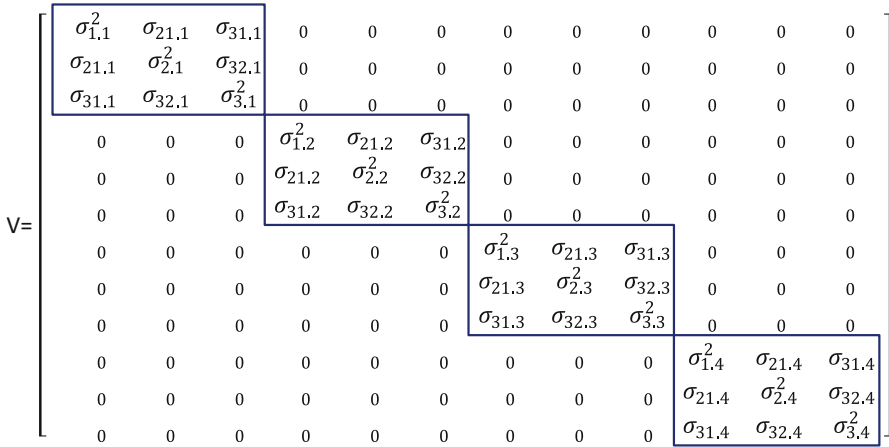
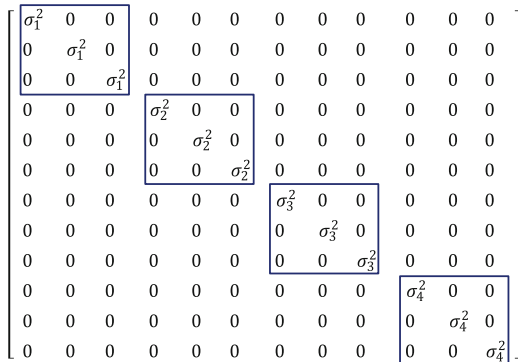


Fig. 4.6 Example of a block diagonal variance-covariance matrix $R_{us,j}$. The rectangles correspond to the different experiments

Fig. 4.7 Example of a block diagonal variance-covariance matrix $R_{id,j}$. The rectangles correspond to different experiments



analysis model or a covariance analysis model (Gilmour 2000; Kempton and Fox 1997). Indeed, in these situations, the analysis of individual experiments results in estimators of treatment means that are not necessarily independent or have the same variance.

$R_{id,j}$ Matrix

A second type of variance-covariance matrix is the block diagonal matrix $R_{id,j}$ in which each block is an $R_j = 0.5 VMC_j I$ matrix, where I is an identity matrix with a dimension equal to the number of treatments, and VMC_j is the mean variance of pairwise comparisons of treatments in experiment j . An example of this type of variance-covariance matrix is shown in Fig. 4.7, for a network of four experiments with three treatments.

Fig. 4.8 Example of a block diagonal variance-covariance matrix R_{id} . The rectangles correspond to different experiments

$$\frac{\overline{VMC}_j}{2} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

An $R_{id, j}$ matrix is suitable for networks of experiments with a randomized complete block design analyzed with model (4.2), for which the experimental error on individual data and/or the number of blocks differs between experiments. In this situation, this type of variance-covariance matrix provides an exact calculation of comparisons between treatments (Möhring and Piepho 2009). This is the variance-covariance matrix used in Sect. 4.2.2.2 for analysis of the “wheat” example.

R_{id} Matrix

The third example of a variance-covariance matrix is that of the R_{id} matrix, in which $R_{id} = 0.5 \overline{VMC}_j \mathbf{I}$, \mathbf{I} is an identity matrix with a dimension equal to the number of treatments multiplied by the number of experiments, and \overline{VMC}_j is the mean for all experiments of the mean variances of pairwise comparisons of treatments for each experiment j . An example of this type of variance-covariance matrix is shown in Fig. 4.8 for a network of four experiments with three treatments.

This type of matrix is the type assumed in model (4.1), in which the variance of the experimental error associated with the mean data is assumed to be the same in all the experiments of the network. This type of model is suitable for networks of experiments with a randomized complete block design analyzed with model (4.2), for which the variance of the experimental errors on individual data σ_e^2 and the number of blocks K are the same in all experiments. In this situation, this type of variance-covariance matrix provides an exact calculation of comparisons between treatments. This is the type of variance-covariance matrix used in Sect. 4.2.2.1 for the analysis of the “wheat” example.

4.2.3.2 Choice of a Variance-Covariance Matrix for Treatment Means

The mixed model can take into account the variance-covariance matrix of treatment means in the two-step analysis of mean data.

The choice of a variance-covariance matrix for the means analyzed in the second step of the analysis of an experimental network depends on the experimental design and the method of analysis used to analyze the results of individual experiments.

However, in situations in which the variance-covariance matrix used for the means should, in theory, be an $\mathbf{R}_{us.j}$ matrix, the two-step analysis of the experimental network is not much simpler than a single-step analysis, and a simpler variance-covariance matrix, such as an $\mathbf{R}_{id.j}$ or \mathbf{R}_{id} matrix, is frequently used to facilitate calculations. For further information on this subject, see Möhring and Piepho (2009).

4.2.3.3 Variance-Covariance Matrix for Random Effects

Different types of variance-covariance matrix can be considered for random effects other than residual errors: “environment” effects, and effects of “environment x studied factor” interaction. For example, certain authors have suggested that varietal experimental networks should be analyzed with a model considering the variance of the variety \times experiment interaction to be non-constant and a function of covariates (Frensham et al. 1997).

4.3 Missing Data

Experimental networks are often incomplete, with some treatments absent from some experiments. For example, in an evaluation of wheat varieties performed over several years, the varieties tested may differ from year to year, because new varieties emerge each year. This situation results in an incomplete data table with a certain number of missing data. In this case, comparisons of the treatments studied are more complicated. Indeed, when two treatments are tested in the same experiment, their effects can be compared directly. By contrast, the comparison of two treatments present in different experiments is indirect, requiring the comparison of each treatment with other treatments present in both experiments. In practice, the comparison of two treatments may consist of a combination of direct and indirect comparisons. The analysis of a table of incomplete data involves the calculation of adjusted means. However, before analyzing such data, it is important to consider the reasons for the absence of certain data.

4.3.1 Origin of Missing Data

Missing data may follow any of the following three patterns:

- MCAR (missing completely at random): the missing data are completely independent of observed or missing values. The missing data can thus be considered a random sample of the entire set of data.

- MAR (missing at random): the missing data are a function of the observed values but not of the missing values. Conditional on the observed values, the missing data are random.
- MNAR (missing not at random): the missing data are a function of the missing values. The values of the non-observed data are the cause of the missing data.

These different situations can be illustrated with the example of a network of experiments evaluating wheat varieties.

- For MCAR, a list of n varieties is tested at p different sites. Purely by chance, the seeds of one or several varieties are not delivered to certain sites in time for the experiment. The resulting missing data are MCAR because we can assume that the missing data are independent of variety and site.
- For MAR, let us consider the case of a network running over several years. Each year, the 50% of varieties giving the best results are maintained for testing in the next year. The other varieties are replaced by new varieties. The worst varieties are rapidly excluded from the network, whereas the best varieties are tested over longer periods to confirm the good results obtained in the first few years. The missing data may be considered MAR, because they are functions of the observed values but not of the (non-observed) missing values.
- For MNAR, we can consider a situation in which varieties are tested at p sites throughout France. Some varieties are tested at all sites, some only in the north and others only in the south. The varieties tested only in the north are late-flowering varieties, because it is known that such varieties generally perform less well in the south. By contrast, the varieties tested only in the south are earlier varieties, because such varieties are known to perform less well in the north. The varieties tested throughout the country generally behave well in all regions. The missing data are thus MNAR because they are a function of the (non-observed) missing values.

Experimental network data are classically analyzed with a mixed model, the parameters of which are generally estimated by the restricted maximum likelihood (REML) method (Pinheiro and Bates 2000). It has been shown that inferences based on the REML method are valid for missing data of the MCAR and MAR types (Piepho and Möhring 2006). Missing data of the MNAR type are more difficult to handle, because, in this case, the missing data are considered to be informative. The analysis thus requires a particular approach involving the modeling of the origin of the missing data.

4.3.2 *Adjusted Means*

In incomplete networks, treatments cannot be compared on the basis of the observed means. Indeed, as not all the treatments are observed in the same experiments, the observed differences between treatments may be confounded with the differences between experiments. Treatment comparisons therefore require the adjustment of treatment means for the effects of the experiments in which they were observed.

Table 4.2 Example of an incomplete data table

	e1	e2	e3
t1	72	86	85
t2	79	87	86
t3	82		90
t4		80	85
t5	76	82	

Table 4.3 Symbolic representation of the data for the example

	e1	e2	e3
t1	y_{11}	y_{12}	y_{13}
t2	y_{21}	y_{22}	y_{23}
t3	y_{31}		y_{33}
t4		y_{42}	y_{43}
t5	y_{51}	y_{52}	

The calculation of adjusted means is illustrated below with a simple example including five treatments and three experiments, for which treatment means are available for each experiment. The data for this example are provided in Table 4.2.

The data are represented in a symbolic manner by y_{ij} , where y is the observed value, the index i corresponds to the treatment and the index j corresponds to the experiment. The symbolic representation of the data for the example studied here is shown in Table 4.3.

For complete experimental networks, the analysis of mean data leads to estimates of treatment means and of variances of comparisons between treatments identical with a fixed-effects or mixed models (see Sect. 4.1.4). By contrast, the same cannot be said for incomplete experimental networks. In an incomplete experimental network, adjusted means are calculated differently for fixed-effects models and for mixed models, which assume the “experiment” effect to be random, and the variance of treatment comparisons is also generally different. The estimation of treatment means by mixed models is generally more precise than that achieved with fixed-effects models, because the mixed model allows the extraction of more information about the data than a fixed-effects model.

Experimental networks are generally analyzed with mixed models, but, for pedagogic reasons, we will first describe the calculation of adjusted means for a fixed-effects model.

4.3.2.1 Fixed-Effects Model

The calculation of adjusted means is based on model (4.6), which can be written as follows, omitting the “*” for the sake of clarity:

$$y_{ij} = \mu + t_i + e_j + \varepsilon_{ij}$$

Table 4.4 Adjusted means

Treatment	Adjusted mean
t1	81.00
t2	84.00
t3	87.21
t4	79.50
t5	80.79

where y_{ij} is the mean value of the quantity of interest for treatment i ($i = 1, \dots, 5$) in experiment j ($j = 1, \dots, 3$) and μ , t_i , e_j , and ε_{ij} are the general mean, the effect of treatment i , the effect of experiment j and the residual associated with y_{ij} , respectively.

The t_i and e_j effects are fixed and only the residual ε_{ij} is random. The residual ε_{ij} is assumed to follow a normal distribution with a mean of zero and a variance of σ_ε^2 . The residuals are also assumed to be independent.

The parameters of the model are estimated by the ordinary least squares method, which consists of using as estimates for unknown parameters the values that minimize the residuals sum of square. The adjusted means are then calculated with the model. For example, the theoretical mean of treatment t_3 , μ_3 , is defined as:

$$\mu_3 = \frac{(\mu + t_3 + e_1) + (\mu + t_3 + e_2) + (\mu + t_3 + e_3)}{3}$$

In the case of the example given here, by introducing the estimated values of the parameters into the equation, we obtain:

$$\hat{\mu}_3 = \frac{(82.5 + 4.71 - 6) + (82.5 + 4.71 + 2.43) + (82.5 + 4.71 + 3.57)}{3} = 87.2$$

Note that the calculation of the adjusted mean for treatment t_3 involves estimating the value for this treatment in all the experiments and then calculating the arithmetic mean of these estimates. The adjusted mean for treatment t_3 is, thus, the mean value for this treatment “as if” it had been observed in all the experiments.

The adjusted means of the five treatments, obtained as shown above, are presented in Table 4.4.

We can check that, for treatments t_1 and t_2 , which were observed in all the experiments, the adjusted means are identical to the unadjusted crude means.

Treatment Comparisons

When we wish to compare two treatments, which is the principal objective of most experimental networks, the estimation of the difference between the two treatments

combines a direct and an indirect comparison. For example, for treatments t_3 and t_4 , it is possible to estimate the difference between the two treatments on the basis of a direct comparison in experiment e_3 :

$$\delta_{direct} = y_{33} - y_{43}$$

By replacing the data by their expression in the model, we obtain:

$$\begin{aligned}\delta_{direct} &= \mu + t_3 + e_3 + \varepsilon_{33} - \mu - t_4 - e_3 - \varepsilon_{43} \\ &= t_3 - t_4 + \varepsilon_{33} - \varepsilon_{43}\end{aligned}$$

This comparison involves no confounding with an experiment effect, and its variance is $2\sigma_e^2$.

It is also possible to compare t_3 and t_4 indirectly, by comparing the difference between t_3 and t_2 in experiment e_1 with the difference between t_4 and t_2 in experiment e_2 :

$$\delta^1_{indirect} = (y_{31} - y_{21}) - (y_{42} - y_{22})$$

Replacing the data with their expression in the model, we obtain:

$$\begin{aligned}\delta^1_{indirect} &= \mu + t_3 + e_1 + \varepsilon_{31} - \mu - t_2 - e_1 - \varepsilon_{21} \\ &\quad - \mu - t_4 - e_2 - \varepsilon_{42} + \mu + t_2 + e_2 + \varepsilon_{22} \\ &= t_3 - t_4 + \varepsilon_{31} - \varepsilon_{21} - \varepsilon_{42} + \varepsilon_{22}\end{aligned}$$

In this case too, there is no confounding with an experiment effect. However, other indirect comparisons are also possible, such as comparing the difference between t_3 and t_1 in experiment e_1 with the difference between t_4 and t_1 in experiment e_2 , or the difference between t_3 and t_5 in experiment e_1 with the difference between t_4 and t_5 in experiment e_2 . It is also possible to perform a single indirect comparison relative to the mean of treatments t_1 , t_2 and t_5 :

$$\delta_{indirect} = \left(y_{31} - \frac{y_{11} + y_{21} + y_{51}}{3} \right) - \left(y_{42} - \frac{y_{12} + y_{22} + y_{52}}{3} \right)$$

By replacing the data with their expression in the model, we obtain:

$$\delta_{indirect} = t_3 - t_4 + \varepsilon_{31} - \varepsilon_{42} - \frac{\varepsilon_{11} + \varepsilon_{21} + \varepsilon_{51}}{3} + \frac{\varepsilon_{12} + \varepsilon_{22} + \varepsilon_{52}}{3}$$

Again, this comparison is not confounded with an experiment effect. The variance of the estimator $\delta_{indirect}$ is $2\sigma_e^2 + \frac{6}{9}\sigma_e^2 = \frac{24}{9}\sigma_e^2$.

It is possible to combine the direct and indirect comparisons in a single estimator, known as the intra-estimator because it is based exclusively on intra-experiment comparisons:

$$\delta_{intra} = \theta_1 \underbrace{(y_{33} - y_{43})}_{\text{Direct comparison}} + \theta_2 \underbrace{\left(y_{31} - y_{42} - \frac{y_{11} + y_{21} + y_{51}}{3} + \frac{y_{12} + y_{22} + y_{52}}{3} \right)}_{\text{Indirect comparison}}$$

δ_{intra} is a linear combination of observations. It is therefore straightforward to calculate its variance from the model. We then look for the values of θ_1 and θ_2 minimizing the variance of δ_{intra} , with the constraint $\theta_1 + \theta_2 = 1$ as a condition of non-bias. We are now dealing with a problem of optimization under constraint, which can be resolved by the method of Lagrange multipliers. For this example, we obtain values for θ_1 and θ_2 of 0.5714 and 0.4286, respectively. By inserting the corresponding numerical values into the expression above, we obtain a value of 7.71 for δ_{intra} , corresponding to the difference between the adjusted means of treatments 3 and 4 presented in Table 4.2.

The notion of indirect comparisons is, thus, equivalent to the notion of adjusted means obtained by the ordinary least squares method (Kempton and Fox 1997).

The statistical model used for the analysis can be used to calculate $var_{\delta_{intra}}$, the variance of the estimator δ_{intra} . In the example given, the variance of the difference between t_3 and t_4 is: $var_{\delta_{intra}} = 7.13$.

4.3.2.2 Mixed Model

This time, the calculation of adjusted means is based on the mixed model (4.1), which can be expressed as follows, omitting the “*” for the sake of clarity:

$$y_{ij} = \mu + t_i + E_j + \varepsilon_{ij}$$

where y_{ij} is the mean value for the quantity of interest for treatment i ($i = 1, \dots, 5$) in experiment j ($j = 1, \dots, 3$) and μ , t_i , E_j , and ε_{ij} are the general mean, the effect of treatment i , the effect of treatment j and the residual associated with y_{ij} , respectively.

The observed experiments are assumed to represent a random sample of the population of possible experiments. This implies that the effect E_j is considered to be random.

The random effects E_j and ε_{ij} are assumed to follow a normal distribution with a mean of zero and a variance of σ_E^2 and σ_ε^2 , respectively. In addition, the different random effects are assumed to be independent of each other.

The parameters of the model are estimated by the REML method. The adjusted means are then calculated from the model, as for the fixed-effects model.

Comparisons Between Treatments

The comparison between two treatments is based on an intra-estimator combining direct and indirect comparisons, as for the fixed-effects model. However, in addition to this intra-estimator, the mixed model also makes use of an inter-estimator based on the difference between the sum of the values for variable y in the experiments in which we observe the compared treatments. These totals contain information about the difference between treatments that can be recovered during analysis with a mixed model.

For example, for treatments t_3 and t_4 , the inter-estimator is given by:

$$\delta_{inter} = (y_{11} + y_{21} + y_{31} + y_{51}) - (y_{12} + y_{22} + y_{42} + y_{52}),$$

Expressed in the model as:

$$\begin{aligned} \delta_{inter} &= (\mu + t_1 + E_1 + \varepsilon_{11} + \mu + t_2 + E_1 + \varepsilon_{21} + \mu + t_3 + E_1 + \varepsilon_{31} + \mu + t_5 + E_1 + \varepsilon_{51}) \\ &\quad - (\mu + t_1 + E_2 + \varepsilon_{12} + \mu + t_2 + E_2 + \varepsilon_{22} + \mu + t_4 + E_2 + \varepsilon_{42} + \mu + t_5 + E_2 + \varepsilon_{52}) \\ &= t_3 - t_4 + (4E_1 + \varepsilon_{11} + \varepsilon_{21} + \varepsilon_{31} + \varepsilon_{51}) - (4E_2 + \varepsilon_{12} + \varepsilon_{22} + \varepsilon_{42} + \varepsilon_{52}) \end{aligned}$$

This comparison includes experiment effects but, in the framework of a mixed model, these effects are considered random and contribute to the variance of the estimator. The variance of δ_{inter} is $var_{\delta_{inter}} = 32 \sigma_E^2 + 8 \sigma_\varepsilon^2$. For the example given, we obtain: $var_{\delta_{inter}} = 32 * 25.54 + 8 * 6.25 = 867.28$.

We can combine the intra- and inter-estimators to obtain an estimator of minimum variance. This estimator, known as δ_{combin} , is the mean of the estimators δ_{intra} and δ_{inter} weighted according to their respective precisions (Kempton and Fox 1997):

$$\begin{aligned} \delta_{combin} &= \gamma_1 \delta_{intra} + \gamma_2 \delta_{inter} \\ \text{with } \gamma_1 &= \frac{1/var_{\delta_{intra}}}{1/var_{\delta_{intra}} + 1/var_{\delta_{inter}}} \text{ and } \gamma_2 = \frac{1/var_{\delta_{inter}}}{1/var_{\delta_{intra}} + 1/var_{\delta_{inter}}} \end{aligned}$$

For the example given, we obtain: $\gamma_1 = \frac{1/7.13}{1/7.13 + 1/867.3} = 0.9918$ and $\gamma_2 = \frac{1/867.3}{1/7.13 + 1/867.3} = 0.0082$.

Thus, for this example, most of the information is provided by the intra comparison.

From the statistical model used for the analysis, it is also possible to calculate the precision of the combined estimator. For the example given, the variance of the difference between t_3 and t_4 estimated with the mixed model is $var_{\delta_{combin}} = 7.08$.

The estimate obtained with the mixed model is, thus, slightly more precise than that obtained with the fixed-effects model in this case.

Estimation of the difference between two treatments has been presented as two different operations, with intra-environment information taken into account through

the δ_{intra} estimator, and inter-environment information taken into account through the δ_{inter} estimator. The two sources of information are then combined. In practice, the effects of treatment are estimated in a single procedure, by the REML method.

If the number of experiments is sufficiently high (at least 5–10), the use of mixed models is recommended for the calculation of adjusted means, which are generally more precise when obtained with mixed models than when obtained with fixed-effects models.

4.4 The Factors “Site” and “Year”

4.4.1 Objective

In a network running over several years, experiments performed in the same year have a common characteristic that is the year effect. In a variety evaluation network, relatively dry years, in which the crops suffer from a strong water deficit, tend to result in yields below the general mean value, for all the experiments performed in the year concerned.

Similarly, in multisite networks, all the experiments performed at the same site have a common characteristic that is the site effect.

In a multisite network running over several years analyzed with model (4.1), the interenvironment variability is described by the “experiment” factor alone, and the effects of the different experiments are considered to be independent. However, this hypothesis of independence is not always realistic, because the effects of experiments performed in the same year or at the same site tend to be more similar than the effects of experiments performed in different years or at different sites.

In this section, we present the analysis of a multisite network run over several years, using a model taking site and year effects into account. This model is an extension of model 4.1. We consider only the analysis of treatment means by experiment, i.e. step 2 in a two-step analysis.

4.4.2 The “Wheat_Multi” Example

Each year, about 30 new winter wheat varieties are registered in France. *Arvalis-Institut du Végétal* is responsible for evaluating the agronomic and technological value of these varieties in field experiments, with a view to making recommendations.

Each variety is generally tested in 2 years, but the varieties with the worst performance may be tested in only 1 year, and those with the best performance may be tested in more than 2 years. Certain control varieties are also followed over a longer period. The varieties are observed for only a limited number of years, to

Table 4.5 Excerpt from the data set for the “wheat_pluri” example

Yield	Variety	Experiment	Site	Year	Vmoyj
87.8	V4	OBERNAI_2005	OBERNAI	2005	5.14
99.3	V8	OBERNAI_2005	OBERNAI	2005	5.14
94.6	V13	OBERNAI_2005	OBERNAI	2005	5.14
93.1	V17	OBERNAI_2005	OBERNAI	2005	5.14
90.2	V19	OBERNAI_2005	OBERNAI	2005	5.14
...

provide an opportunity to evaluate the newly registered varieties experimentally. In experiments of this type, no individual experiment includes all the varieties tested.

The dataset corresponds to the results for 80 varieties in 113 experiments. These experiments were performed over a period of 6 years, at 54 different sites. Over the period covered by the dataset, there were, therefore, on average, about two experiments per site, and just under 20 experiments per year.

For each variety in each experiment, the data table shows the adjusted mean for the variety obtained by analysis of the data of that experiment. The data table also shows the name of the variety, the name of the experiment, the name of the site, the year and the variance of the mean data for each experiment, calculated with Eq. (4.7).

The total number of observations is 1942. If all the varieties had been present in all the experiments, the dataset would have included $80 \times 113 = 9040$ observations. The proportion of cells filled in the dataset is thus $1942/9040 = 0.21$. Table 4.5 shows the first few lines of the dataset.

The missing data in this dataset are missing at random (MAR) because the missing data depend on the observed data and not on the (non-observed) missing data. The analysis of these data with a mixed model is therefore valid (Piepho and Möhring 2006).

4.4.3 Model for the Analysis of Mean Data

The factors “year” and “site” are crossed factors and variability between experiments is described as the sum of the effects of year, site and the interaction between year and site (Dagnelie 1998).

In model (4.1), the effects of the variety \times experiment interaction are completely confounded with the experimental errors. In the model breaking down the effect of an experiment in terms of year and site, it is also possible to break down the effect of the variety \times experiment interaction into the sum of the variety \times year interaction, the variety \times site interaction and a variety \times year \times site interaction. In this case, only the variety \times year \times site interaction is confounded with the experimental error.

This model can, thus, be used to describe some of the variability between experiments, and the variety \times experiment interaction, using the factors “year” and “site”.

The model used for data analysis is as follows:

$$y_{irs} = \mu + t_i + A_r + L_s + AL_{rs} + tA_{ir} + tL_{is} + \varepsilon_{irs}^* \quad (4.10)$$

where y_{irs} is the mean value for treatment i ($i = 1, \dots, I$), in year r ($r = 1, \dots, R$) at site s ($s = 1, \dots, S$) and μ , t_i , A_r , L_s , AL_{rs} , tA_{ir} , tL_{is} , and ε_{irs}^* are the general mean, the effect of treatment i , the effect of year r , the effect of site s , the effect of the interaction between year r and site s , the effect of the interaction between treatment i and year r , the effect of the interaction between treatment i and site s and the residual associated with y_{irs} , respectively. This residual includes both the experimental error and a treatment \times year \times site interaction effect.

The observed sites are assumed to be a random sample of a population of possible sites. The years are assumed to be a random sample of the population of possible years, although, in practice, the years are generally successive and, thus, never really chosen at random. Nevertheless, in practice, it is usually assumed that the effect of year is random. If we consider the effects of “year” and “site” to be random, this implies that all the interactions involving these factors are also random.

The effects A_r , L_s , AL_{rs} , tA_{ir} , tL_{is} and ε_{irs}^* are considered to be random and assumed to follow a normal distribution with a mean of zero and a variance of, σ_A^2 , σ_L^2 , σ_{AL}^2 , σ_{tA}^2 , σ_{tL}^2 and $\sigma_{\varepsilon^*}^2$, respectively. Furthermore, the various random effects are assumed to be independent of each other.

4.4.4 Estimation of the Variance of the Treatment \times Year \times Site Interaction

The residual ε_{irs}^* encompasses both the experimental error associated with the mean y_{irs} and the effect of the treatment \times year \times site interaction. We can obtain an estimate of the variance of the treatment \times year \times site interaction by calculating the difference between the variance of the residuals ε_{irs}^* and the variance of the estimation errors of the mean data:

$$\hat{\sigma}_{tAL}^2 = \hat{\sigma}_{\varepsilon^*}^2 - Vmean$$

where $\hat{\sigma}_{\varepsilon^*}^2$ is the estimated residual variance of model (4.10) and $Vmean$ is calculated with Eq. (4.8).

Generally, the value for the variance of the treatment \times year \times site interaction obtained by difference, as described above, is an approximation and will be different from the estimate obtained in a single-step analysis of individual data.

4.4.5 Variance of the Difference Between Two Treatments

For model (4.10), the variance of the difference between two treatment means is given by the following formula:

$$\text{Var}(\hat{\mu}_i - \hat{\mu}_{i'}) = 2 \left(\frac{\sigma_{iL}^2}{S} + \frac{\sigma_{iA}^2}{R} + \frac{\sigma_{\varepsilon^*}^2}{SR} \right)$$

In the case of a network of experiments with a randomized complete block design and the same number of blocks in each experiment, we have:

$$\sigma_{\varepsilon^*}^2 = \sigma_{iAL}^2 + \frac{\sigma_{\varepsilon}^2}{K}$$

where σ_{iAL}^2 is the variance of the treatment \times year \times site interaction, σ_{ε}^2 is the variance of the experimental errors for the individual data and K is the number of blocks per experiment (σ_{ε}^2 and K are assumed to be identical in all experiments). The variance of the difference between two treatment means is thus given by the following equation (Kempthorne 1960):

$$\text{Var}(\hat{\mu}_i - \hat{\mu}_{i'}) = 2 \left(\frac{\sigma_{iL}^2}{S} + \frac{\sigma_{iA}^2}{R} + \frac{\sigma_{iAL}^2}{SR} + \frac{\sigma_{\varepsilon}^2}{SRK} \right) \quad (4.11)$$

Equation (4.11) will be used in Chap. 5 for the dimensioning of a multisite network of experiments run over several years.

4.4.6 Analysis of the “Wheat_Multi” Example, with R Script

The R script used to analyze the “wheat_multi” example is very similar to that used to analyze the “wheat” example in Chap. 3. For this reason, we present the entire script for the analysis of the “wheat_multi” example, but we do not comment on the parts of the script common to the two examples, or on outputs equivalent to those for the “wheat” example.

```
# R version 3.4.3
# loading of the packages
library(lme4) # version 1.1-14
library(emmeans) # version 1.0
library(car) # version 2.1-6
library(outliers) # Version: 0.14
```

```

Linear mixed model fit by REML ['lmerMod']
Formula: yield ~ variety + (1 | year) + (1 | site) + (1 | year:site) +
      (1 | variety:year) + (1 | variety:site)
Data: DF
REML criterion at convergence: 11468.3
Random effects:
Groups      Name      Std.Dev.
variety:site (Intercept) 1.888
variety:year (Intercept) 1.517
year:site    (Intercept) 8.756
site         (Intercept) 8.180
year         (Intercept) 4.208
Residual                    3.802
Number of obs: 1942, groups:
variety:site, 1410; variety:year, 184; year:site, 113; site, 54; year, 6
Fixed Effects:
(Intercept)      V10      V11      V12      V13      V14
      94.13293      0.72601      1.35587      2.02108      2.38906      -4.06531

```

Fig. 4.9 Information about the mixed model

```

# adjustment of the model
res.lmer <- lmer(rendement ~ variete + (1|annee) + (1|lieu) +
  (1|annee:lieu) + (1|variete:annee) + (1|variete:lieu),
  data=DFmoy,
  na.action=na.exclude)
res.lmer

```

The result of the adjustment of the model is presented in Fig. 4.9. This adjustment includes an estimation of the standard deviation of the different random effects and an estimation of the fixed effects, only some of which are shown in this figure.

```

# Validation of the model
plot(fitted(res.lmer), residuals(res.lmer), abline(h=0))
hist(residuals(res.lmer))

```

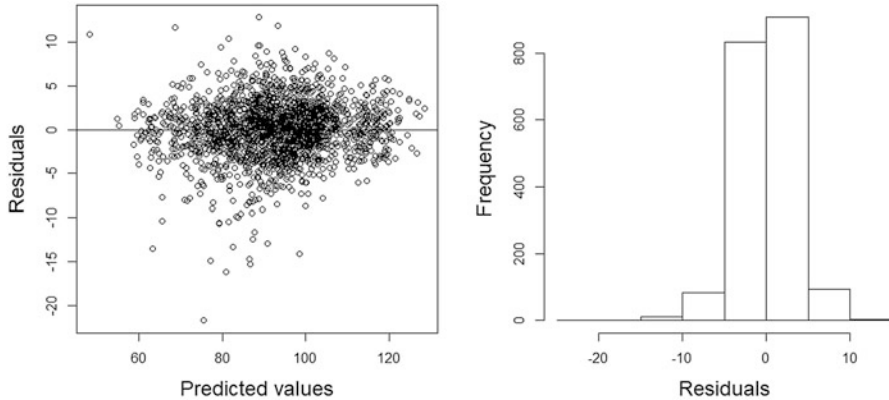


Fig. 4.10 Graphical representations for model validation. On the left: scatter plot of the residuals as a function of predicted values. On the right: histogram of the residuals

Figure 4.10 provides a graphical representation of the model residuals. These graphs reveal the presence of one or several suspect data points worthy of further investigation to check that there is no error affecting the value of these data.

```
DFmoy$residus <- residuals(res.lmer)
grubbs.test(DFmoy$residus)
DFmoy[which.min(DFmoy$residus),]
```

Grubbs’ test, the results of which are shown in Fig. 4.11, revealed the presence of a suspect residual, and Fig. 4.12 shows the observation corresponding to the residual with the lowest value.

```
par(mfrow=c(2,2))
hist(ranef(res.lmer)$lieu[, "(Intercept)"])
hist(ranef(res.lmer)$`annee:lieu`[, "(Intercept)"])
hist(ranef(res.lmer)$`variete:annee`[, "(Intercept)"])
hist(ranef(res.lmer)$`variete:lieu`[, "(Intercept)"])
par(mfrow=c(1,1))
```

The lines of code above are new relative to the script for the analysis of the “wheat” example. They can be used to display histograms of the random effects of “site” and of the year \times site, variety \times year and variety \times site interactions. These effects are centered on zero and are assumed to follow a normal distribution. The number of effects is sufficient, in the “wheat_multi” example, for the graphs shown in Fig. 4.13 to be plotted.


```
Grubbs test for one outlier

data: DF$residus

G = 6.7033, U = 0.9768, p-value = 1.52e-08

alternative hypothesis: lowest value -21.6932254973021 is an outlier
```

Fig. 4.11 Grubbs' test

yield	variety	experiment	site	year	Vmoyj	residuals
95253.71	V53	LONGVIC_2008	LONGVIC	2008	3.63	-21.69323

Fig. 4.12 Observation corresponding to the residual with the lowest value

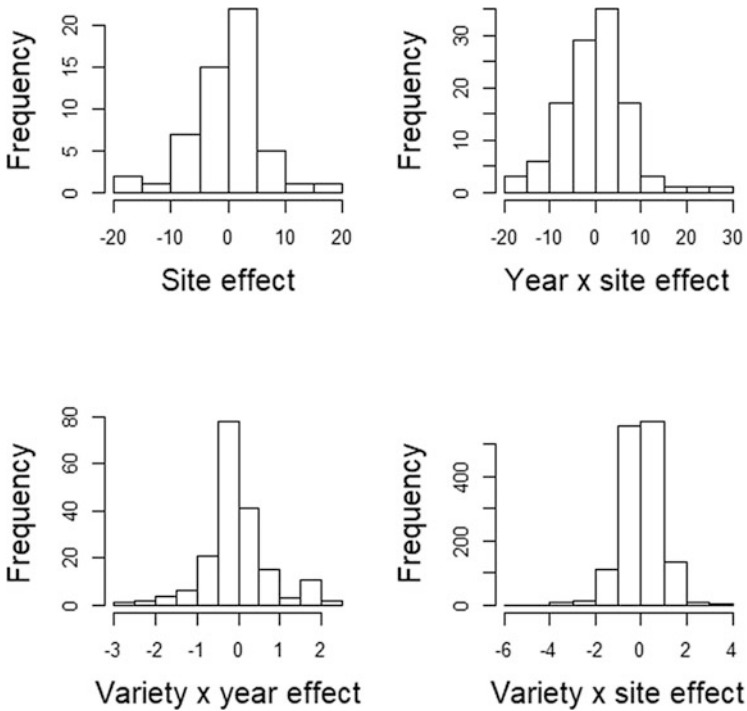


Fig. 4.13 Histograms of the different random effects

Figure 4.14 shows the result of an F test of the equality of the means of the 80 varieties. The probability associated with the test statistic ($2.974e-11$) indicates that the differences between the varieties are highly significant.

```
Analysis of Deviance Table (Type II Wald F tests with Kenward-Roger df)
```

```
Response: yield
```

```

              F Df Df.res    Pr(>F)
variety 4.2039  79  95.601 2.974e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Fig. 4.14 F test

```
# ANOVA table
Anova(res.lmer, test.statistic="F")
```

```
# Adjusted means
moy_var <- lsmeans(res.lmer, ~variete)
moy_var
```

Figure 4.15 shows the start of the adjusted means table for the varieties.

```
# Comparisons with the general mean
contrast(moy_var, method="eff", adjust="sidak")
```

We do not use the **pairs** function of the **emmeans** library, which performs all the possible pairwise comparisons, here because, with 80 varieties, such an analysis would involve $(80 \times 80 - 80) / 2 = 3160$ comparisons, and would entail the use of an excessive amount of computer time. Instead, we use the **contrast** function with the argument **method = "eff"** to compare each variety with the general mean. This approach requires 80 tests in total for this example. The difference between the variety mean and the general mean corresponds to the effect of the variety. The first few results for this approach are shown in Fig. 4.16. A statistical test is performed to test the null hypothesis of an absence of variety effect. The probability value for each test is adjusted to account for multiple testing.

```
# Comparisons with a control
contrast(moy_var, method="trt.vs.ctrl", ref=2)
```

Fig. 4.15 Adjusted means

variety	lsmean	SE	df	lower.CL	upper.CL
V1	94.13293	2.948956	22.88	88.03079	100.23506
V10	94.85893	2.468206	11.89	89.47585	100.24201
V11	95.48879	3.514057	46.21	88.41621	102.56137
V12	96.15400	2.898815	21.27	90.13018	102.17782
V13	96.52198	2.405186	10.78	91.21485	101.82912
V14	90.06761	2.697724	16.72	84.36863	95.76659
V15	96.05092	2.460441	11.74	90.67702	101.42481
V16	95.21234	2.898815	21.27	89.18852	101.23616
V17	94.17538	2.630766	15.34	88.57886	99.77191
V18	74.64226	3.416976	43.22	67.75228	81.53224
V19	92.47296	2.398402	10.65	87.17315	97.77277
...					

Confidence level used: 0.95

Figure 4.17 shows comparisons between variety means and the mean of the control variety V10.

```
# Random effects
ranef(res.lmer)
```

The `ranef` function of the `lme4` library can be used to display the random effects. We do not show this output here to avoid overloading the text.

```
# Variances of the means and of the interaction
Vmoy <- mean(by(DFmoy$Vmoyj, DFmoy$expérimentation, unique))
Vmoy

var_intera <- summary(res.lmer)$sigma^2 - Vmoy
var_intera
```

contrast	estimate	SE	df	t.ratio	p.value
V1 effect	1.130022531	1.8758398	61.50	0.602	1.0000
V10 effect	1.856029391	0.9860050	72.07	1.882	0.9949
V11 effect	2.485889880	2.6595996	238.98	0.935	1.0000
V12 effect	3.151100706	1.7901772	51.23	1.760	0.9991
V13 effect	3.519081447	0.8197594	78.88	4.293	0.0040
V14 effect	-2.935292120	1.4586923	81.15	-2.012	0.9796
V15 effect	3.048014529	0.9669837	66.73	3.152	0.1768
V16 effect	2.209434039	1.7901772	51.23	1.234	1.0000
V17 effect	1.172480400	1.3357035	109.33	0.878	1.0000
V18 effect	-18.360641577	2.5361994	584.81	-7.239	<.0001
V19 effect	-0.529945261	0.8011623	72.18	-0.661	1.0000
...					

Fig. 4.16 Comparison of variety means with the general mean

contrast	estimate	SE	df	t.ratio	p.value
V1 - V10	-0.72600686	2.099362	61.74	-0.346	1.0000
V11 - V10	0.62986049	2.843741	199.53	0.221	1.0000
V12 - V10	1.29507131	2.029242	53.59	0.638	1.0000
V13 - V10	1.66305206	1.256569	70.67	1.323	0.9820
V14 - V10	-4.79132151	1.732767	75.76	-2.765	0.2448
V15 - V10	1.19198514	1.332355	65.10	0.895	0.9995
V16 - V10	0.35340465	2.029242	53.59	0.174	1.0000
V17 - V10	-0.68354899	1.683198	92.91	-0.406	1.0000
V18 - V10	-20.21667097	2.740451	396.15	-7.377	<.0001
V19 - V10	-2.38597465	1.243528	67.77	-1.919	0.7839
...					

Fig. 4.17 Comparison with the control variety V10

The first line of the script shown above calculates a common estimate of the variance of the mean data and stores this value in the object **Vmoy**. In the case of this example, we obtain $Vmoy = 3.107889$.

The residual variance of the mixed model is obtained with the **summary(res.lmer)\$sigma^2** command. It is thus possible to calculate the variance of the variety \times year \times site interaction, **var_intera**, by simple subtraction, using the residual variance of the mixed model and **Vmoy**, the variance of the mean data. For the sample studied here, we obtain $var_intera = 11.34765$. This estimate will be used to dimension a network of experiments in Chap. 5.

References

- Dagnelie P (1998) *Statistique Théorique et Appliquée. Tome 2: inférence statistique à une et à deux dimensions*. De Boeck & Larcier, Paris/Bruxelles, p 659
- Frensham A, Culis B, Verbila A (1997) Genotype by environment variance heterogeneity in a two-stage analysis. *Biometrics* 53:1373–1383
- Gilmour AR (2000) Post blocking gone too far! Recovery of information and spatial analysis in field experiments. *Biometrics* 56:944–946
- Kempthorne O (1960) *The design and analysis of experiments*. Wiley, New York, 631 p
- Kempton RA, Fox PN (1997) *Statistical methods for plant variety evaluation*. Chapman & Hall, London, 191 p
- Möhring J, Piepho H-P (2009) Comparison of weighting in two-stage analysis of plant breeding trials. *Crop Sci* 49:1977–1988
- Monod H (2001) *Alpha-plans, carrés semi-latins et autres plans en répliques. Comment les utiliser?* ITCF, Paris, 44p. ISBN:2-86492-446-3
- Piepho H-P (1996) Comparing cultivar means in multilocation trials when the covariance structure is not circular. *Heredity* 76:198–203
- Piepho H-P, Möhring J (2006) Selection in cultivar trials – is it ignorable? *Crop Sci* 46:192–201
- Piepho H-P, Büchse A, Emrich K (2003) A hitchhiker’s guide to mixed models for randomized experiments. *J Agron Crop Sci* 189:310–322
- Pinheiro JC, Bates DM (2000) *Mixed-effects models in S and S-PLUS*. Springer, New York, 528 p
- Smith A, Cullis B, Thompson R (2001a) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* 57:1138–1147
- Smith A, Cullis B, Gilmour A (2001b) The analysis of crop variety evaluation data in Australia. *Aust NZ J Stat* 43(2):129–145
- Spike J, Piepho H-P, Hu X (2005) Analysis of unbalanced data by MIXED linear models using MIXED procedure of the SAS system. *J Agron Crop Sci* 191:47–54

Chapter 5

Planning an Experimental Network



5.1 Objective

When planning a network of experiments, it is generally important to know **how many** experiments should be performed. However, it is often forgotten that it is just as important to know **how** to choose the environments in which the experiments will be performed.

This question of how to choose environments is both simple and complex. If the objective of the experimental network is to obtain conclusions that can be generalized to a population of environments, then the environments included in the network should be chosen at random from the population of possible environments. This involves the definition of the target population of environments, which is not as simple as it might first appear. Furthermore, it is not always possible to select the environments at random, essentially for practical reasons. So, for example, costs are often limited by performing experiments in areas close to the experimental station. Another example concerns the choice of farmers whose land will be used for the experiments, in situations in which the farmer will be required to act during the experiment. In such cases, it is common for researchers to collaborate regularly with the same farmers, to guarantee the success of the experiments.

The question of how many experiments should be performed is more straightforward but its resolution requires some idea of the experimental variability at the various scales observed. Where possible, an existing dataset is used as a basis for estimating this variability, to make it possible to plan the acquisition of new data.

The question of the number of replicates is multi-faceted. In a multisite network, the number of replicates per experiment and the number of experiments must be determined. In the case of a multi-local and multi-year network, the number of years must also be determined.

Generally, the dimensions of an experiment or a network of experiments are determined so as to satisfy an objective in terms of statistical power (Gozé 1992). Power is the capacity of the experiment or network of experiments to detect a

difference between treatments, if there is one. However, the calculation of power requires the formulation of hypotheses concerning the real differences between treatments, which is generally difficult to achieve in a realistic manner. In addition, calculations of power are more complicated for mixed models than for linear fixed-effects models. Readers can obtain more information about this approach from Littell et al. (2006).

We describe below a simpler approach involving calculation of the variance of the difference between two treatments. The principle is simple. The variance of the difference between two treatments depends on the experimental variability at the different scales observed, and the variances at these different scales must therefore first be estimated. The variance of the difference between two treatments is then calculated from the estimated variances obtained in the first step, by varying the number of replicates. Finally, these simulations are used to determine the number of replicates providing the best compromise between the precision of the comparisons and the cost of the network. The precision of the comparisons is estimated as the least significant difference (LSD) at 5%, which is about equal to twice the standard deviation of the difference between two treatments:

$$LSD = 2\sqrt{\widehat{Var}(\hat{\mu}_i - \hat{\mu}_j)}.$$

This approach is applied, in Sects. 5.2.1 and 5.2.2 to the case of a multisite network and a multisite multiyear network of experiments with a randomized complete block design. Section 5.3 shows how to optimize an experimental network based on the calculation of the variance of a contrast other than the difference between two treatments.

In this chapter, we assume that the residual variance of individual experiments is identical for all the experiments of the network. We also restrict this approach to the dimensioning of a complete network, i.e. one in which all the treatments are present in all the experiments. However, we think that this approach should provide in most cases, even if these conditions are not fulfilled, an order of magnitude of the number of replicates to be performed within the network, which is, in practice, often sufficient.

5.2 Comparison of Two Treatments

5.2.1 Case of a Multisite Network

5.2.1.1 Variance of the Difference Between Two Treatments

The variance of the difference between two treatments in the case of a multisite network described by model (3.1) is given by Eq. (3.2), which we recall here:

$$\text{Var}(\hat{\mu}_i - \hat{\mu}_{i'}) = 2 \left(\frac{\sigma_{tE}^2}{J} + \frac{\sigma_\varepsilon^2}{JK} \right)$$

where $\text{Var}(\hat{\mu}_i - \hat{\mu}_{i'})$ is the variance of the estimator of the difference between the means of treatments i and i' , and σ_{tE}^2 , σ_ε^2 , J and K are the variance of the treatment \times experiment interaction, the variance of the residuals, the number of experiments and the number of blocks per experiment, respectively.

σ_ε^2 and σ_{tE}^2 can be estimated with model (3.1), by analyzing the individual data of a network of experiments with a complete randomized block design. In a two-step analysis, σ_{tE}^2 can be estimated with Eq. (4.9), and σ_ε^2 can be estimated with the following formula: $\frac{\sum_{j=1}^J 0.5 n_j \text{VMC}_j}{J}$, where VMC_j , n_j and J are the mean variance of pairwise comparisons of treatments for experiment j , the number of replicates in experiment j (i.e. the number of blocks for an experiment with a complete randomized block design) and the number of experiments, respectively.

5.2.1.2 “Wheat” Example

The analysis of the “wheat” example in Chap. 3 provided estimates of σ_ε^2 and σ_{tE}^2 : $\hat{\sigma}_\varepsilon^2 = 6.10$ and $\hat{\sigma}_{tE}^2 = 6.53$ (Fig. 3.12; N.B., this figure shows the standard deviations). By replacing the theoretical values with these estimates in Eq. (3.2), we obtain: $\widehat{\text{Var}}(\hat{\mu}_i - \hat{\mu}_{i'}) = 2 \left(\frac{6.53}{J} + \frac{6.10}{JK} \right)$.

Figure 5.1 shows the LSD between two treatments for a number of experiments J ranging from 3 to 20, and a number of blocks K ranging from 2 to 5. If we know the precision desired, we can simply refer to this graph to determine the size of the network required to achieve this level of precision. It is also possible to obtain a certain amount of information from Fig. 5.1 that can then be used to guide the design of an experimental network. For example, this figure indicates that the gain in precision provided by each additional block decreases as the number of blocks increases. A similar observation can be made concerning the number of experiments. In addition, it shows that for a given total number of blocks, it is better to perform a large number of experiments with only a few blocks per experiment than a small number of experiments with a large number of blocks. Thus, for example, it is better to perform five experiments with three blocks per experiment than three experiments with five blocks per experiment.

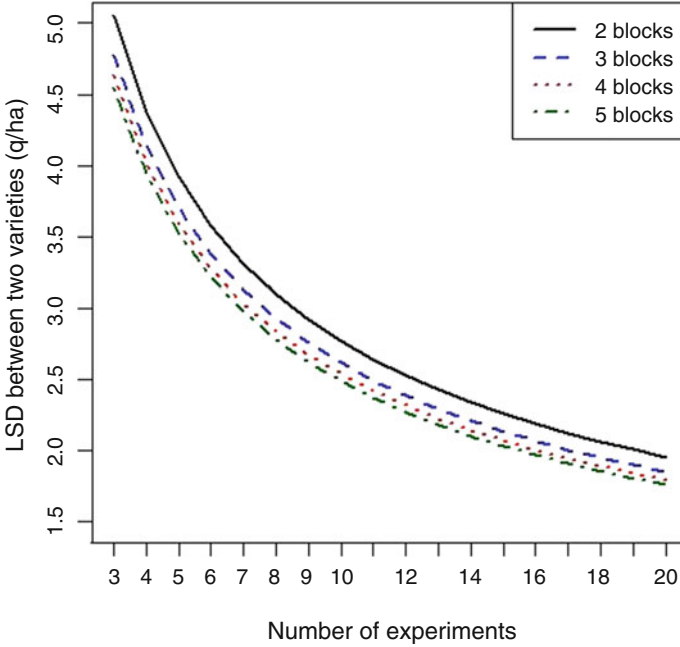


Fig. 5.1 LSD between two varieties as a function of the number of blocks per experiment and the number of experiments. Simulations based on the “wheat” example

5.2.2 Case of a Multisite and Multiyear Network

5.2.2.1 Variance of the Difference Between Two Treatments

The variance of the difference between two treatments in a multisite and multiyear network of experiments with a randomized complete block design is given by Eq. (4.11), which we recall here:

$$Var(\hat{\mu}_i - \hat{\mu}_{i'}) = 2 \left(\frac{\sigma_{iL}^2}{S} + \frac{\sigma_{iA}^2}{R} + \frac{\sigma_{iAL}^2}{SR} + \frac{\sigma_e^2}{SRK} \right)$$

where $Var(\hat{\mu}_i - \hat{\mu}_{i'})$ is the variance of the estimator of the difference between the means of treatments i and i' , and $\sigma_{iL}^2, \sigma_{iA}^2, \sigma_{iAL}^2, \sigma_e^2, S, R$ and K are the variance of the treatment \times site interaction, the variance of the treatment \times year interaction, the variance of the treatment \times site \times year interaction, the variance of the experimental errors on the individual data, the number of sites, the number of years and the number of blocks per experiment, respectively.

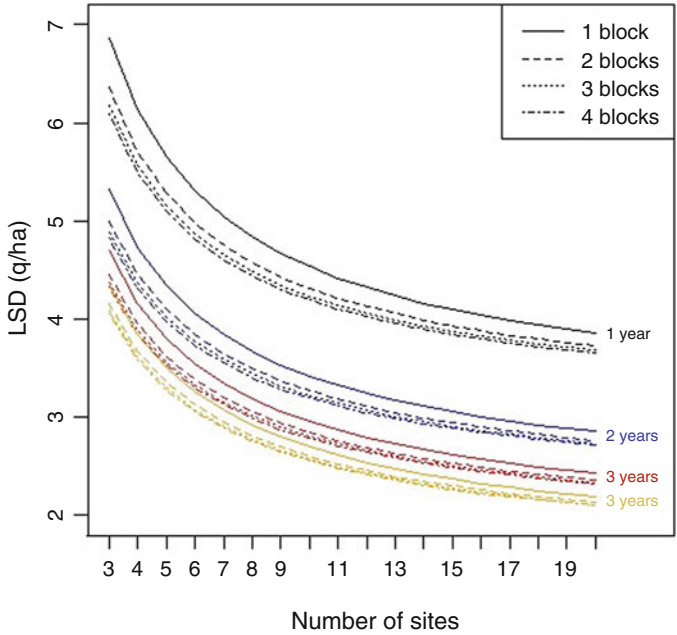


Fig. 5.2 LSD between two varieties as a function of the number of blocks per experiment, the number of sites and the number of years. Simulations based on the “wheat_multi” example

5.2.2.2 The “Wheat_Multi” Example

The analysis of the “wheat_multi” example shown in Sect. 4.4.6 provides estimates of σ_{tL}^2 , σ_{tA}^2 and σ_{tAL}^2 . We can estimate σ_ϵ^2 with the following formula: $\hat{\sigma}_\epsilon^2 = \frac{\sum_{j=1}^J 0.5 n_j VMC_j}{J}$, where VMC_j , n_j and J are the mean variance of pairwise treatment comparisons for experiment j , the number of replicates of experiment j (i.e. the number of blocks for an experiment with a randomized complete block design), and the number of experiments, respectively. For the “wheat_multi” example, we obtain $\hat{\sigma}_\epsilon^2 = 9.94$. By replacing the theoretical values with these estimates in formula (4.11), we obtain: $\widehat{Var}(\hat{\mu}_i - \hat{\mu}_{i'}) = 2 \left(\frac{3.56}{S} + \frac{2.30}{R} + \frac{11.35}{SR} + \frac{9.94}{SRK} \right)$.

The variance components used above were estimated from an incomplete dataset, but that does not prevent them from being used to plan a complete experimental network.

Figure 5.2 shows the LSD between two treatments for a number of sites S ranging from 3 to 20, a number of years R ranging from 1 to 4 and a number of blocks per experiment ranging from 1 to 4. If we know the precision desired, we can simply refer to this graph to determine the size of the network required to achieve this level

of precision. It can be noted that the gain in precision provided by an additional replicate decreases as the number of replicates increases. The gain in precision is small when we pass from three to four blocks. Four-block experiments do not, therefore, appear to be justified, and we could even reasonably pose the question as to the relevance of including more than two blocks per experiment. A similar sort of reasoning can be applied to the number of sites and the number of years.

5.3 Other Contrasts

It is possible to optimize an experimental network based on a calculation of the variance of a contrast other than the difference between two treatments. For example, we may wish to compare the mean of a particular treatment with the mean of several control treatments. Another frequently used type of comparison involves comparing the mean of a treatment with the general mean, defined as the mean of the treatment means. This type of contrast is called the effect of treatment. Whatever the type of contrast chosen, the LSD is defined as follows: $ppds = 2\sqrt{\widehat{Var}_{contrast}}$, where $\widehat{Var}_{contrast}$ is the estimated variance for the contrast chosen.

5.3.1 Comparison to the Mean of Several Controls

5.3.1.1 Case of a Multisite Network

The variance of the difference between a treatment and the mean of several control treatments in a multisite network described by model (3.1) is given by the following formula:

$$Var(\hat{\mu}_i - \widehat{tem}) = \left(1 + \frac{1}{n_{tem}}\right) \left(\frac{\sigma_{iE}^2}{J} + \frac{\sigma_{\epsilon}^2}{JK}\right)$$

where $\hat{\mu}_i$, \widehat{tem} , n_{tem} , σ_{iE}^2 , σ_{ϵ}^2 , J and K are the estimated mean of treatment i , the mean of the estimated means of the control treatments, the number of control treatments, the variance of the treatment x experiment interaction, the variance of the residuals, the number of experiments and the number of blocks per experiment, respectively. If there is only one control treatment, we obtain the variance of the difference between two treatments given by Eq. (3.2).

The LSD for this contrast can be calculated by replacing the theoretical values of the variances σ_{tE}^2 , σ_ε^2 by their estimates from a dataset.

5.3.1.2 Case of a Multisite and Multiyear Network

The variance of the difference between a treatment and the mean of several control treatments in a multisite and multiyear network of experiments with a randomized complete block design is given by the following formula:

$$\text{Var}(\hat{\mu}_i - \widehat{tem}) = \left(1 + \frac{1}{n_{tem}}\right) \left(\frac{\sigma_{tL}^2}{S} + \frac{\sigma_{tA}^2}{R} + \frac{\sigma_{tAL}^2}{SR} + \frac{\sigma_\varepsilon^2}{SRK}\right)$$

where $\hat{\mu}_i$, \widehat{tem} , n_{tem} , σ_{tL}^2 , σ_{tA}^2 , σ_{tAL}^2 , σ_ε^2 , S , R and K are the estimated mean of treatment i , the mean of the estimated means of the control treatments, the number of control treatments, the variance of the treatment \times site interaction, the variance of the treatment \times year interaction, the variance of the treatment \times site \times year interaction, the variance of the experimental errors on the individual data, the number of sites, the number of years and the number of blocks per experiment, respectively. When there is only one control treatment, we obtain the variance of the difference between two treatments given by Eq. (4.11).

The LSD for this contrast can be obtained by replacing the theoretical values of the variances σ_{tL}^2 , σ_{tA}^2 , σ_{tAL}^2 and σ_ε^2 by their estimates based on a dataset.

5.3.2 Comparison with the General Mean

5.3.2.1 Case of a Multisite Network

The variance of the difference between a treatment and the general mean in the case of a multisite network described by model (3.1) is given by the following equation:

$$\text{Var}(\hat{\mu}_i - \widehat{GM}) = \left(1 - \frac{1}{n_{trait}}\right) \left(\frac{\sigma_{tE}^2}{J} + \frac{\sigma_\varepsilon^2}{JK}\right)$$

where $\hat{\mu}_i$, \widehat{GM} , n_{trait} , σ_{tE}^2 , σ_ε^2 , J and K are the estimated mean of treatment i , the estimated general mean, the total number of treatments, the variance of the treatment \times experiment interaction, the variance of the residuals, the number of experiments and the number of blocks per experiment, respectively.

The LSD for this contrast can be calculated by replacing the theoretical values of the variances σ_{tE}^2 and σ_ε^2 with their estimates from a dataset.

5.3.2.2 Case of a Multisite and Multiyear Network

The variance of the difference between a treatment and the general mean in the case of a multisite and multiyear network of experiments with a randomized complete block design is given by the following formula:

$$\text{Var}(\hat{\mu}_i - \widehat{GM}) = \left(1 - \frac{1}{n_{\text{trait}}}\right) \left(\frac{\sigma_{iL}^2}{S} + \frac{\sigma_{iA}^2}{R} + \frac{\sigma_{iAL}^2}{SR} + \frac{\sigma_\varepsilon^2}{SRK}\right)$$

where $\hat{\mu}_i$, \widehat{GM} , n_{trait} , σ_{iL}^2 , σ_{iA}^2 , σ_{iAL}^2 , σ_ε^2 , S , R and K are the estimated mean for treatment i , the estimated general mean, the total number of treatments, the variance of the treatment \times site interaction, the variance of the treatment \times year interaction, the variance of the treatment \times site \times year interaction, the variance of the experimental errors on the individual data, the number of sites, the number of years and the number of blocks per experiment, respectively.

As in the previous cases, the LSD for this contrast can be calculated by replacing the theoretical values of the variances σ_{iL}^2 , σ_{iA}^2 , σ_{iAL}^2 , and σ_ε^2 with their values estimated from a dataset.

References

- Gozé E (1992) Détermination de la dimension des réseaux d'essais. *Coton Fibres Trop* 47(2):81–94
 Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O (2006) SAS system for mixed models, 2nd edn. SAS Institute Inc, Cary, 814 p

Part II

Meta-analysis

This second part has two chapters and concerns meta-analysis.

Chapter 6 presents the objectives and principal steps of a meta-analysis. This chapter explains the importance of defining precisely the quantity to be estimated by the meta-analysis and of performing a rigorous systematic search to identify the studies available. This chapter then describes two statistical approaches for analyzing the data extracted from the selected studies: (i) estimation of mean effect size and (ii) meta-regression. These two approaches are illustrated with several examples.

Chapter 7 deals with several specific problems relating to meta-analysis: the definition of effect size and the various associated biases, the analysis of count data with generalized linear models and the analysis of non-linear responses. Finally, this chapter presents an alternative statistical approach based on the use of Bayesian models. These subjects are illustrated with real data.

Chapter 6

Basic Concepts in Meta-analysis



6.1 Definition, Origin and Principal Steps of Meta-analysis

Meta-analysis is one of the methods used for synthetic analyses of knowledge. It combines two approaches:

- Systematic literature review
- Statistical analysis

Chalmers et al. (2002) defined a “systematic review” as the application of strategies to limit bias in the collection, critical evaluation and synthesis of all the relevant studies dealing with a particular subject. ANSES recently (2016) identified 10 different definitions of “systematic review” and proposed the following definition: “A systematic review of the literature involves the exhaustive assembly, evaluation and synthesis of all the relevant studies dealing with a specific question, even if these studies are contradictory. Systematic reviews should be based on a detailed protocol established in advance, favoring a transparent and reproducible approach.”

Systematic reviews are frequently performed in the framework of academic research and in analyses of health and environmental risks. One of the principal limitations of such reviews is their qualitative nature: the result of a systematic review is a narrative conclusion summarizing the studies reviewed. Such analyses provide no quantitative results concerning the force of causal relationships, the size of the effects studied or the level of uncertainty.

Meta-analysis combines statistical analysis with systematic review. This has two advantages: it limits the risks of bias by recovering all the relevant studies on the basis of an explicit protocol and it provides results in a quantitative form. Meta-analysis has been defined in a number of different ways. Glass (1976) defined meta-analysis as the statistical analysis of a large collection of results from individual studies for the purpose of integrating the findings. Other definitions have been proposed, including those of Chalmers et al. (2002) and Koricheva et al. (2013).

All stress the notion of combining different studies with common objectives, with the goal of estimating a quantity of interest.

Historically, meta-analysis was first used at the start of the twentieth century. Karl Pearson (1904) collected and analyzed the results of different studies evaluating the efficacy of a vaccine against typhoid. Yates and Crowther (1941) is often presented as the first paper reporting results of an agronomic meta-analysis (Chalmers et al. 2002). The objective was to determine the optimal dose of phosphate fertilizer in the context of a trade blockade and shortages during the Second World War. A large number of trials testing the effects of various doses of fertilizer on crop yield were collected and a global analysis was performed to calculate, as accurately as possible, the dose of fertilizer required. Meta-analysis became a standard method in the medical domain in the 1990s, under the impetus of the Cochrane Organization in particular, which promoted this approach and organizes workshops on systematic reviews and meta-analyses (www.cochrane.org). Meta-analysis has become an essential technique for identifying risk factors for disease and evaluating the efficacy of medical treatments (Sutton et al. 2000; Borenstein et al. 2009). Certain authors consider meta-analysis to be an approach for limiting the risks of false discovery (Ioannidis 2005). Meta-analysis is increasingly being used outside the medical domain, particularly in ecology, environmental science and agronomy (Philibert et al. 2012a, b; Koricheva et al. 2013; Makowski et al. 2014a, b).

There are two major types of meta-analysis. The first aims to estimate the mean size of a treatment effect (e.g. for a method of disease control or the effect of a cropping system) on a variable of interest (e.g. yield), using all the available relevant studies (see Hossard et al. 2016 for a recent example from the domain of agronomy). The second type of meta-analysis (often referred to as meta-regression) aims to estimate the relationship between a response variable and one or more explanatory variables, making use of a collection of experimental studies including observed values for the variables considered (Philibert et al. 2012b; Yu et al. 2015). These two types of meta-analysis have a number of steps in common:

- Definition of the objective of the meta-analysis: definition of the effect size, response variables, explanatory variables and population.
- Systematic review: definition of the protocol for bibliographic research, identification of relevant studies.
- Selection of studies and data extraction: definition of the selection criteria, application of the criteria to the studies, extraction of data from the selected studies.
- Statistical analysis: definition of one or several statistical models, estimation of parameters, estimation of the quantities of interest, analysis of uncertainty.
- Critical analysis: analyses of sensitivity and of publication bias.
- Communication of the results.

These steps are presented in detail below and are illustrated with examples.

6.2 Estimation of Mean Effect Size

6.2.1 Objective

In this type of meta-analysis (which is by far the most common), the objective is to estimate the effect of an experimental treatment E on a variable of interest X relative to a control C . This requires the definition of several elements:

- The variable of interest X ,
- The control C ,
- The experimental treatment E ,
- The population of studies,
- The effect size allowing comparison of C and E .

Depending on the subject of the meta-analysis, the quantity X may be a continuous variable, such as crop yield, or a discrete variable, such as the number of diseased plants in a given sample. In agronomy, the treatments C and E generally correspond to agricultural practices (e.g. C = plot not treated with pesticides, E = treated plot) or cropping systems (e.g. C = conventional cropping system and E = organic system).

The effect of treatment E is studied by defining a quantity called “effect size”, which compares the value of X obtained when the experimental treatment E is applied (X_E) with that obtained when the control treatment is applied (X_C). Different effect sizes combining X_E and X_C are widely used, including, in particular, the difference between X_E and X_C , $X_E - X_C$, and the $\frac{X_E}{X_C}$ ratio. This ratio can be used to manage data expressed in different units, which is often the case when studies of different origins are used.

It is important to define the population for which the effect size must be estimated as precisely as possible. This population includes all the studies that we wish to include for the estimation of mean effect size. In practice, only a subset of these studies is recovered. This subset constitutes a sample that is used to make inferences for the total population (Fig. 6.1). For example, if the objective of the meta-analysis is to estimate the mean yield loss to be expected when switching from conventional to organic maize production in Europe, then the population corresponds to all

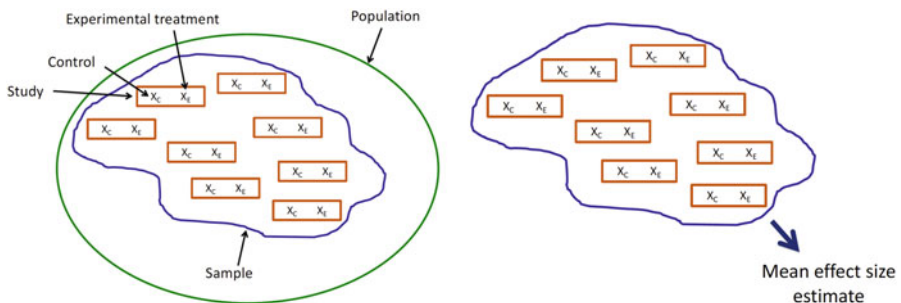


Fig. 6.1 Estimation of mean effect size from a sample of studies from a population

European studies that have measured maize yield in at least one organic and one conventional cropping system (treatment). In practice, a sample of available studies is recovered and used to estimate the mean loss in Europe. This loss is determined by estimating the expected difference in yield or the expected ratio of yields for organic and conventional treatments.

6.2.2 Systematic Search for Studies, Selection of References and Data Extraction

Systematic searches for studies are generally performed with bibliographic databases. It is important to describe the search strategy used as transparently and precisely as possible, to ensure that the procedure used is repeatable. It is therefore necessary to present the search equation, the time period covered and the type of documents sought. The use of two databases or more is recommended (Efsa 2010; Anses 2016).

Studies should then be selected according to explicit criteria concerning the study subject, the quality of the experimental protocol used and the data available. If sufficient time is available, it is often recommended to have two different people apply the selection procedure (Efsa 2010; Anses 2016). However, as the study selection procedure can be long (several weeks or months in practice) it is not always possible to have the studies analyzed, one-by-one, by two different people. If the number of studies identified is large (several hundreds), an initial selection is frequently made on the basis of the titles and abstracts of the articles. A second selection is then applied, based on the entire texts of the studies selected during the first round of screening. The different steps in the selection process should be described, with a Prisma-type diagram (www.prisma-statement.org), for example. Such a diagram can be used to present:

- The number of studies recovered from bibliographic databases,
- The studies from other sources,
- The studies eliminated on the basis of their titles or abstracts,
- The studies eliminated during the reading of the entire text,
- The studies eliminated due to a lack of availability of data essential for the analysis.

The data must then be extracted from the selected studies. All the data required for the calculation of effect size should be extracted (i.e. all data corresponding to measured values of X_E and X_C). When available, data providing information about the precision of the measurements of X_E and X_C (standard deviation, sample size) should also be extracted. The data may be extracted from the tables or figures presented in the studies. Specialist software is required to extract data from figures (e.g. <https://automeris.io/WebPlotDigitizer/>). The extracted data should be archived in computer files. In general, standard spreadsheets are used for this purpose. For complex data structures, it may be useful to archive the data in a relational database.

Certain programs facilitate the steps of study identification, selection and data extraction. For example, the `metagear` package of R (Lajeunesse 2016) includes functions for analyzing article abstracts, for drawing Prisma diagrams and for the automatic extraction of data from graphs.

6.2.3 Estimation of Mean Effect Size with a Fixed-Effects Model

6.2.3.1 Description of the Individual Effect Sizes

The first step is to calculate individual effect sizes from the data extracted from the studies. In most cases, the studies provide mean values of X_E and X_C , denoted \bar{X}_E and \bar{X}_C , respectively. For each study, it is thus possible to estimate different effect sizes, including, in particular, the ratio defined by $R = \frac{\bar{X}_E}{\bar{X}_C}$. In practice, the logarithm of this ratio is frequently used, rather than the ratio itself. The logarithm of the ratio is easier to manipulate mathematically and its distribution is often more symmetric than that of the ratio itself. The log ratio is defined as follows: $L = \ln\left(\frac{\bar{X}_E}{\bar{X}_C}\right)$. Another effect size, known as Hedges' g (Hedges and Olkin 1985), is sometimes used to quantify the difference between \bar{X}_E and \bar{X}_C whilst taking into account measurement variability and sample sizes for the experimental treatment E and the control C , but the values of this effect size are more difficult to interpret than those of the ratio. The examples presented in this chapter are based on the ratio R (or its logarithm), but other types of effect size are presented in Chap. 7.

When the standard deviations of X_E and X_C and the sample sizes used to calculate the mean values are available for the studies considered, it is possible to estimate the standard deviation of the log ratio L as follows: $\sigma_L = \sqrt{\frac{\sigma_C^2}{n_C \bar{X}_C^2} + \frac{\sigma_E^2}{n_E \bar{X}_E^2}}$ (Hedges et al. 1999), where n_E and n_C are the sample sizes and σ_E and σ_C are the standard deviations of X_E and X_C . The standard deviation of the log ratio can then be used to estimate a confidence interval for L . For example, the 95% confidence interval can be approximated as follows: $CI_{95\%} = (L - 1.96\sigma_L; L + 1.96\sigma_L)$. The confidence interval of the ratio can be obtained by taking the exponential of the limits.

This procedure must be applied study-by-study. If P studies are available, we obtain a series of P ratios (R_1, R_2, \dots, R_P), a series of P log ratios (L_1, L_2, \dots, L_P), and the associated confidence intervals. The distributions of the ratios, log ratios and confidence intervals are often described graphically, on a particular type of graph known as a “forest plot”. This type of graph is illustrated in an example below.

Example: Ratio of the Yields of Organic and Conventional Cropping Systems

We consider here 65 studies used in the meta-analysis of Seufert et al. (2012). Each study includes experimental values for yield obtained with conventional and organic cropping systems. The studies cover different sites, different years and different crops. The effect size is the following ratio: organic yield/conventional yield. Yields,

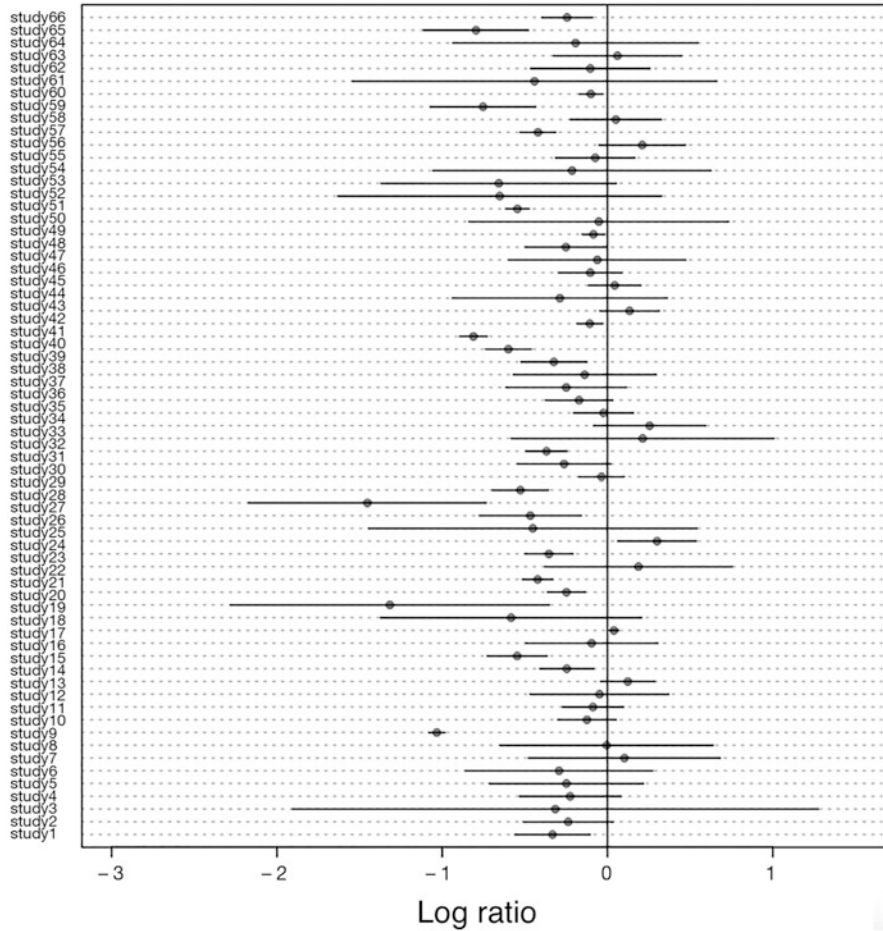


Fig. 6.2 Logarithms of estimated organic yield/conventional yield ratios for 65 studies. The horizontal bars indicate the 95% confidence intervals. The data used were obtained from the database of Seufert et al. (2012). Study 36 is missing

sample sizes and standard deviations were extracted from each study. The values were used to calculate log ratios and their 95% confidence intervals (Fig. 6.2).

6.2.3.2 Test of the Heterogeneity of Effect Size

The null hypothesis is $H_0: L_1 = L_2 = \dots = L_p$. If this hypothesis is correct, then effect size is the same in each of the studies considered in the meta-analysis. This null hypothesis is generally tested in Q tests, as described by Hedges et al. (1999) and Borenstein et al. (2009), for example. Q tests are based on the quantity Q defined as follows:

$$Q = \sum_{i=1}^P w_i (L_i - \bar{L})^2$$

$$\text{with } \bar{L} = \frac{\sum_{i=1}^P w_i L_i}{\sum_{i=1}^P w_i} \text{ and } w_i = \frac{1}{\sigma_{L_i}^2}$$

$$Q = \sum_{i=1}^P w_i L_i^2 - \frac{\left(\sum_{i=1}^P w_i L_i \right)^2}{\sum_{i=1}^P w_i}$$

According to the null hypothesis H_0 , the statistic Q is distributed according to a Chi^2 distribution with $P-1$ degrees of freedom: $Q \sim \chi_{P-1}^2$. The null hypothesis is rejected if $Q > q_{\alpha}$, where q_{α} is the quantile $1-\alpha$. If the null hypothesis is not rejected, we cannot rule out the possibility that effect size is identical in the various studies and, in this case, it is possible to estimate mean effect size with a fixed-effects model (see Sect. 6.2.3.3). If the null hypothesis is rejected, we can conclude that effect size differs between studies. In this case, it is preferable to use a random-effects model to estimate mean effect size (see Sect. 6.2.4).

Example: Ratio of Organic Yield to Conventional Yield (Continued)

The Q test was applied to the difference between the 65 log ratios presented in Fig. 6.2. The Q value obtained was 1982.31. This value is much higher than the 95th percentile of a Chi^2 distribution with 64 degrees of freedom (84.82). We can therefore reject the null hypothesis that these 65 log ratios for yield are equal. Here is the R code used:

```
#Calculation of vectors, weights and log ratios
w <- -1/TAB$Var_InR
L <- -TAB$lnR
#Calculation of the Q statistic
Q <- sum(w*(L)^2)-((sum(w*L))^2)/sum(w)
#Calculation of the p-value
1-pchisq(Q,length(L)-1)
```

6.2.3.3 Estimation of the Weighted Mean Effect Size

If we assume that the effect sizes are homogeneous, then we can consider all the studies included in the meta-analysis to have the same effect size. The true value of this effect size is unknown but can be estimated from the sample of P studies available, using the following very simple statistical model:

$$L_i = \mu + \varepsilon_i$$

where L_i is the estimated individual effect size for study i (e.g. log ratio), μ is the true effect size, ε_i is the difference between L_i and μ (residual error) and $\text{var}(L_i) = \text{var}(\varepsilon_i) = \sigma_i^2$ is the variance of the residual error. In practice, this variance is assumed to be known and is estimated from the data extracted from the studies, according to the method indicated in 6.2.3.1. The value of μ can be estimated by calculating a weighted mean, as follows:

$$\mu_{est}^{FE} = \frac{\sum_{i=1}^P w_i L_i}{\sum_{i=1}^P w_i}$$

$$w_i = \frac{1}{\sigma_i^2}$$

The variance of this estimator and its 95% confidence interval are estimated as follows:

$$\text{var}(\mu_{est}^{FE}) = \frac{1}{\sum_{i=1}^P w_i}$$

$$CI_{95\%} = \left(\mu_{est}^{FE} - 1.96\sqrt{\text{var}(\mu_{est}^{FE})}; \mu_{est}^{FE} + 1.96\sqrt{\text{var}(\mu_{est}^{FE})} \right)$$

Example: Yield Ratios for Organic and Conventional Cropping Systems (Continued)

Although the Q test is significant, we use here the fixed-effects model for illustrative purposes. The value of the log ratio estimated with the fixed-effects model described below is -0.25 . Its standard deviation is 0.0082. The estimated yield ratio is 0.78 and its 95% confidence interval is (0.76–0.79). This result reveals that the loss of yield is probably between 21 and 24%, according to the fixed-effects model. The R code used for the statistical analysis is presented below:

```
#Calculation of weights from the inverse of the variances of the log ratios
  from the TAB file
w <- 1/TAB$Var_InR
#Creation of a vector including the log ratios
L <- TAB$lnR
#Calculation of the weighted mean log ratio
MEF <- sum(w*L)/sum(w)
#Calculation of the standard error of the mean
SE <- sqrt(1/(sum(w)))
```

```
#Estimated ratio
R <- -exp(MEF)
#Calculation of the upper and lower limits of the 95% confidence interval
R_lb <- -exp(MEF-1.96*SE)
R_ub <- -exp(MEF + 1.96*SE)
```

6.2.4 Estimation of the Mean Effect Size with a Random-Effects Model

In the presence of heterogeneity, the hypothesis underlying the previous model is not realistic. We cannot consider all the studies included in the meta-analysis to have the same effect size μ . In this case, it is preferable to use a random-effects model, defined as follows:

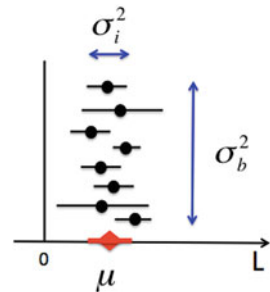
$$L_i = \mu + b_i + \varepsilon_i$$

where L_i is the estimated individual effect size for study i (e.g. log ratio), μ is the expected effect size for all the studies of the population considered, b_i is a random effect of study describing the difference in effect size between study i ($\mu + b_i$) and the expected value μ , and ε_i is the difference between L_i and $\mu + b_i$ (residual error). With this model, each study has a true individual effect size (unknown) equal to $\mu + b_i$. This model includes one fixed term (μ) and three random terms (b_i , ε_i and L_i), with variances equal to:

$$\begin{aligned} \text{var}(b_i) &= \sigma_b^2 \\ \text{var}(\varepsilon_i) &= \sigma_i^2 \\ \text{var}(L_i) &= \sigma_b^2 + \sigma_i^2 \end{aligned}$$

The principle underlying this random-effects model is presented in the diagram in Fig. 6.3.

Fig. 6.3 Diagram illustrating the principle of the random-effects model. The black points correspond to individual studies. The red point corresponds to the estimated expected effect size. The horizontal bars represent the 95% confidence intervals



As for the previous model, the first step is to calculate the individual effect sizes L_i (e.g. the log ratios) and their variances σ_i^2 from the data extracted from the individual studies. The individual sizes P are then used to estimate the two unknown parameters of the model: μ and σ_b^2 . Two methods are widely used to estimate σ_b^2 : the method of Dersimonian and Laird (the method of moments), and the restricted maximum likelihood method. The first of these methods is non-iterative and based on application of the following formula:

$$\sigma_{est,b}^2 = \frac{Q - (P - 1)}{\sum_{i=1}^P w_i - \frac{\sum_{i=1}^P w_i^2}{\sum_{i=1}^P w_i}}, \quad \text{where } w_i = \frac{1}{\sigma_i^2}$$

where Q is the same quantity as that used to test the homogeneity of effect size (see above).

The second method (restricted maximum likelihood) is iterative and can be applied with statistical software, such as the `lme` function of the `nlme` R package (Pinheiro and Bates 2000), the `lmer` function of the `lme4` R package (Bates et al. 2015), or the `rma` function of the `metafor` R package (Viechtbauer 2010). This second method can give more precise results and is increasingly frequently used in practice.

The parameter μ is estimated by calculating a weighted mean individual effect size

$$\mu_{est}^{FE} = \frac{\sum_{i=1}^P w_i L_i}{\sum_{i=1}^P w_i}$$

where $w_i = \frac{1}{\sigma_b^2 + \sigma_i^2}$. Note that the weight, w_i , is different from that used in the fixed-effects model. It depends on two variances: the variances of b_i and ε_i . If the variance of b_i tends toward zero, then the fixed-effect and random-effect models yield the same result. If the variance of ε_i tends toward zero, then the weights w_i are all approximately equal ($w_i \rightarrow \frac{1}{\sigma_b^2}$) and μ can be estimated by calculating the simple mean of individual effect size:

$$\mu_{est}^{FE} \rightarrow \frac{1}{P} \sum_{i=1}^P L_i$$

In practice, the variance of b_i is generally located somewhere between these two extreme cases.

Example: Yield Ratios for Organic and Conventional Cropping Systems (Continued)

The value of the expected log ratio estimated with the random-effects model described above is -0.25 and its standard deviation is 0.054 , when the method of Dersimonian and Laird is used. The estimated expected value is, therefore, similar to the mean effect size estimated with the fixed-effects model, but with a much larger standard deviation of the estimator. The yield ratio estimated with the random-effects model is 0.78 , with a 95% confidence interval of $(0.70-0.87)$, which is much larger than that obtained with the fixed-effects model. This result indicates that the uncertainty is underestimated when a fixed-effects model is used in the presence of strong heterogeneity. The results obtained with the random-effects model indicate that the yield loss associated with a shift from conventional to organic cropping systems is probably between 13 and 30%.

The restricted maximum likelihood method, implemented in `lme`, yields a confidence interval similar to that obtained with the method of moments $(0.73-0.85)$, and an estimated ratio identical to that obtained with the method of moments (0.78) .

Here is the R code used:

```
##Method of Dersimonian and Laird##
#Estimation of interstudy variance
Vb <- -(Q-(length(L)-1))/(sum(w)-(sum(w^2)/sum(w)))
#Calculation of weights
wT <- -1/(TAB$Var_InR + Vb)
#Estimation of the mean effect size
MEF <- -sum(wT*L)/sum(wT)
#Estimation of the standard error of the estimated mean effect size
SE <- -sqrt(1/(sum(wT)))
#Estimated ratio
R <- -exp(MEF)
#Upper and lower limits of the 95% confidence interval (95%)
R_lb <- -exp(MEF-1.96*SE)
R_ub <- -exp(MEF + 1.96*SE)
##Use of the lme function of the nlme package##
library(nlme)
#Definition of the structure of the data with the groupedData function
#The observed variable is lnR and the data are grouped according to the
  variable "Study"
Data <- -groupedData(lnR~1|Study, data=TAB)
#Estimation of parameters with lme
Fit <- -lme(lnR~1, random = ~ 1, data=Data, weight= varFixed(~Var_InR),
  method="REML")
summary(Fit)
#Estimated ratio and 95% confidence interval
R <- -exp(Fit$coefficients$fixed)
R_lb <- -exp(Fit$coefficients$fixed-1.96*sqrt(Fit$varFix))
R_ub <- -exp(Fit$coefficients$fixed + 1.96*sqrt(Fit$varFix))
```

The residual variances σ_i^2 , $i = 1, \dots, P$, are not estimated in exactly the same way with the Dersimonian and Laird method and the `lme` function of R. With the Dersimonian and Laird method, the value of each σ_i^2 is estimated as follows: $\sigma_i^2 = \sigma_{Li}^2 = \frac{\sigma_{Ci}^2}{n_{Ci}\lambda_{Ci}^2} + \frac{\sigma_{Ei}^2}{n_{Ei}\lambda_{Ei}^2}$ (Hedges et al. 1999), i.e. by a quantity calculated directly from the available data for study i . With the `lme` function and the argument `weight=varFixed`, the residual variances σ_i^2 are assumed to be proportional to the values of $\sigma_{Li}^2 = \frac{\sigma_{Ci}^2}{n_{Ci}\lambda_{Ci}^2} + \frac{\sigma_{Ei}^2}{n_{Ei}\lambda_{Ei}^2}$, i.e. $\sigma_i^2 = K\sigma_{Li}^2$, where K is a parameter estimated by `lme`. This difference in definition may partly explain the small difference in the results obtained with the DerSimonian and Laird method and those obtained with `lme`.

The restricted maximum likelihood method can be applied with $\sigma_i^2 = \sigma_{Li}^2$ (rather than $\sigma_i^2 = K\sigma_{Li}^2$) with the `rma` function of the `metafor` package (Viechtbauer 2010). The R code for the use of `rma` is as follows:

```
library(metafor)
Fit <- rma(yi=Data$lnR, vi=Data$Var_lnR, method="REML")
summary(Fit)
R <- -exp(Fit$b)
R_lb <- -exp(Fit$ci.lb)
R_ub <- -exp(Fit$ci.ub)
```

For our example, the results are identical to those obtained with `lme`. However, one of the advantages of the `metafor` package is that it can be used to perform many relevant graphical analyses. The `rma` function can also be used to estimate the mean effect size by the Dersimonian and Laird method, by replacing the argument `method="REML"` with `method="DL"`. The `forest` function of `metafor` can be used to display individual effect sizes (as in Fig. 6.2), the associated confidence intervals and the mean effect size estimated from all the studies simultaneously in a single figure (Fig. 6.4).

6.3 Meta-regression

6.3.1 Objective

The objective of meta-regression is to explain a part of the interstudy variability (called heterogeneity) of a variable of interest (generally an effect size) using one or several covariables describing the characteristics of the individual studies. As in classical regression, the covariables may be continuous (e.g. rainfall, soil depth) or discrete (e.g. variety, type of climate). The data extracted from the individual studies are used to estimate the parameters of a regression model linking the variable "effect size" to the covariables. For a continuous covariable, the model can be used to calculate the impact on effect size of increasing the covariable by one or several units. For categorical covariables, the regression model can be used to estimate the

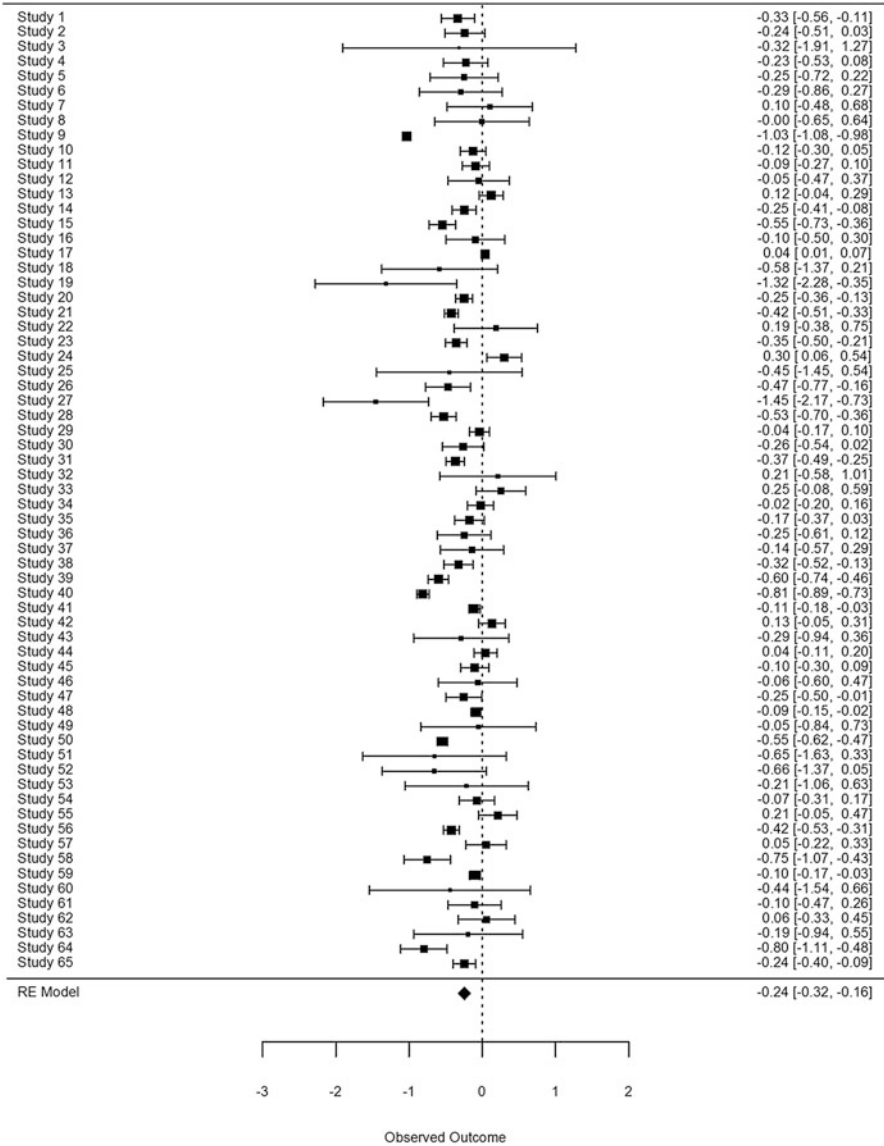


Fig. 6.4 Individual effect sizes of the 65 studies (log ratios for organic vs. conventional yields), associated confidence intervals and mean effect size estimated with a random effects model (diamond at the base of the graph, and associated confidence interval on the right). These results were generated with the `rma` function of the R package `metafor` (`method="REML"`)

value of effect size for subgroups of studies corresponding to the different categories. The regression parameters can be estimated with statistical models with and without random effects.

6.3.2 Example

We present here an example consisting of a set of virtual data including the results of eight studies with a continuous covariable (X_1) and a binary covariable ($X_2 = 0$ or $X_2 = 1$). For each of the eight studies, the estimated effect size is reported, together with its standard error (SE), a value of X_1 and a value of X_2 . This example was inspired by a real meta-analysis published by Van de Putte et al. (2010). The effect size corresponds to a ratio of yields obtained without soil tillage and after conventional soil tillage. SE is the standard error of this ratio, the covariable X_1 is water balance (in mm) and X_2 is a binary variable indicating the type of crop (0 for a winter cereal, 1 for a spring cereal).

The data file is displayed as shown below (“effect” is the yield ratio and “SE” is its standard error):

X1	X2	Effect	SE
-101	0	0.55	0.07
-75	0	0.59	0.05
-52	1	0.61	0.09
55	0	0.78	0.06
10	1	0.81	0.02
82	0	0.91	0.08
75	1	0.95	0.07
98	1	0.99	0.09

Figure 6.5a reveals the existence of considerable variability in effect size between studies. The Q statistic (Hedges et al. 1999) is 45. In a test of the null hypothesis that all the studies have the same effect size, the p -value (calculated assuming a χ^2 distribution with 7 degrees of freedom) is below 1.4×10^{-7} . The interstudy variability is, therefore, highly significant and it would be reasonable to consider the interstudy variability presented in Fig. 6.5a to reflect real variability between studies rather than estimation errors for the effect sizes of individual studies. This figure suggests that a large part of this heterogeneity can be explained by the X_1 values (Fig. 6.5b). The individual effect sizes seem to be strongly correlated with X_1 . The relationship between effect size and X_2 is less clear (Figs. 6.5c, d). Below, we show how to perform a formal test for the existence of a relationship between effect size and X_1 or X_2 by meta-regression analysis.

6.3.3 Regression Models with and Without Random Effects

Meta-regression can be performed with fixed-effects or random-effects models. A fixed-effects model is appropriate if the covariables explain all of the variability between studies. Conversely, a random-effects model should be used if only part of this variability can be explained by the covariables included in the regression model (Borenstein et al. 2009; Mergensen et al. 2013; Thompson and Higgins 2002).

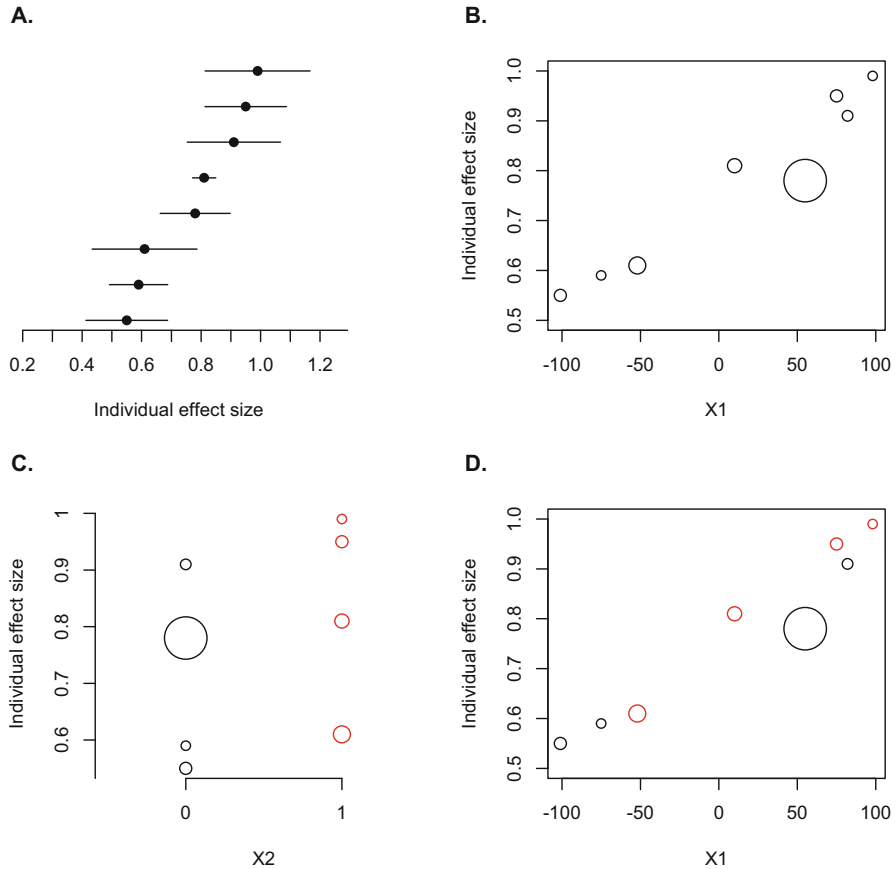


Fig. 6.5 Description of the dataset. (a): Forest plot showing the individual effect sizes and their 95% confidence intervals. (b): Bubble plot showing the individual effect sizes as a function of X_1 (circle size proportional to $1/SE$, i.e., to precision). (c): Bubble plot showing individual effect sizes as a function of X_2 . (d): Bubble plot showing individual effect sizes as a function of X_1 and X_2 ($X_2 = 0$ corresponds to the black circles, $X_2 = 1$ corresponds to the red circles)

The Gaussian linear fixed-effects regression model can be defined as follows:

$$y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ki} + \varepsilon_i, \quad \text{where } \varepsilon_i \sim N(0, \sigma_i^2).$$

β_0 is the intercept, β_k , $k = 1, \dots, p$, are the p coefficients of regression, y_i is the effect size of study i , x_{ki} is the value of the k^{th} covariable of study i , ε_i is the random residual term, and σ_i is the standard deviation of study i , $i = 1, \dots, n$. In meta-regression, intra-study standard deviations σ_i are assumed to be known and are fixed as the standard errors of the individual effect sizes obtained during the data extraction step. The parameters β_0 and β_k , $k = 1, \dots, p$, can, therefore, be estimated by the weighted least squares method, with the `glm` function of R, for example.

The Gaussian random-effects linear regression model is defined as follows:

$$y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ki} + b_i + \varepsilon_i, \quad \text{where } b_i \sim N(0, \tau^2) \text{ and } \varepsilon_i \sim N(0, \sigma_i^2).$$

This model includes a random effect of study, b_i , with a variance of τ^2 . This study effect describes the proportion of the interstudy variability not explained by the p covariables of the regression model. As for the fixed-effects model, the intra-study standard deviations, σ_i , are generally assumed to be known. By contrast, the interstudy variance, τ^2 , must be estimated from data. In practice, the variance τ^2 can be estimated by the restricted maximum likelihood method, and β_0 , β_k , $k = 1, \dots, p$, can be estimated by the generalized least squares method (Lindstrom and Bates 1988), with, for example, the lme function of R (package nlme, Pinheiro and Bates 2000).

Different approaches can be used to determine whether to use a fixed-effects model or a random-effects model. One approach involves first adjusting a fixed-effects model and testing the residual interstudy variability with, for example, a statistical test based on the sum of residual squares of the model (Borestein et al. 2009). This test is a modified version of the Q test described above. Another approach involves adjusting the two types of model and comparing them on the basis of different model choice criteria, such as Akaike's information criterion (AIC).

The random-effects model described above constitutes an extension of the model presented in Sect. 6.2.4 for classical meta-analysis. Indeed, if we remove the covariables of the model, we obtain the model defined in the framework of classical meta-analysis:

$$y_i = \beta_0 + b_i + \varepsilon_i,$$

where $b_i \sim N(0, \tau^2)$, $\varepsilon_i \sim N(0, \sigma_i^2)$, and β_0 is mean effect size. The Gaussian random-effects linear regression model described above can be made more complex by defining the regression coefficients β_k , $k = 1, \dots, p$, as random variables. In this case, the effects of the covariables are assumed to vary between studies. This amounts to considering there to be interactions between the effects of study and the effects of the covariables.

The Gaussian linear model presented above can be used in many situations, but is not always appropriate for analyses of effect sizes defined on the basis of count data (e.g. disease incidence, presence/absence of a pest organism on an agricultural plot), such as odds ratios. In this case, the use of generalized linear models (with or without random effects) is recommended (see, for example, Makowski et al. 2014b). The Gaussian linear model may also be inappropriate if the covariables are related to effect size in a non-linear manner. In this case, it is preferable to use non-linear regression models, with or without random effects (see, for example, Philibert et al. 2012b).

6.3.4 Example (Continued)

Here, we will develop meta-models to explain part of the variability of the individual effect sizes presented in Figure 6.5a, using the covariables X_1 and X_2 . Three fixed-effects models are first adjusted for the data:

- Model M1 linking effect size to X_1 , defined by $y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$,
- Model M2 linking effect size to X_2 , defined by $y_i = \beta_0 + \beta_2 x_{2i} + \varepsilon_i$,
- Model M3 linking effect size to X_1 and X_2 , defined by $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$,

The results obtained for the three models are presented in Fig. 6.6a, b, c. Figure 6.6a shows the result of the adjustment of M1. The estimated value of β_1 is 0.002 ($p < 0.01$), indicating that the effect size increases by 0.002 units for each additional unit of X_1 (water balance). As X_2 is binary, the result of the adjustment of M2 is not a curve, but two mean effect size values, one for $X_2 = 0$ (equal to β_0) and one for $X_2 = 1$ (equal to $\beta_0 + \beta_2$) (Fig. 6.6b). The difference between these two values is not significant ($p > 0.1$). Model M3 generates two regression lines, one for $X_2 = 0$ and the other for $X_2 = 1$ (Fig. 6.6c). These two lines have a slope β_1 of 0.0019 ($p < 0.01$) and, with this model, the effect of covariable X_2 is significant ($p < 0.05$).

A statistical test based on the weighted sum of the residuals is performed for each model, to determine whether a proportion of the interstudy variability remains unexplained by the covariables. The p -values for M1, M2 and M3 are 0.6, 3.2×10^{-5} , and 0.87, respectively. Thus, M1 and M3 can explain almost all the interstudy variability, whereas M2 is not able to do so.

These results indicate that the use of a random-effects model is not entirely justified in this case when X_1 is taken into account. However, for the purposes of illustration, we have also developed a random-effects model (M4) including the covariable X_2 and defined by $y_i = \beta_0 + \beta_2 x_{2i} + b_i + \varepsilon_i$, where $b_i \sim N(0, \tau^2)$ and $\varepsilon_i \sim N(0, \sigma_i^2)$. Like M2, model M4 generates two mean values for effect size, one for $X_2 = 0$ (equal to β_0) and one for $X_2 = 1$ (equal to $\beta_0 + \beta_2$) (Fig. 6.6d). These two values are very similar to those obtained with M2 (Fig. 6.6b).

Here is the R code used:

```
#Adjustment of the fixed-effects models with the glm function
#W is a vector including the weights associated with the Y data (inverse of
  the variance)
W <- 1/V
mod4 <- glm(Y~X2,weights=W)
summary(mod4)
#Adjustment of the random-effects models with the lme function of the nlme
  package
Data <- groupedData(Y~X1 + X2|Studies)
mod4RE <- lme(Y~X2, random=~1,weights=varFixed(~V), data=Data)
summary(mod4RE)
```

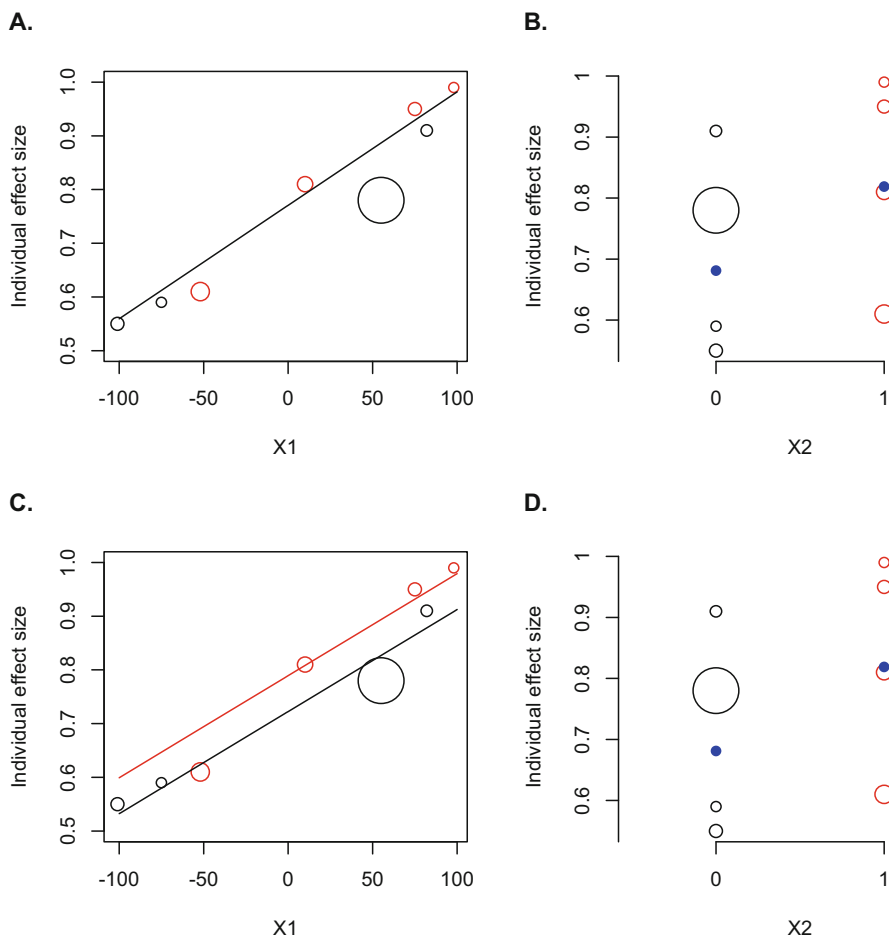


Fig. 6.6 Meta-models adjusted for the data from eight studies. Bubble plots showing individual effect sizes as a function of the covariable X_1 (**a**, **c**), and the covariable X_2 (**b**, **c**). Circle size is proportional to $1/SE$. $X_2 = 0$ corresponds to the black circles, $X_2 = 1$ corresponds to the red circles. (**a**): The black line corresponds to the fixed-effects meta-model including covariable X_1 . (**b**): The blue points are the estimated mean effect sizes for $X_2 = 0$ and $X_2 = 1$ based on a fixed-effects model. (**c**): The black and red lines show the fixed-effects meta-model including both X_1 and X_2 ($X_2 = 0$ in black, $X_2 = 1$ in red). (**d**): The blue points indicate the estimated mean effect sizes for $X_2 = 0$ and $X_2 = 1$ with a random-effects regression model

As indicated in Sect. 6.2.4, it is difficult to set predefined residual variances with the `lme` function. When the argument `weight=varFixed(~V)` is used in `lme`, the residual variances σ_i^2 are assumed to be proportional to the variance values of the individual studies, but not strictly equal to those values. It is possible to fix the residual variances at predefined values with the `rma` function of the `metafor` package, with the following R code:


```
Mod4rma <- rma(yi=Y, vi=V, mods=~factor(X2),method="REML")
summary(mod2rma)
```

The results obtained with `lme` and `rma` are very similar. The estimated values of β_0 and β_2 are 0.68 and 0.14, respectively, with `lme`, and 0.70 and 0.14, respectively, with `rma` (REML method). The effect of X_2 is non-significant with both `lme` and `rma`.

6.4 Critical Analysis of the Results

The critical analysis of the results includes several aspects. The first relates to the choice of the studies. It is important to ensure that the systematic bibliographic review was correctly performed. In particular, the study search and selection procedures must be described in a transparent manner, and the selection criteria used must not be likely to bias the conclusions. Certain scientific journals (e.g. PLoS One and Environmental Evidence) have requirements in terms of the traceability of the studies selected. These high standards help to increase the quality of systematic reviews and of the meta-analyses performed by scientists.

Even if the systematic review was well-performed, some studies presenting particular characteristics may have been excluded by the publication process. Such exclusions can generate a specific bias often referred to as *publication bias*. This bias exists when, due to imperfections in the process of dissemination for research results, the effect sizes taken into account in the meta-analysis lead to conclusions different from those that would have been drawn if all the correctly calculated effect sizes had been taken into account (Duval and Tweedie 2000; Koricheva et al. 2013). Such a bias can lead, in particular, to the non-publication of results that are not statistically significant. These missing data can result from self-censoring by the authors or from a rejection of non-significant results by scientific journals (contentious, but unfortunately quite common behavior).

One approach frequently used to detect the existence of a publication bias involves the plotting of a particular type of graph called a funnel plot. This type of graph represents the effect sizes of individual studies on the x axis, and the level of precision associated with the effect sizes (the inverse of sample size, inverse of variance or inverse of the standard error of effect sizes) on the y axis. In a funnel plot, the most precise studies are shown at the top, and the least precise appear at the bottom. In the absence of a publication bias, the graph is symmetric around the mean effect size. By contrast, a publication bias would tend to result in an asymmetric funnel plot characterized by missing effect sizes at the bottom of the graph, generally corresponding to the side opposite to the effect of interest. Thus, if the effect sought is expected to be positive, a meta-analysis presenting a publication bias would be characterized by a funnel plot for which the most extreme effect sizes would be located at the bottom on the right, whereas some of the least extreme effect sizes would be missing at the bottom on the left. Such a bias would lead to an

overestimation of mean effect size. This asymmetry can be assessed statistically with Egger's method (Koricheva et al. 2013). This method involves adjusting the regression between a normalized effect size (effect size divided by its standard error) and the precision of effect size (inverse of the standard error of effect size). A publication bias is detected if the intercept value of the regression line is significantly different from zero. Another approach consists in estimating the number of missing studies that might exist in a meta-analysis (Duval and Tweedie 2000).

Example: Yield Ratios for Organic and Conventional Cropping Systems (Continued)

Figure 6.7 shows a funnel plot describing effect size (log ratio, x axis) and the precision of effect size (inverse of the standard error, y axis). The relationship between normalized effect size and precision is not significant ($p > 0.1$). The calculations were performed with the following R code (note that funnel plots can also be drawn with the function `funnel` of the R package `metafor`):

```
LogRatio <- -TAB$lnR
Precision <- 1/sqrt(TAB$Var_InR)
plot(LogRatio,Precision, xlab="log ratio", ylab="Precision")
abline(v=Fit$coefficients$fixed)
LRnorm <- -LogRatio*Precision
summary(lm(LRnorm~Precision))
```

In addition to attending to the points mentioned above, care is required in evaluations of the validity of the hypotheses underlying the statistical models used. The question of the choice between a fixed-effects and a random-effects

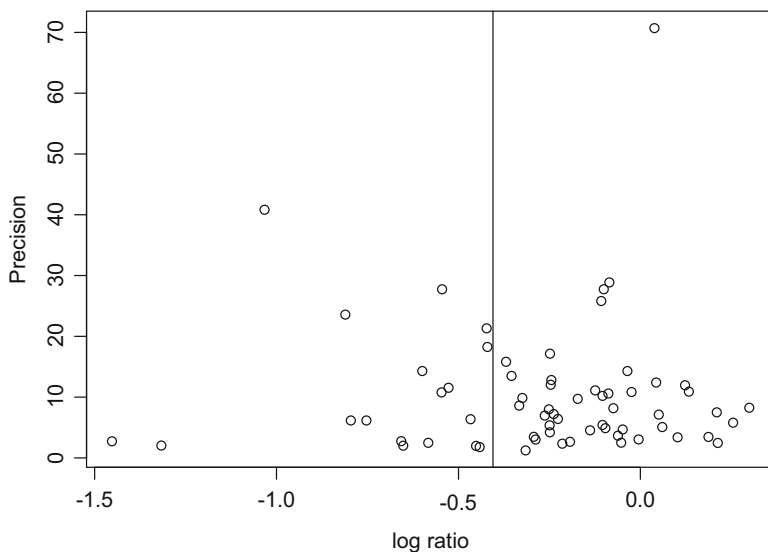


Fig. 6.7 Funnel plot for the yield ratio example. Each point corresponds to a study. The vertical bar indicates the mean effect size estimated with the `lme` function of R

model is often posed in meta-analyses. The Q test presented above provides useful elements for justifying this choice. In addition to the Q test, it is recommended to compare the results of different types of models, to determine whether the choice of a particular model is likely to affect the conclusions of the meta-analysis (Hossard et al. 2016). The AIC can also be used to determine what kind of statistical model to use. In addition to criteria of this type, the choice of model must also take into account the context in which the meta-analysis is being performed. If strong heterogeneity between studies is suspected, it may be preferable to use a random-effects model by default for quantitative analyses of the variability between studies.

An analysis of the residuals is strongly recommended when a meta-regression is performed. This type of analysis can be used to evaluate the validity of the hypotheses underlying the regression model adjusted for the data, particularly those concerning the linearity of the data and the distribution of residuals. It is also recommended to check for the existence of influential data or studies. Jackknife approaches are useful in this respect (Philibert et al. 2012a). In such approaches, the studies (or data) are removed one-by-one, with re-estimation of the quantities of interest without the study (or datum) concerned.

Philibert et al. (2012a) proposed a list of eight criteria for the analysis of meta-analysis quality:

- Presentation of a repeatable study selection procedure
- Presentation of a complete list of the studies selected
- Analysis of the interstudy variability of effect size
- Analysis of the sensitivity of the conclusions to the statistical method and data used
- Studies weighted according to their precision
- Analysis of publication bias
- Accessibility of the data
- Accessibility of the computer code used.

This list makes it possible to check rapidly that the principal quality criteria are satisfied when performing a meta-analysis.

References

- Anses (2016) Évaluation du poids des preuves à l'Anses: revue critique de la littérature et recommandations à l'étape d'identification des dangers. Rapport du groupe de travail « Méthodes pour l'évaluation des risques »
- Bates D, Maechler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67:1–48
- Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009) Introduction to meta-analysis. Chapter 20. Wiley, Chichester
- Chalmers I, Hedges LV, Cooper H (2002) A brief history of research synthesis. *Eval Health Prof* 25 (1):12–37

- Duval S, Tweedie R (2000) Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56:455–463
- Efsa (2010) Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA J* 8(6):1637
- Glass GV (1976) Primary, secondary and meta-analysis of research. *Educ Res* 10:3–8
- Hedges LV, Olkin I (1985) *Statistical methods for meta-analysis*. Academic, Orlando
- Hedges LV, Gurevitch J, Curtis P (1999) The meta-analysis using response ratios in experimental ecology. *Ecology* 80:1150–1156
- Hossard L, Archer DW, Bertrand M, Colnenne-David C, Debaeke P, Erfors M, Jensen ES, Jeuffroy MH, Munier-Jolain N, Nilsson C, Sanford GR, Snapp SS, Makowski D (2016) A meta-analysis of maize and wheat yields in low-input vs. conventional and organic systems. *Agron J* 108:1155–1167
- Ioannidis JPA (2005) Why most published research findings are false. *PLoS Med* 2(8):e124
- Koricheva J, Gurevitch J, Mengersen K (eds) (2013) *Handbook of meta-analysis in ecology and evolution*. Princeton University Press, Princeton
- Lajeunesse MJ (2016) Facilitating systematic reviews, data extraction and meta-analysis with the metagear package for R. *Methods Ecol Evol* 7:323–330
- Lindstrom MJ, Bates DM (1988) Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J Am Stat Assoc* 83:1014–1022
- Makowski D, Nesme T, Papy F, Doré T (2014a) Global agronomy, a new field of research. *Agron Sustain Dev* 34:293–307
- Makowski D, Vicent A, Pautasso M, Stancanelli G, Rafoss T (2014b) Comparison of statistical models in a meta-analysis of fungicide treatments for the control of citrus black spot caused by *Phyllosticta citricarpa*. *Eur J Plant Pathol* 139:79–94
- Mengersen K, Schmid CH, Jennions MD, Gurevitch J (2013) Statistical models and approaches to inference. In: Koricheva J, Gurevitch J, Mengersen K (eds) *Handbook of meta-analysis in ecology and evolution*. Princeton University Press, Princeton, pp 89–108
- Pearson K (1904) Report on certain enteric fever inoculation statistics. *Br Med J* 3:1243–1246
- Philibert A, Loyce C, Makowski D (2012a) Assessment of the quality of the meta-analysis in agronomy. *Agric Ecosyst Environ* 148:72–82
- Philibert A, Loyce C, Makowski D (2012b) Quantifying uncertainties in N₂O emission due to N fertilizer application in cultivated areas. *PLoS One* 7(11):e50950. <https://doi.org/10.1371/journal.pone.0050950>
- Pinheiro JC, Bates DM (2000) *Mixed-effects models in S and S-PLUS*. Springer, New York
- Seufert V, Ramankutty N, Foley JA (2012) Comparing the yields of organic and conventional agriculture. *Nature* 485:229–232
- Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F (2000) *Methods for meta-analysis in medical research*. Wiley, New York
- Thompson SG, Higgins JPT (2002) How should meta-regression analyses be undertaken and interpreted. *Stat Med* 21:1559–1573
- Van den Putte A, Govers G, Diels J, Gillijns K, Demuzere M (2010) Assessing the effect of soil tillage on crop growth: a metaregression analysis on European crop yields under conservation agriculture. *Eur J Agron* 33(3):231–241
- Viechtbauer W (2010) Conducting meta-analyses in R with the metafor package. *J Stat Softw* 36: 1–48. <http://www.jstatsoft.org/v36/i03/>
- Yates F, Crowther EM (1941) Fertilizer policy in wartime: The fertilizer requirements of arable crops. *Empire J Exp Agric* 9:77–97
- Yu Y, Stomph T-J, Makowski D, van der Werf W (2015) Temporal niche differentiation increases the land equivalent ratio of annual intercrops: a meta-analysis. *Field Crop Res* 184:133–144

Chapter 7

Statistical Problems Specific to Meta-analysis



7.1 Definition of Effect Size

The definition of effect size and its estimation study-by-study are important steps in meta-analysis. In Chap. 6, effect size was defined as the log ratio of the mean for observations for treatment E and for the control C . We present here other effect sizes that may prove useful in certain contexts, and we explain how they can be estimated, study-by-study.

7.1.1 Correction of Biases Linked to the Use of Ratios

The log ratio of the mean for the experimental treatment E and the mean for the control C is frequently used in meta-analysis. As indicated in Chap. 6, the classic estimator of this effect size is $L = \ln\left(\frac{\overline{X_E}}{\overline{X_C}}\right)$ and the variance of the estimator can be estimated as $\sigma_L^2 = \frac{\sigma_C^2}{n_C \overline{X_C}^2} + \frac{\sigma_E^2}{n_E \overline{X_E}^2}$.

For small sample sizes (i.e. when only small numbers of data are available for calculating each mean), these estimators are biased. It is possible to correct this bias by using variants of the classic estimators, as defined by Lajeunesse (2015):

$$L_b = \ln\left(\frac{\overline{X_E}}{\overline{X_C}}\right) + \frac{1}{2} \left[\frac{(\sigma_E^2)}{n_E \overline{X_E}^2} - \frac{(\sigma_C^2)}{n_C \overline{X_C}^2} \right], \text{ with } \sigma_{L_b}^2 = \sigma_L^2 + \frac{1}{2} \left[\frac{\sigma_E^4}{n_E^2 \overline{X_E}^4} + \frac{\sigma_C^4}{n_C^2 \overline{X_C}^4} \right].$$

It is, however, less straightforward to interpret the estimator L_b than to interpret L , because L_b is also dependent on the variance of the observations and sample size. A value of L_b greater than 1 does not necessarily indicate that the mean of the observations for treatment E is greater than that for the control C .

Another problem concerns the correlations between log ratios based on the same mean \overline{X}_C . This situation arises when at least two experimental treatments are compared with the same control in an experiment. These log ratios cannot be considered independent if they share the same control. It is, therefore, useful to take into account the correlation between log ratios based on the same control in the statistical analysis. The methods proposed by Lajeunesse (2011) make it possible to do this. They can be applied with the assistance of the `covariance_commonControl()` function of the R package `metagear` (Lajeunesse 2016).

7.1.2 *Difference Between Observation Means*

In addition to the log ratio presented in Chap. 6, the effect of a treatment on a continuous variable (e.g. yield) can be measured by calculating the difference between the mean of observations for treatment E and the mean of observations for the control C , in other words, by calculating $D = \overline{X}_E - \overline{X}_C$. If the theoretical standard deviation for measurements of X is the same for C and E , then the standard deviation of D can be estimated as $S = \sqrt{\frac{(n_E-1)S_E^2 + (n_C-1)S_C^2}{n_E+n_C-2}}$, where S_E and S_C are the standard deviations of measurements of X for E and C , respectively.

The use of D is justified when all the data of the various studies are expressed at the same scale. If this is not the case, it is preferable to normalize D by dividing it by S , and, thus, to use the effect size d defined as $d = D/S$. Unlike D , the quantity d has no units. Its standard deviation can be expressed as $S_d = \frac{n_E+n_C}{n_E n_C} + \frac{d^2}{2(n_E+n_C)}$.

The quantity d defined above is biased (it overestimates the treatment effect if the sample size is small). This bias can be corrected by multiplying d by a corrective factor. This correction yields an effect size generally referred to as g (Hedges and Olkin 1985) and defined as $g = Jd$, where $J = 1 - \frac{3}{4(n_E+n_C-2)-1}$. The standard deviation of g is estimated as $S_g = JS_d$. However, the interpretation of g is not straightforward, because its value depends on the observation means, their standard deviations and sample size.

7.1.3 *Effect Sizes for Binary and Count Data*

Count data, corresponding to the number of occurrences of an event of interest, are widely used. This is the case, in particular, when the objective is to analyze the results of experiments testing one or several treatments for the control of a disease or pest. In this case, the data for each experiment are presented as a contingency table of the type presented in Table 7.1.

Table 7.1 Hypothetical results of an experiment to test the effect of a treatment on the number of diseased plants

	Control (untreated plants)	Experimental treatment (treated plants)	
Number of healthy plants	$A = 5$	$B = 19$	$A + B = 24$
Number of diseased plants	$C = 15$	$D = 1$	$C + D = 16$
	$n_C = A + C = 20$	$n_E = B + D = 20$	

It is possible to calculate different effect sizes to measure the efficacy of treatment from the data presented in Table 7.1. For example, it is possible to estimate the difference in the proportion of diseased plants, defined as $DP = P_E - P_C = \frac{D}{n_E} - \frac{C}{n_C}$, and its standard deviation, $\sigma_{DP} = \sqrt{\frac{P_E(1-P_E)}{n_E} + \frac{P_C(1-P_C)}{n_C}}$.

It is also possible to calculate the logarithm of the ratio of proportions, defined as $LRP = \ln(P_E/P_C) = \ln\left(\frac{D/n_E}{C/n_C}\right)$ and its standard deviation, $\sigma_{LRP} = \sqrt{\frac{P_E(1-P_E)}{n_E P_E} + \frac{P_C(1-P_C)}{n_C P_C}}$.

Odds ratios (*OR*) are another type of effect size often used to measure the efficacy of a treatment. The odds ratio is the ratio of the odds for the treatment *E* to the odds for the control *C*: $OR = \frac{O_E}{O_C}$ where $O_E = \frac{P_E}{(1-P_E)}$ and $O_C = \frac{P_C}{(1-P_C)}$.

For example, with the data from Table 7.1, $P_E = 0.05$, $P_C = 0.75$, $O_E = 0.053$, and $O_C = 3$. This means that, in the control group, a plant is three times more likely to be diseased than to be healthy. The odds in the experimental group (*E*) are much lower, just over 0.05. We can also say that a plant has a 19 times higher chance (0.95/0.05) of not being diseased than of being diseased in the treatment group (*E*). The odds ratio is equal to $0.053/3 = 0.018$ and thus lower than one, which indicates that a plant has a markedly lower chance of being diseased in the treatment group (*E*) than in the control group (*C*).

In general, we work with the logarithm of the *OR*. The standard deviation of log *OR* (*LOR*) can be calculated in several ways, the simplest being $\sigma_{LOR} = \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}$.

In a meta-analysis, the effect sizes presented above and their standard deviations must be calculated study-by-study. The results can be presented as forest plots, as explained in Chap. 6, and it is then possible to estimate the mean effect size with fixed-effects or random-effects models, depending on the heterogeneity between studies.

7.1.4 Correlation Coefficient

In some meta-analyses, it is useful to define the effect size as a coefficient of correlation between two variables. This is the case when the objective is to quantify

the strength of a relationship between two variables studied in different individual studies. Each study reports a correlation coefficient estimated from a certain number of data, and the meta-analysis makes it possible to combine the various correlation coefficients.

Several precautions are required for the analysis of correlation coefficients. In particular, when their values are close to one, their distribution tends to deviate strongly from the Gaussian. For this reason, it is preferable to transform them by calculating the following quantity (known as Fisher's z): $z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right)$, where r is Pearson's correlation coefficient. The standard deviation of z can be estimated as $\sigma_z = \frac{1}{n-3}$, where n is the number of data used to calculate r (Sokal and Rohlf 1995). The values of z and σ_z can be analyzed with the models presented in Chap. 6.

7.1.5 Effect Sizes Based on Variance

Most meta-analyses study the effect of treatments on the mean response of a variable of interest. However, several recent studies have highlighted the value of not concentrating solely on studies of mean values, but of also studying the effect of treatments on variance (Nakagawa et al. 2015; Lesur et al. 2017). For example, Lesur et al. (2017) studied the differences in the interannual variability of yield between organic and conventional cropping systems, with the aim of determining whether organic systems resulted in higher or lower levels of yield stability.

Various metrics can be used for this purpose, including, in particular, the log ratio of variances or the log ratio of coefficients of variation. The calculation and weighting of these quantities were described by Nakagawa et al. (2015). It is also possible to take the relationship between variance and mean into account with the following statistical model based on the Taylor's power law:

$$\ln(\hat{\sigma}_{ij}) = \beta_0 + b_{0i} + (\beta_1 + b_{1i})X_{ij} + \beta_2 \ln(\mu_{ij}) + \varepsilon_{ij}$$

where $\hat{\sigma}_{ij}$ is the estimated standard deviation of the response variable studied in treatment j of study i ($j = 1$ for the control C , $j = 2$ for the experimental treatment E), μ_{ij} is the mean value of the response variable, β_0 , β_1 , and β_2 are fixed effects, b_{0i} and b_{1i} are random study effects, X_{ij} is a binary variable (equal to zero for the control C and equal to one for the experimental treatment E), and ε_{ij} is the residual error. An unbiased estimate of $\ln(\hat{\sigma}_{ij})$ can be obtained by calculating $\ln(\hat{\sigma}_{ij}) = \ln(s_{ij}) + \frac{1}{2(n-1)}$, where s is the empirical standard deviation and n is sample size. This estimate must be calculated for each study i and treatment j , and weighted for its variance $s_{\ln(\hat{\sigma}_{ij})}^2 = \frac{1}{2(n-1)}$. This model is a mixed meta-regression model (including both fixed and random effects). It can be fitted to the set of values of $\ln(\hat{\sigma}_{ij})$ calculated for all studies with the procedures described in Chap. 6 using the nlme or metafor R packages. The estimated value of β_2 measures the effect of the treatment on the variance of the response; a positive (negative) estimated value

indicates that the treatment increases (decreases) the variance. Several variants of this model can be developed to take into account the interaction between the effect of the treatment and the mean effect, for example.

7.2 Generalized Linear Models for the Analysis of Discrete Data

Generalized linear models generalize those presented in Chap. 6, for the analysis of data with a non-Gaussian distribution, including discrete and count data in particular (Agresti 2002). We raised the issue of this type of data in Sect. 7.1.3 and we have already presented several quantities that can be computed from count data in Table 7.1.

For data corresponding to measurements of proportions, such as those presented in Table 7.1, one widely used approach involves calculating odds ratios for each individual study and then analyzing all the odds ratios together with a Gaussian model of the linear mixed model type. However, the Gaussian approximation is not always realistic, particularly when the observed proportions are close to zero or one. In such cases, it is preferable to analyze all the data with a generalized linear model.

Below, we present a particular type of generalized linear model: the binomial logit random-effects model. This model is useful for analyses of the effects of treatments on the proportion of diseased individuals (plants, animals, humans) (Makowski and Monod 2011; Makowski et al. 2014), and for analyses of epidemiological surveillance networks (Michel et al. 2016, 2017).

7.2.1 Binomial Logit Random-Effects Model for Analyzing the Effect of a Treatment

This model has two levels. The first describes the distribution of count data within a study (e.g. measurements of the proportion of diseased plants in a plot), and the second describes the variability of the true proportion between studies (e.g. variability of the true proportion of diseased plants between plots) as a function of treatment type. The model can be expressed formally in the following fashion:

$$Y_{ij} \sim \text{Binomial}(N_{ij}, \pi_{ij}) \quad i = 1, \dots, I, j = 1, 2 \quad (7.1)$$

$$\text{logit}(\pi_{ij}) = \ln \left(\frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \alpha_i + \beta_i X_{ij}, \quad \text{with} \quad \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \Sigma \right] \quad (7.2)$$

where N_{ij} is the number of measurements collected for the j^{th} treatment of the i^{th} study, Y_{ij} is the number of positive cases (e.g. the number of diseased plants among the N_{ij} plants observed in the i^{th} study), π_{ij} is the true proportion (unknown) of positive cases in the i^{th} study, X_{ij} is a binary variable equal to 1 if Y_{ij} was collected in

the experimental treatment and equal to zero if Y_{ij} was collected in the control plot (untreated), α_i and β_i are the parameters determining the value of π_{ij} for the control and treatment in the i^{th} study, respectively, μ_α and μ_β are the expected values of the parameters α_i and β_i (the “mean” values for all studies), and Σ is the two-by-two variance-covariance matrix of α_i and β_i (which determines the between-study variability of these parameters).

In this model, the odds ratio for the treatment relative to the control is $\exp(\beta_i)$ for the i^{th} study, and the mean odds ratio for all studies is $\exp(\mu_\beta)$. The parameters of model (7.1 and 7.2) can be estimated by maximum likelihood methods, taking into account all the measurements from all the studies, with the `glmer` function of the R package `lme4`, for example. The estimated values of the parameters can then be used to estimate the odds ratios, both study-by-study, and as a mean value for all studies. It is also possible to use the model to estimate the incidence study-by-study $\frac{\exp(\alpha_i + \beta_i X)}{1 + \exp(\alpha_i + \beta_i X)}$ and the global incidence $\frac{\exp(\mu_\alpha + \mu_\beta X)}{1 + \exp(\mu_\alpha + \mu_\beta X)}$.

Several variants of model (7.1 and 7.2) can be defined. It is, for example, possible to include different types of covariable X , describing the characteristics of the environment (variety, soil type, climatic conditions etc.), for example (Michel et al. 2016, 2017). It is also possible to include additional random effects, to distinguish site and year effects, for example (Michel et al. 2016, 2017). Finally, it is sometimes useful to replace the binomial probability distribution by a Poisson distribution when the measurements are not proportions but counts without fixed upper limit (e.g. the number of insects observed on a plot).

7.2.2 Example

The objective of this example is to estimate the efficacy of a fungicide treatment for controlling *Phyllosticta citricarpa*, a fungus that infects citrus crops (Makowski et al. 2014). For this purpose, we will use the results of 16 trials performed in different citrus groves located in different regions of the world. Each trial included two treatments: a control treatment (untreated, $X = 0$) and a fungicide treatment ($X = 1$). For each treatment, 300 to 2000 fruits were observed, and the number of diseased fruits was determined.

The proportions of diseased fruits, the odds and odds ratios were calculated from the observations, as indicated in Sect. 7.1.3. The results are presented in Fig. 7.1. According to the trial results, between 14 and 94% of the fruits in the untreated zones, and between 0 and 60% of those in the treated zones were diseased. The odds are almost all greater than one for the controls (indicating a higher probability of obtaining a diseased fruit than a healthy fruit), whereas they are almost all below one in the treated zones (indicating a higher probability of obtaining a healthy fruit than a diseased fruit).

The fungicide treatment systematically decreased the proportion of diseased fruits (the odds ratios are all markedly lower than one on Fig. 7.1d), but the efficacy of

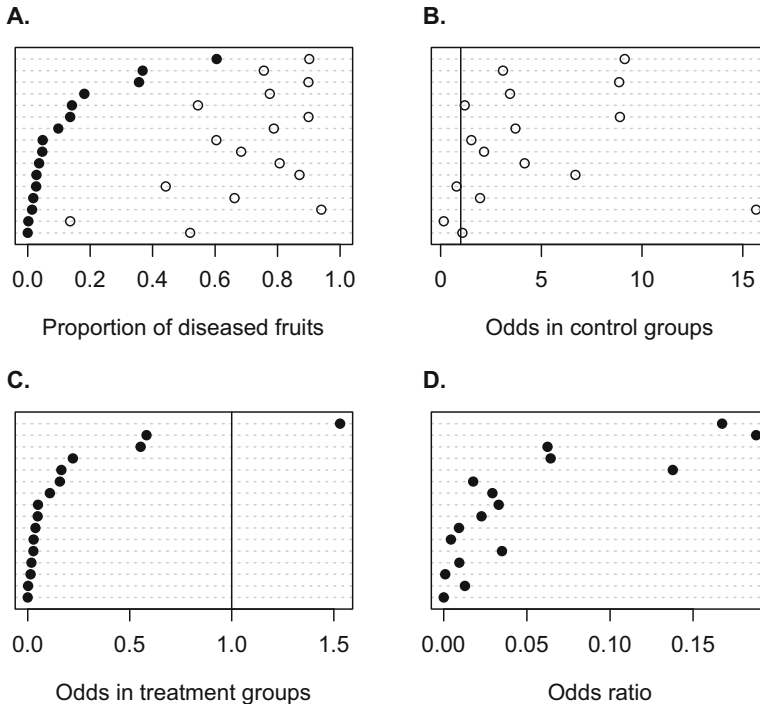


Fig. 7.1 Proportions of diseased fruits in the control (white points) and fungicide-treated (black points) zones (a), odds for the control zones (b), odds for the treated zones (c), and odds ratios (d). Each horizontal line on each graph corresponds to a trial including an untreated zone (control) and a zone treated with fungicide to control *Phyllosticta citricarpa*. Odds = proportion of diseased fruits/proportion of healthy fruits. Odds ratio = odds in the treated zone/odds in the untreated zone

fungicide treatment appears to vary considerably between situations. In some cases, the proportion of diseased fruits was zero or close to zero, whereas, in others, it remained high (Fig. 7.1a).

We will use model (4.1 and 4.2) here to estimate the mean efficacy of fungicide treatment. The model was adjusted for the data from the 16 trials with the `glmer` function of the R package `lme4`, with the following code:

```
Mod <- glmer(cbind(NbDiseasedF, NbFruits-NbDiseasedF) ~ Fung_Gp + (1
+ Fung_Gp | Code), family=binomial, data=DataSetT)
summary(Mod)
```

The results obtained with the `summary()` function are presented below:

```
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: cbind(NbDiseasedF, NbFruits - NbDiseasedF) ~ Fung_Gp + (1 +
```

```

Fung_Gp | Code)
Data: DataSetT

    AIC   BIC   logLik   deviance df.resid
372.9  380.2  -181.4  362.9    27

Scaled residuals:
Min      1Q      Median      3Q      Max
-1.02235 -0.08930   0.00707  0.03915  0.45596

Random effects:
Groups Name      Variance Std.Dev. Corr
Code (Intercept)  1.241   1.114
Fung_Gp          2.317   1.522  -0.02
Number of obs: 32, groups: Code, 16

Fixed effects:
      Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.0221     0.2798   3.653  0.000259 ***
Fung_Gp     -3.8618     0.3932  -9.822 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
(Intr)
Fung_Gp -0.028

```

The values indicated in the *Random effects* section correspond to the values of two variances, standard deviations and the correlation of the random effects α_i and β_i (i.e. the content of the matrix Σ). The values indicated in the *Fixed effects* section correspond to estimates of μ_α and μ_β , the standard deviations of the estimators and the results of tests of equality to zero for μ_α and μ_β . These tests clearly indicate, in this case, that the estimated values are significantly different from zero. The value of -3.8618 corresponds to the log odds ratio. The negative sign indicates that the fungicide treatment decreases the risk of disease. The odds ratio is equal to the exponential of -3.8618 , i.e. 0.021. Its 95% confidence interval can be calculated with the following R code:

```

exp(fixef(Mod)[2]-1.96*sqrt(vcov(Mod)[4]))
exp(fixef(Mod)[2]+ 1.96*sqrt(vcov(Mod)[4]))

```

We obtain a 95% confidence interval of [0.0097–0.045] for this example. As this interval does not include the value 1, the odds ratio can be considered to be significantly lower than one, confirming the ability of the fungicide to decrease the incidence of the disease. It is also possible to estimate the mean proportion of diseased fruits in the treated and untreated zones, and their associated 95% confidence intervals:

```

###Untreated
#Estimated incidence
exp(fixef(Mod)[1])/(1 + exp(fixef(Mod)[1]))
#Lower limit of the 95% CI
exp(fixef(Mod)[1]-1.96*sqrt(vcov(Mod)[1]))/(1 + exp(fixef(Mod)[1]-1.96*sqrt
  (vcov(Mod)[1])))
#Upper limit of the 95% CI
exp(fixef(Mod)[1] + 1.96*sqrt(vcov(Mod)[1]))/(1 + exp(fixef(Mod)[1]
  + 1.96*sqrt(vcov(Mod)[1])))

###Treated
#Variance of the estimated expected logit value
VAR <- -vcov(Mod)[1] + vcov(Mod)[4] + 2*vcov(Mod)[2]
#Estimated incidence
exp(fixef(Mod)[1] + fixef(Mod)[2])/(1 + exp(fixef(Mod)[1] + fixef(Mod)[2]))
#Lower limit of the 95% CI
exp(fixef(Mod)[1] + fixef(Mod)[2]-1.96*sqrt(VAR))/(1 + exp(fixef(Mod)[1]
  + fixef(Mod)[2]-1.96*sqrt(VAR)))
#Upper limit of the 95% CI
exp(fixef(Mod)[1] + fixef(Mod)[2] + 1.96*sqrt(VAR))/(1 + exp(fixef(Mod)[1]
  + fixef(Mod)[2] + 1.96*sqrt(VAR)))

```

The estimated proportion of diseased plants is 0.74 [0.62, 0.83] for the untreated zone and 0.06 [0.02, 0.13] for the treated zone. We would therefore expect there to be, on average, about 2–13% diseased fruits after the application of a fungicide treatment. Although the treatment is able to reduce the disease incidence substantially, it cannot completely control the disease.

7.3 Non-linear Mixed Models

7.3.1 *Utility and Definition*

Non-linear mixed models are another type of extension of the linear mixed models presented in Chap. 6. Like linear models, they can be used to analyze measurements of a continuous response variable, but non-linear models can also be used to take the non-linear relationships between one or several explanatory variables X and the response variable of interest Y , into account when modeling. If linear models provide an unrealistic description of the relationship between X and Y , it may be interesting to use a non-linear model. Such models can be formulated in different ways. We provide a simple formulation below:

$$Y_{ij} = f(X_{ij}, \theta_i) + \epsilon_{ij} \quad i = 1, \dots, I; j = 1, \dots, n_i \quad (7.3)$$

$$\theta_i \sim N(\mu, \Sigma) \quad \epsilon_{ij} \sim N(0, \sigma_\epsilon^2) \quad (7.4)$$

where Y_{ij} is the j^{th} measurement obtained in the i^{th} study, $f()$ is a non-linear function, θ_i is a parameter vector specific to the i^{th} study, X_{ij} is a variable vector (for continuous or discrete variables) with an effect on the response variable, μ is the vector for the expected values of the parameters θ_i (“mean” values of the parameters in the population of studies), Σ is the variance-covariance matrix of θ_i (which determines the between-study variability of the parameters), and σ_ϵ^2 is the residual intra-study variance (representing the variability unexplained by X in the studies).

Equation (7.3) describes the response to Y as a function of X in each study i . Eq. (7.4) describes the between-study variability. There are many variants of model (7.3 and 7.4) based on other hypotheses concerning the distribution of the parameters and the residual error, but model (7.3 and 7.4) is appropriate for use in numerous situations (see the example below). The parameters of non-linear mixed models can be estimated with the `nlme` package (function `nlme()`) or with the `saemix` package.

7.3.2 Example

The objective is to estimate the response, in terms of N_2O (a greenhouse gas produced in large amounts by agricultural activities, including the application of nitrogen fertilizer in particular) emissions, to the dose of nitrogen fertilizer applied (Philibert et al. 2012; Gerber et al. 2016). We use here measurements of N_2O emissions collected in 203 experimental studies (corresponding to different site-years). In each study, several doses of fertilizer were applied to different experimental plots and N_2O emissions were measured (in $\text{kg ha}^{-1} \text{ year}^{-1}$) on each plot. In total, 985 values for N_2O emission are available. As an illustration, the measurements obtained in nine of the 203 experimental studies are presented in Fig. 7.2. This figure shows that the form of the emission response to fertilizer dose varies considerably between studies. The response is very strong in some studies and quite weak in others. The response does not seem to be linear: the emissions increase strongly for doses above 200 or 250 $\text{kg} \cdot \text{ha}^{-1}$.

For estimation of the response, we adjust model (7.3 and 7.4) for the data, using an exponential function (Philibert et al. 2012; Gerber et al. 2016) defined as follows:

$$f(X_{ij}, \theta_i) = \exp(\theta_{0i} + \theta_{1i}X_{1ij} + \mu_2X_{2ij}) \quad (7.5)$$

where X_{1ij} is a continuous variable representing the dose of fertilizer applied to the j^{th} plot of the i^{th} study, and X_{2ij} is a binary variable indicating whether the crop implemented in the i^{th} study is inundated rice. An “inundated rice” effect was added to the model because N_2O emissions are known to be lower for this crop.

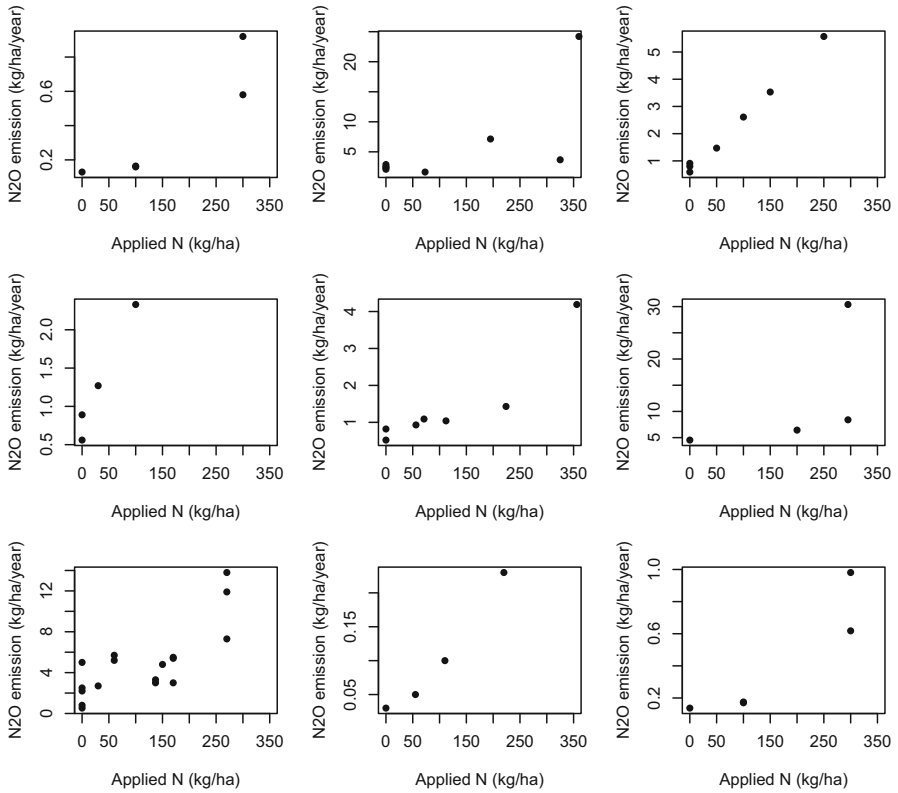


Fig. 7.2 Examples of the measurements of N₂O emissions obtained in nine experiments, following the application of various doses of N fertilizer

With this model, θ_{0i} and θ_{1i} follow a normal distribution, with expected values μ_0 and μ_1 . The “rice” effect is assumed to be fixed. Other variants have been tested, but the results showed them to be less satisfactory (Philibert et al. 2012; Gerber et al. 2016). According to Eq. (7.5), the mean emission response to fertilizer application at dose X_1 is $\exp(\mu_0 + \mu_1 X_1 + \mu_2)$ for rice crops, and $\exp(\mu_0 + \mu_1 X_1)$ for other crops.

The model was adjusted for the data with the nlme function, with the code below:

```
#Definition of dataset structure
groupedTAB <- groupedData(N20 ~ N_rate + Rice | Ref_num, data = TAB)
#Adjustment of the model
model <- nlme(N20~exp(theta0 + theta1*N_rate + theta2*Rice),
data=groupedTAB,
fixed=theta0 + theta1 + theta2~1,random=pdDiag(theta0 + theta1~1),
start=c(theta0= 1.46, theta1= 0.002, theta2=0))
```

By contrast to linear models and generalized linear models, it is necessary to initialize the estimation algorithm of the non-linear model with initial values. In

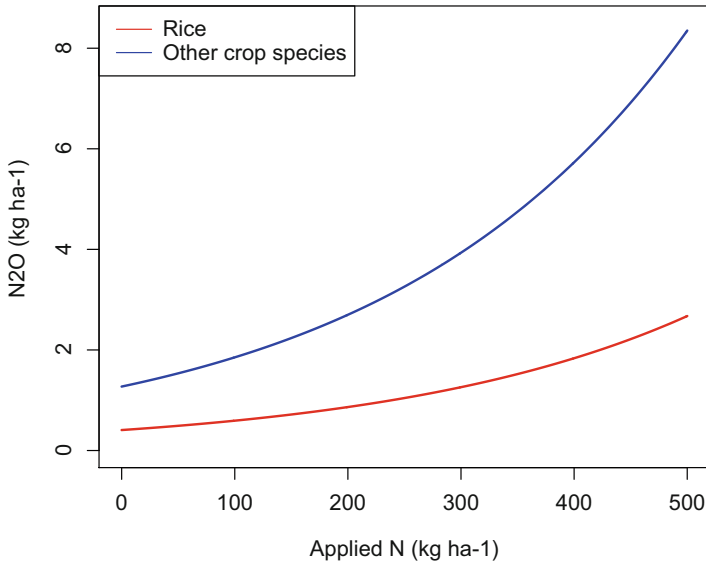


Fig. 7.3 Estimated responses of N₂O emissions to the dose of N fertilizer, for rice and for other crops

nlme, these values are specified with the argument `start`. The estimated values of μ_0 , μ_1 and μ_2 are 0.24, 0.0038, and -1.14 , respectively. All these values are significantly different from zero ($p < 0.01$). The positive sign of the value estimated for μ_1 indicates that the emissions increase with the dose of N fertilizer. The negative sign of the value estimated for μ_2 indicates that the emissions are lower for inundated rice, as expected. These results are presented in Fig. 7.3, which shows the estimated responses for the two types of crop.

7.4 Bayesian Models

7.4.1 Definition

There are many works describing the principles of Bayesian statistics and the methods for their use (e.g. Gelman et al. 2014; Biobayes 2015). We provide here only a few basic elements.

The Bayesian approach can be used to integrate two types of information for estimation of the parameters of interest: (i) prior information (available before the data are obtained), which may originate from an expert opinion, the scientific literature or previous analyses, for example, and (ii) the data.

In Bayesian statistics, the parameters are not considered to be fixed, but are instead treated as random variables distributed according to certain probability

distributions. These probability distributions describe our uncertainty about the values of the parameters. Each parameter is characterized by two probability distributions: the prior and posterior distributions. The prior distribution represents the initial uncertainty on the value of the parameter (before the acquisition of data). When little information other than the collected data is available, the prior distribution is considered to be poorly informative, as it can provide little information about the possible values of the parameter or parameters to be estimated. The posterior distribution describes the final uncertainty, once the data have been taken into account. The posterior distribution is calculated from a combination of the prior distribution and the available data, in the framework of probability theory, using Bayes' theorem.

All the statistical models described in Chaps. 2, 3, 4, 5 and 6 can be formulated in a Bayesian framework. This formulation requires definition of the prior distribution of the parameters. We illustrate this approach here for the random-effects model defined in Sect. 6.2.4 to calculate mean effect size in a meta-analysis. As in classical statistics, this model is defined as follows:

$$L_i = \mu + b_i + \varepsilon_i$$

where L_i is the estimated effect size for study i (e.g. log ratio), μ is the expected effect size across all the studies of the population considered, b_i is a random effect of "study" describing the difference between the effect size of study i ($\mu + b_i$) and the expected effect size μ , and ε_i is the difference between L_i and $\mu + b_i$ (residual error). The variances of b_i , ε_i and L_i are:

$$\begin{aligned}\text{var}(b_i) &= \sigma_b^2 \\ \text{var}(\varepsilon_i) &= \sigma_i^2 \\ \text{var}(L_i) &= \sigma_b^2 + \sigma_i^2\end{aligned}$$

For estimation of the parameters of this model in a Bayesian framework, we need to define the prior distribution for each unknown parameter. For example, the prior distribution of μ may be defined as a normal distribution. If we know nothing about μ , then the expected mean of the normal distribution can be fixed at zero and its variance can be fixed at a very large value (e.g. 10^8), so as to define a very flat distribution. The prior distribution of σ_b^2 cannot be defined as a normal distribution because the variance can only take positive values (normal distributions have no bound). An inverse-gamma distribution is often used to define the prior distributions of variances, because such distributions are defined for $[0, +\infty)$, making it possible to have a weakly informative distribution provided an appropriate parametrization is chosen (see the example below). If the intra-study variances σ_i^2 are assumed to be known (which is often the case in meta-analyses, in which they are extracted from the articles), it is not necessary to define a prior distribution for these parameters. By contrast, if these variances are unknown, a prior distribution must be defined for each, as for σ_b^2 .

Once the model has been defined, the posterior distribution is calculated from the data, for all the unknown parameters. In practice, the posterior distribution can only rarely be calculated analytically, and this determination requires an algorithm. Markov-chain Monte Carlo (MCMC) methods are the most widely used algorithms for this purpose. They can be implemented with statistical software, such as the `MCMCglmm` package (Hadfield 2010). This package can be used to adjust the linear and generalized models described in Chaps. 2, 3, 4, 5, 6 and 7 in a Bayesian framework. An application is presented below.

7.4.2 Example: Meta-analysis with `MCMCglmm`

We will come back here to the problem of estimating yield losses due to a switch from conventional to organic cropping systems based on the data from 65 studies extracted from the scientific literature (see Chap. 6). We will use here the random-effects model defined above (Sect. 7.4.1) to estimate mean effect size (yield ratio) in a Bayesian framework, with the `MCMCglmm` package (Hadfield 2010).

The prior distribution for the expected mean log yield ratio (μ) is defined as a normal distribution with an expected mean value of zero and a variance of 10^8 . This variance is sufficiently large for the prior distribution to be weakly informative. The prior distribution of the interstudy variance (σ_b^2) is defined as a weakly informative inverse-gamma distribution with parameters fixed at 0.5 ($V = 1$ and $\text{nu} = 1$, according to the parameterization used by Hadfield 2010). The intra-study variances (σ_i^2) should be set equal to the variances of the log yield ratios extracted from the articles (as in Sect. 6.2.4). Nevertheless, the `MCMCglmm` package adds, by default, an additive Gaussian overdispersion defined by a residual variance estimated from the data (denoted R in `MCMCglmm`). The prior distribution of this variance is also defined as a gamma distribution, with parameters fixed at 0.5 here.

During the use of `MCMCglmm`, certain parameters must be specified to tune the MCMC algorithm, including, in particular, the total number of iterations (argument `nitt`), the burn-in period (argument `burnin`, indicating the number of iterations eliminated from the front of the chain), and the number of consecutive iterations eliminated along the entire length of the chain, to decrease autocorrelation (argument `thin`). The values of these regulatory parameters can be fixed on the basis of a diagnosis performed with the `gelman.plot` and `autocorr` functions of the `coda` package (see below).

The model is adjusted with the following code:

```
library(MCMCglmm)
#Definition of priors, B for mu, R for residual variance and G for the
#interstudy variance
prior1 <- list(B=list(mu=0,V=10^8), R=list(V=1,nu=1),G=list(G1=list(V=1,
  nu=1)))
#Adjustment of the model with 50000 iterations of MCMC, a burn-in period of
```

```
#10000 iterations, and the elimination of 90% of the iterations to reduce auto-
correlations
Mod_mcmc <- MCMCglimm(lnR~1,random=~Study, mev=Data$Var_InR,
  data=Data, verbose=F, nitt=50000, thin=10, burnin=10000,
  prior=prior1,pr=TRUE)
#Visualization of the adjustment summary
summary(Mod_mcmc)
#Graph of the value chain for the parameter mu
plot(Mod_mcmc)
```

The results obtained for the parameter μ are presented in Fig. 7.4. The values presented in this figure are the values of μ generated by the MCMC algorithm. These values can be extracted from the object Sol generated by MCMCglimm (Mod_mcmc\$Sol[,1]). They can be used to describe the prior distribution of yield (the exponential of μ), including, in particular, its mean and quantiles:

```
mean(exp(Mod_mcmc$Sol[,1]))
quantile(exp(Mod_mcmc$Sol[,1]), 0.025)
quantile(exp(Mod_mcmc$Sol[,1]), 0.975)
```

These results were obtained with 4000 values from the MCMC presented in Fig. 7.4 (50,000 values in total – 10,000 values for the burn-in period = 40,000; 40,000/10 = 4000). The results obtained are similar to those obtained with the Dersimonian and Laird method (Chap. 6; the posterior mean is 0.78 and the 95% interval of credibility is [0.70–0.87]).

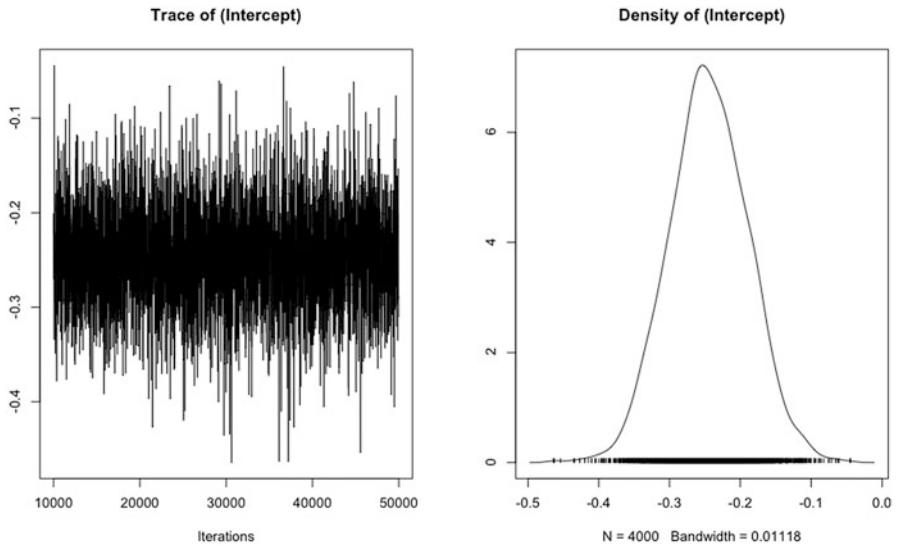


Fig. 7.4 Values generated by MCMCglimm for the parameter μ (log ratio for yield) (graph on the left) and visualization of the posterior probability density of μ (graph on the right)

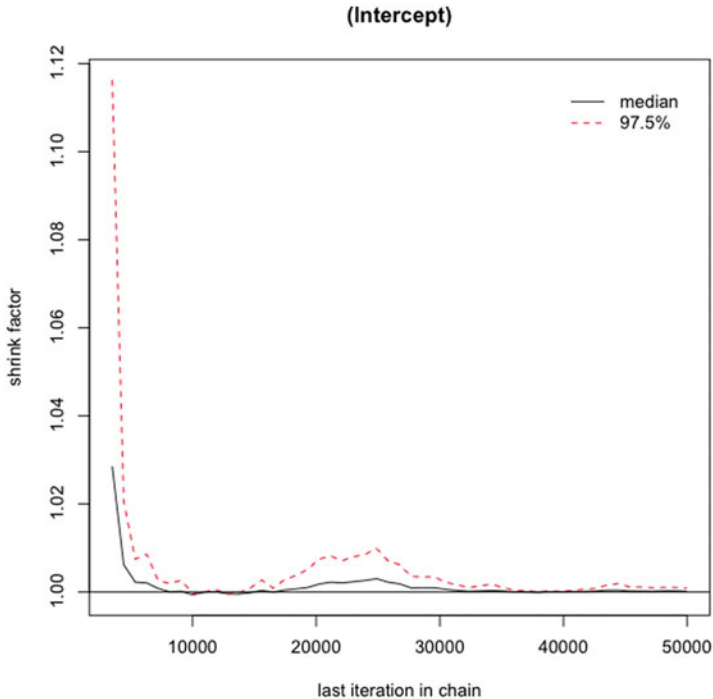


Fig. 7.5 Graph of the Gelman and Rubin factor for the diagnosis of chain convergence. A value below 1.02 is considered satisfactory. Here, the algorithm converges after a few thousand iterations. In this case, a burn-in period of 10,000 iterations is, thus, largely sufficient

The convergence of the MCMC algorithm should be checked by the generation of several chains of values (at least three) and the comparisons of variances within and between chains with the `gelman.plot` function of the `coda` package. An example based on three chains is presented below, and the results are presented in Fig. 7.5:

```
Mod_mcmc_1 <- MCMCglmm(lnR~1, random=~Study, mev=Data$Var_InR,
  data=Data, verbose=F, nitt=50000, prior=prior1)
Mod_mcmc_2 <- MCMCglmm(lnR~1, random=~Study, mev=Data$Var_InR,
  data=Data, verbose=F, nitt=50000, prior=prior1)
Mod_mcmc_3 <- MCMCglmm(lnR~1, random=~Study, mev=Data$Var_InR,
  data=Data, verbose=F, nitt=50000, prior=prior1)
ChainList <- mcmc.list(Mod_mcmc_1$Sol, Mod_mcmc_2$Sol,
  Mod_mcmc_3$Sol)
gelman.plot(ChainList)
```

In Bayesian statistics, it is recommended to analyze the sensitivity of the results to the posterior distributions. Here, we repeat the analysis with a prior distribution that

is slightly more informative for the two variances, corresponding to a gamma distribution with $\nu/2 = 0.001$:

```
prior1 <- list(B=list(mu=0,V=10^8), R=list(V=1,nu=0.002),G=list(G1=list
  (V=1,nu=0.002)))
```

The results are very similar. The posterior mean is 0.78 and the 95% credibility interval is [0.72–0.85].

One of the advantages of the Bayesian approach is that it can be used for sophisticated analyses of uncertainty. It is, thus, possible to describe the posterior distribution of any quantity dependent on the parameters of the model considered. For example, Fig. 7.6 presents the ranks of the 65 studies according to their estimated yield ratios (organic/conventional). The studies were ranked with the aid of 4000 values of $\mu + b_i$, $i = 1, \dots, 65$, generated by MCMCgImm. This approach generated 4000 classifications, corresponding to 4000 ranks for each study. Each rank distribution can therefore be summarized by the median rank, and the first and third quartile ranks (other choices are also possible). The code used is described below and the results are presented in Fig. 7.6:

```
SolRank <- apply(Mod_mcmc$Sol[,2:65],1,rank)
ResultSolRank <- matrix(nrow=64,ncol=3)

for (i in 1:64) {
  ResultSolRank[i,] <- c(median(SolRank[i,]),      quantile(SolRank[i,],0.25),
    quantile(SolRank[i,], 0.75))
}

TABrank <- data.frame(row.names(SolRank),ResultSolRank)

TABrank <- TABrank[order(TABrank[,2]),]
dotchart(TABrank[,2],labels=TABrank[,1], xlim=c(0,65),xlab="Ranking of
  experimental studies", pch=19)

for (i in 1:64) {

lines(c(TABrank[i,3],TABrank[i,4]),c(i,i))
}
```

Figure 7.6 shows that the ranks are generally uncertain. Some studies tend to better ranked than others. Studies 33, 56 and 43 are the highest ranked because their estimated ratios are the highest. Studies 9, 27 and 41 are the lowest ranked because they have lower ratios (i.e. greater yield losses associated with the switch to organic cropping systems). However, most of the studies cannot be ranked relative to each other with a high degree of confidence.

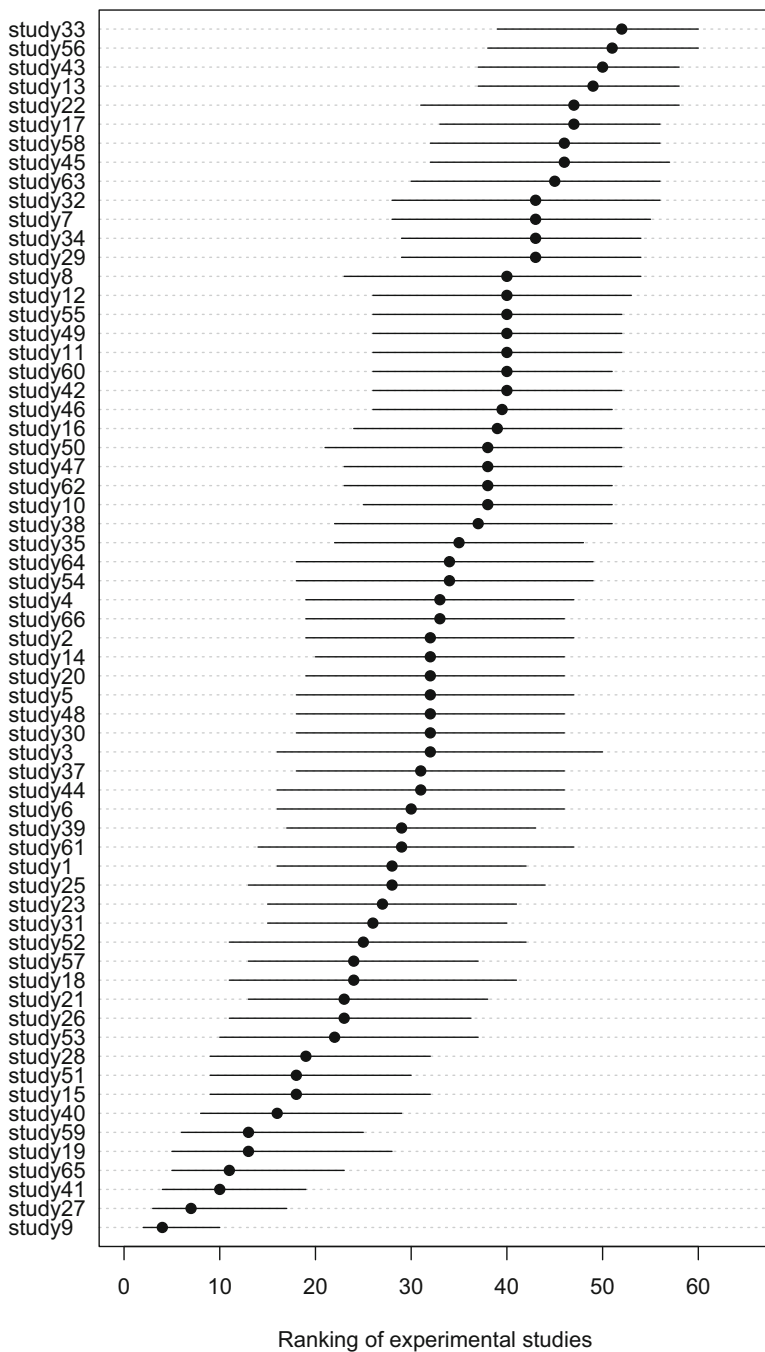


Fig. 7.6 Ranks of the 65 studies according to the ratio of organic to conventional yields. The studies with the highest ratios are the highest ranked. The points correspond to the median rank and the bars indicate the interquartile range

References

- Agresti A (2002) *Categorical data analysis*, 2nd edn. Wiley, Hoboken
- Collectif Biobayes (2015) *Initiation à la statistique bayésienne*. Ellipses, Paris
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2014) *Bayesian statistical analysis*, 3rd edn. Chapman & Hall/CRC Press, Boca Raton
- Gerber JS, Kimberly MC, Makowski D, de Cortazar-Atauri IG, Havlík P, Herrero M, Launay M, Mueller ND, O'Connell CS, Smith P, West PC (2016) Spatially explicit estimates of N₂O emissions from croplands suggest climate mitigation opportunities from improved fertilizer management. *Glob Chang Biol* 22:3383–3394
- Hadfield JD (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J Stat Softw* 33(2):1–22
- Hedges LV, Olkin I (1985) *Statistical methods for meta-analysis*. Academic, Orlando
- Lajeunesse MJ (2011) On the meta-analysis of response ratios for studies with correlated and multi-group design. *Ecology* 92:2049–2055
- Lajeunesse MJ (2015) Bias and correction for the log response ratio in ecological meta-analysis. *Ecology* 96:2056–2063
- Lajeunesse MJ (2016) Facilitating systematic reviews, data extraction and meta-analysis with the metagear package for R. *Methods Ecol Evol* 7:323–330
- Lesur-Dumoulin C, Malézieux E, Ben-Ari T, Langlais C, Makowski D (2017) Lower average yields but similar yield variability in organic versus conventional horticulture. A meta-analysis. *Agron Sustain Dev* 37:45
- Makowski D, Monod H (2011) *Analyse statistique des risques agro-environnementaux*. Springer, Paris
- Makowski D, Vicent A, Pautasso M, Stancanelli G, Rafoss T (2014) Comparison of statistical models in a meta-analysis of fungicide treatments for the control of citrus black spot caused by *Phyllosticta citricarpa*. *Eur J Plant Pathol* 139:79–94
- Michel L, Brun F, Piraux F, Makowski D (2016) Estimating the incidence of *Septoria tritici* in wheat crops from in-season measurements. *Eur J Plant Pathol*. <https://doi.org/10.1007/s10658-016-0887-9>
- Michel L, Brun F, Makowski D (2017) A framework based on generalised linear mixed models for analysing pest and disease surveys. *Crop Prot* 94:1–12
- Nakagawa S, Poulin R, Mengersen K, Reinhold K, Engquist L, Lagisz M, Senio AM (2015) Meta-analysis of variation: ecological and evolutionary applications and beyond. *Methods Ecol Evol* 6:143–152
- Philibert A, Loyce C, Makowski D (2012) Quantifying uncertainties in N₂O emission due to N fertilizer application in cultivated areas. *PLoS One* 7(11):e50950. <https://doi.org/10.1371/journal.pone.0050950>
- Sokal RR, Rohlf FJ (1995) *Biometry: the principles and practice of statistics in biological research*, 3rd edn. W.H. Freeman, New York

Appendix: R Resources to Implement Network Analysis and Meta-analysis Methods

This appendix presents the computer resources used in this book to implement network analysis and meta-analysis methods. It details the content of the KenSyn R package (R Core Team 2019), developed specifically for this book. This package contains both the data sets used in the examples and the R functions used in the statistical analyses. Some of the functions presented here are essential to implement the mixed model. Other functions facilitate the extraction of information from adjusted models, the analysis of uncertainties associated with estimates or the formatting of results according to conventional standards.

KenSyn Package: R Code and Data Sets of the Examples Presented in the Different Chapters

All the examples presented in this book have been gathered together into a KenSyn R package for “Knowledge Synthesis in Agriculture”.

Installation

You can install the KenSyn package from the CRAN repository (<http://cran.r-project.org/web/packages/KenSyn>), if you have an Internet connection, by following these instructions:

- Step 1: click on “Packages” in the R menu and select “Install the package(s)...”.
- Step 2: Choose the "0-Cloud" location or a location near you from the list of mirror download sites.
- Step 3: Choose “KenSyn” from the list of available packages.

- Step 4: Check the message in the console window indicating that this installation has been successful. Any dependencies are installed.

Another way to install the **KenSyn** package is to execute the following instruction in the R console.

```
install.packages ("KenSyn", rest = "http://cran.rstudio.com")
```

You can also use an R editor like Rstudio (<https://www.rstudio.com>), which makes it easier to write code, explore results and install R packages.

The current version of **KenSyn** is version 0.3 (April 2019). It is fully functional with the current version of R (version 3.5.3). As **nlme** and **lmer** packages have evolved very regularly, it is recommended to have a recent version of R and it is possible that slightly different results may be obtained depending on the versions of R and the different packages.

In addition to **KenSyn**, depending on the examples, you may need to install other packages used in the different example of this book. They are then loaded at the beginning of the various scripts by the function **library(package_name)**. To do this, proceed in the same way as for **KenSyn** package.

Content and Use

The **KenSyn** package is an R package (a library) that brings together data sets, examples, and functions in a shareable and documented form. This package includes the code in the form of a demo for all the examples that are provided in this book (thus you can run them yourself) and the necessary data sets.

Examples in the Form of Demos

To see all the demo scripts, you can run the following lines

```
library(KenSyn)
demo(package="KenSyn")
```

A table similar to the one presented in Table 1 will be displayed. The beginning of the name (chxxx_) allows you to match the chapters of this book.

Then, to execute a demo script, you must specify the name and execute the following instruction.

```
demo(ch01_mixedmodel_nlme, package="KenSyn")
```

Table 1 List of scripts containing the examples as R demo files in the KenSyn package

Name of the demo	Description
ch01_mixedmodel_nlme	ch01. Simple mixed-effects model example on wheat yield (with nlme and lme4)
ch03_network_compblock	ch03. Network of experiments: complete blocks. Wheat varieties on a single year
ch04a_network_mean_data	Ch04a. Network of experiments: Wheat varieties on a single year. Analysis of means of data
ch04b_network_hetero_var	ch04b. Network of experiments: Wheat varieties on a single year. Heterogeneous variances
ch04c_network_pluriannual	ch04c. Network of experiments: Wheat varieties on several years
ch06a_metaanalysis_organic	ch06. Meta-analysis comparing organic vs conventional cropping system with (nlme and metafor)
ch06b_metaanalysis_metareg	ch06. Meta-analysis: code for illustrating the main principles of meta-regression (with nlme)
ch06c_metaanalysis_metareg_m	ch06. Meta-analysis: code for illustrating the main principles of meta-regression (with metafor)
ch07a_metaanalysis_citrus	ch07. Meta-analysis to estimate the effectiveness of a fungicidal treatment to control <i>Phyllosticta citricarpa</i> , a citrus disease (with lme4)
ch07b_metaanalysis_N2O	ch07. Meta-analysis to estimate the response of N2O emissions to the applied N fertilizer rate (with nlme)
ch07c_metaanalysis_bayesian	ch07. Meta-analysis with Bayesian approach: comparing organic vs conventional cropping system (with MCMCglmm)
others_network_machines	Others. Network of experiments: complete blocks. Test of worker performance on machine
others_network_var_itk	Others. Network of experiments: Wheat varieties with different crop managements
others_network_var_soil	Others. Network of experiments: Wheat varieties with 2 different soils

The script is then executed step by step in an interactive way. Simply press the *Enter* key on the keyboard in the console area to execute the following instructions and see the results appear as they are computed, either in the console area or in the graphics area.

It is quite convenient to get the r script directly. This script is stored in the installation directory of the package on your computer, you can know the path with the following instruction, remembering to add the extension `.r` after the name of the example.

```
system.file("demo", "ch01_mixedmodel_nlme.r", package = "KenSyn")
[1] "D:/XXXXXX/Documents/R/win-library/3.5/KenSyn/demo/
ch01_mixedmodel_nlme.r"
```

Table 2 List of data sets in the KenSyn package

Data set	Description
N2O	Meta-analysis dataset to estimate the response of N2O emissions to the applied N fertilizer rate
Citrus	Meta-analysis dataset: estimation the effectiveness of a fungicidal treatment to control <i>Phyllosticta citricarpa</i> , a citrus fungus
machines	Workers testing a new machine.
Organic	Meta-analysis dataset on comparison of organic to conventional crop systems.
wheat_var	Network of experiment to evaluate Wheat varieties on one single year
wheat_var_itk	Network of experiment to evaluate Wheat varieties on one single year with different cropping systems
wheat_var_soil	Network of experiment to evaluate Wheat varieties on two contrasted soils
wheat_var_years	Network of experiment to evaluate Wheat varieties on several years (2005–2010)
wheatyield	Wheat Yield at a regional scale (fake data)

You can then open it in your code editor and save it in your working directory.

Another way is to access the example code via the R help interface. To do this, you execute the instruction.

```
help(KenSyn)
```

The package help page appears in your browser. You must then go to the very bottom of this page, to access the *index*. Then a “Code Demo” link allows you to access the entire list of examples as a demo. This script can then be copied to work on it. Note that this access does not work under current R studio version.

Data Sets

To see all the data sets used in the different examples and mobilized in the demo scripts, you can execute the following instruction.

```
library(KenSyn)
data(package="KenSyn")
```

A table similar to the one presented in Table 2 is then displayed.

Then by launching a specific dataset help, you have a brief description of the origin and structure of the data.

```
help(wheat_var)
```

The data set is accessible as soon as the package is loaded and can be used, for example, as follows:

```
# see the structure
str(wheat_var)
# see the first lines of the data set
head(wheat_var)
# or a histogram with the values of all the yields
hist(wheat_var$rendement)
```

Implement the Mixed-Effects Model Under R

Adjust a Mixed-Effects Model

Two main R packages are available to implement the mixed-effects model under R. These are packages that are widely used and well maintained by the community. Nevertheless, the evolution dynamics of these packages has remained strong in recent years and in a number of cases we regularly face code problems related to incompatibilities between versions.

To fit a mixed-effect model, there are two packages that are widely used in the R community.

nlme Package

The package **nlme** (Nonlinear Mixed-Effects Models, Pinheiro et al. 2017) is dedicated to the adjustment and comparison of linear and non-linear Gaussian mixed-effects models.

- The generic **lme** function (Linear Mixed-Effects Models) allows non-linear mixed-effects models (fixed and random) to be adjusted according to the formulation described by Laird and Ware (1982) by allowing nested random effects. Intra-group errors can be correlated and/or have unequal variances. It returns an object containing the fitted model.

In the example above, we model the variable *Rdt* by an intercept with a random effect on this intercept by *Site*. The data set containing the variables is *wheatyield*.

```
lme(Rdt~1, random=~1|Site, data=wheatyield)
```

- The **groupedData** function allows you to structure a dataset according to a precise formula and the object created is then directly usable by modeling functions as **lme**.

In the example above, we have the same example as before, but with the preliminary step of structuring the data.

```
gData= groupedData(Rdt~1|Site, data= wheatyield)
lme(gData)
```

lme4 Package

The **lme4** package (Linear, generalized linear, and nonlinear mixed models, Bates et al. 2015) is dedicated to the adjustment and analysis of mixed-effect models, with different functions for each type of model.

- The **lmer** function is used to adjust linear mixed-effect models.

The example shows you the same simple model with the **lmer** function.

```
lmer(Rdt~1+(1|Site), data= wheatyield)
```

- The **glmer** function is used to adjust generalized linear models with mixed effects.
- The **nlmer** function is used to adjust non-linear mixed-effect models.

Work with the Results of the Mixed Models Under R

Basic Features

Various generic functions allow to extract useful information in a readable format from objects resulting from mixed-effect model adjustments (but also from other models)

Let's save the result of the adjustment of a model defined with **nlme** or **lmer** into an object named *fit*.

```
fit = lme(Rdt~1, random=~1|Site, data= wheatyield)
```

- **summary** formats the main results of a fitted model in a readable form.

```
summary (fit)
```

- **fitted** retrieves the adjustments by the model.

```
fitted (fit)
```

- **residuals** retrieves the residuals.

```
residuals (fit)
```

- **ranef** display random effects.

```
ranef (fit)
```

- **predict** make predictions with the model, on the same data as used for the adjustment. In this case, this corresponds to the results of the fitted function.

```
predict (fit)
```

Or **predict** with new data, with in this case, the site effect that can be taken into account if the site was in the data set used to fit the model. Thus, in the example, the prediction for data from site 3 is computed.

```
predict (fit, newdata= data.frame (Site=3, Rdt=NA) )
```

Emmeans Package

The **emmeans** package (Lenth 2017) computes the estimated marginal means (EMM) for specified factors or combinations of factors in a linear model and, possibly, comparisons or contrasts between them. EMMs are also known as "least square means". It replaces the obsolete Ismeans package.

- **emmeans** calculates adjusted means for each level of a factor.
- **pairs** performs all 2–2 comparisons between means using the "**tukey**" method to adjust the probability of tests to take into account the multiplicity of tests.
- **cld** build homogeneous groups.
- **contrast**, used with the "**eff**" method, calculates all the differences between each level and the overall mean. The "**sidak**" adjustment method adapts the probability of tests to take into account the multiplicity of tests. Used with the argument **method="trt.vs.ctrl"** and **ref=n**, contrast allows to compare all levels to level n.
- **confint** computes confidence intervals of the differences between the levels and a control level or the confidence intervals for the effects of a variable, depending on the **contrast** method.

This package requires the **pbkrtest** and **lmerTest** packages for some tests.

Car Package

The **car** package (Companion to Applied Regression) contains functions and example of a book (Fox and Weisberg 2011). The **Anova** function is mainly used to calculate the analysis of variance table (with a capital A, different from the basic **anova** function).

Outliers Package

The **outliers** package (Komsta 2011) contains tests commonly used to identify outliers and in our case it can be used to detect suspicious residues.

The Metafor Package, to Performing Meta-analyses Under R

The **metafor** package (A Meta-Analysis Package for R) provides a complete collection of functions for performing meta-analyses in R (Viechtbauer 2010). The package includes functions to calculate different effect sizes or results frequently used in meta-analyses. It allows the user to adjust different models with fixed, random or mixed-effects. Finally, it provides analyses and graphs traditionally used in meta-analyses.

- **escalc** is used to calculate the effect size.
- **rma.uni** adjusts linear models for meta-analyses with fixed, random or mixed-effects.
- **forest** allows you to plot the forest graph with the effect sizes very often used in meta-analysis.
- **funnel** allows you to draw a plot to evaluate possible publication biases.

Bayesian Approach with the Mixed-Effect Model

MCMCglmm Package

The **MCMCglmm** package (MCMC Generalised Linear Mixed Models, Hadfield 2010) is used to implement generalised linear mixed-effects models in a Bayesian framework using Markov Chain Monte Carlo (MCMC). The **MCMCglmm** function applies the Markov Chain Monte Carlo sampler to adjust generalized linear mixed-effects models.

Coda Package

The **coda** package (Plummer et al. 2006) provides functions to summarize and plot the chains sampled by Markov Chain Monte Carlo methods and to perform diagnostic tests of convergence to the equilibrium distribution of the Markov chain. The **gelman.plot** function plots the evolution of the Gelman and Rubin factor as a function of the number of iterations to diagnose graphically the convergence of the MCMC algorithm.

References

- Bates D, Maechler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67(1):1–48. <https://doi.org/10.18637/jss.v067.i01>
- Fox J, Weisberg S (2011) *An R companion to applied regression*, 2nd edn. Sage, Los Angeles
- Hadfield JD (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J Stat Softw* 33(2):1–22. www.jstatsoft.org/v33/i02
- Komsta L (2011) *Outliers: Tests for outliers*. R package version 0.14. <https://CRAN.R-project.org/package=outliers>
- Laird NM, Ware JH (1982) Random-effects models for longitudinal data. *Biometrics* 38:963–974
- Lenth R (2017) *emmeans: Estimated marginal means, aka least-squares means*. R package version 0.9.1. CRAN.R-project.org/package=emmeans
- Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2017) *nlme: linear and nonlinear mixed effects models*. R package version 3.1-131, CRAN.R-project.org/package=nlme
- Plummer M, Best N, Cowles K, Vines K (2006) CODA: convergence diagnosis and output analysis for MCMC. *R News* 6:7–11
- R Core Team (2019) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. www.R-project.org
- Viechtbauer W (2010) Conducting meta-analyses in R with the metafor package. *J Stat Softw* 36(3):1–48. www.jstatsoft.org/v36/i03