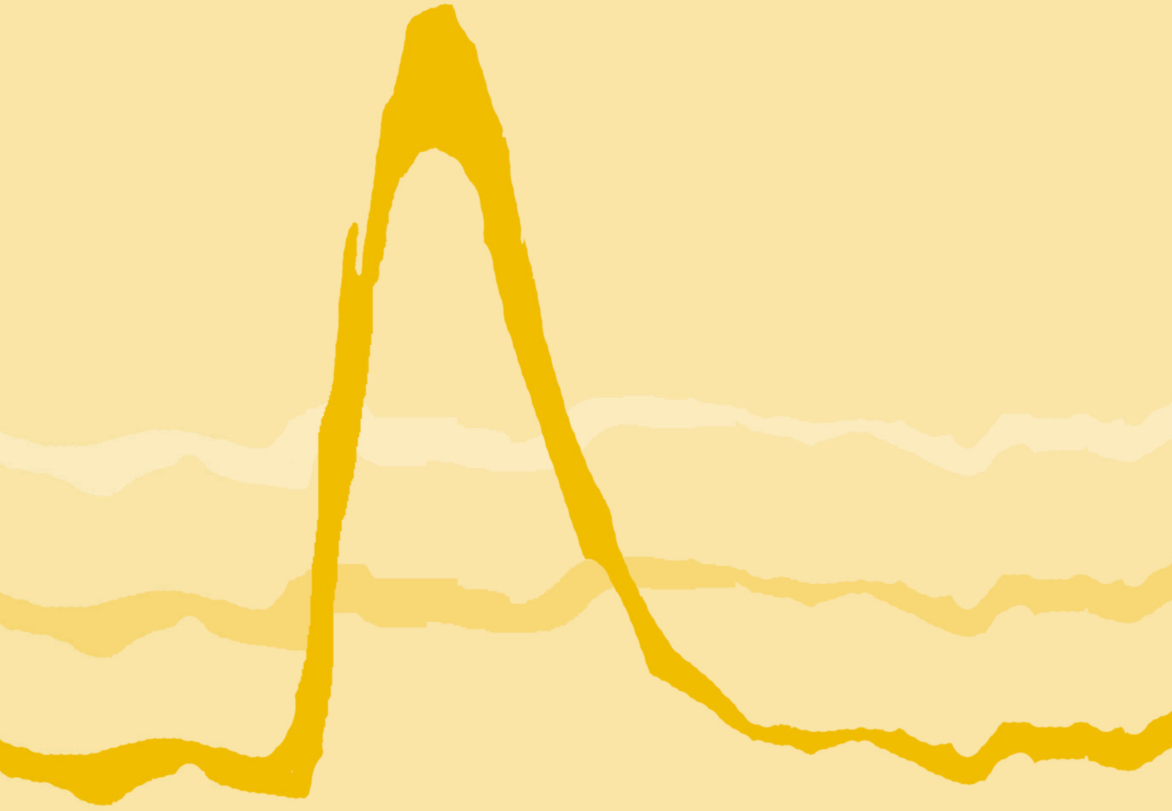
Statistics Applied to Clinical Trials

Third Edition

T.J. Cleophas, A.H. Zwinderman and
T.F. Cleophas



Springer

STATISTICS APPLIED TO CLINICAL TRIALS
THIRD EDITION

Statistics Applied to Clinical Trials, Third Edition

by

TON J CLEOPHAS, MD, PhD, Associate-Professor,
President American College of Angiology,
Co-Chair Module Statistics Applied to Clinical Trials,
European Interuniversity College of Pharmaceutical Medicine Lyon, France,
Internist-clinical pharmacologist,
Department Medicine, Albert Schweitzer Hospital, Dordrecht, The Netherlands

AEILKO H ZWINDERMAN, Math D, PhD, Professor,
Co-Chair Module Statistics Applied to Clinical Trials,
European Interuniversity College of Pharmaceutical Medicine Lyon, France,
Professor of Statistics,
Department Biostatistics and Epidemiology, Academic Medical Center Amsterdam, The Netherlands

and

TOINE F CLEOPHAS, D Techn,
Technical University, Delft, The Netherlands

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4020-4229-4 ISBN 978-1-4020-4650-6 (eBook)
DOI 10.1007/978-1-4020-4650-6

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

All Rights Reserved
© 2006 Springer
No part of this work may be reproduced, stored in a retrieval system, or transmitted
in any form or by any means, electronic, mechanical, photocopying, microfilming, recording
or otherwise, without written permission from the Publisher, with the exception
of any material supplied specifically for the purpose of being entered
and executed on a computer system, for exclusive use by the purchaser of the work.

TABLE OF CONTENTS

CHAPTER 5 / STATISTICAL POWER AND SAMPLE SIZE

CHAPTER 6 / INTERIM ANALYSES

CHAPTER 7 / MULTIPLE STATISTICAL INFERENCES

CHAPTER 8 / CONTROLLING THE RISK OF FALSE POSITIVE CLINICAL TRIALS

CHAPTER 13 / CURVILINEAR REGRESSION

CHAPTER 14 / LOGISTIC AND COX REGRESSION, PROBLEMS WITH REGRESSION MODELING, MARKOW MODELS

CHAPTER 15 / REGRESSION MODELING FOR IMPROVED PRECISION

CHAPTER 16 / POST-HOC ANALYSIS IN CLINICAL TRIALS, A CASE FOR LOGISTIC REGRESSION ANALYSIS

CHAPTER 25 / TESTING CLINICAL TRIALS FOR RANDOMNESS

CHAPTER 26 / CLINICAL DATA WHERE VARIABILITY IS MORE IMPORTANT THAN AVERAGES

CHAPTER 27 / TESTING REPRODUCIBILITY

CHAPTER 28 / ACCURACY OF DIAGNOSTIC TESTS

CHAPTER 29 / ADVANCED ANALYSIS OF VARIANCE

PREFACE

The European Interuniversity Diploma of Pharmaceutical Medicine is a postacademic course of 2-3 years sponsored by the Socrates program of the European Community. The office of this interuniversity project is in Lyon and the lectures are given there. The European Community has provided a building and will remunerate lecturers. The institute which provides the teaching is called the European College of Pharmaceutical Medicine, and is affiliated with 15 universities throughout Europe, whose representatives constitute the academic committee. This committee supervises educational objectives. Start lectures February 2000.

There are about 20 modules for the first two years of training, most of which are concerned with typically pharmacological and clinical pharmacological matters including pharmacokinetics, pharmacodynamics, phase III clinical trials, reporting, communication, ethics and, any other aspects of drug development. Subsequent training consists of practice training within clinical research organisations, universities, regulatory bodies etc., and finally of a dissertation. The diploma, and degree are delivered by the Claude Bernard University in Lyon as well as the other participating universities.

The module "Statistics applied to clinical trials" wil be taught in the form of a 3 to 6 day yearly course given in Lyon and starting February 2000. Lecturers have to submit a document of the course (this material will be made available to students). Three or 4 lecturers are requested to prepare detailed written material for students as well as to prepare examination of the students. The module is thus an inportant part of a postgraduate course for physicians and pharmacists for the purpose of obtaining the European diploma of pharmaceutical medicine. The diploma should make for leading positions in pharmaceutical industry, academic drug research, as well as regulatory bodies within the EC. This module is mainly involved in the statistics of randomized clinical trials .

The chapters 1-9, 11, 17, 18 of this book are based on the module "Medical statistics applied to clinical trials" and contain material that should be mastered by the students before their exams. The remaining chapters are capita selecta intended for excellent students and are not included in the exams.

The authors believe that this book is innovative in the statistical literature because, unlike most introductory books in medical statistics, it provides an explanatory rather than mathematical approach to statistics, and, in addition, emphasizes non-classical but increasingly frequently used methods for the statistical analyses of clinical trials, e.g., equivalence testing, sequential analyses, multiple linear regression analyses for confounding, interaction, and synergism.The authors are not aware of any other work published so far that is comparable with the current work, and, therefore, believe that it does fill a need.

August 1999
Dordrecht, Leiden, Delft

PREFACE TO SECOND EDITION

In this second edition the authors have removed textual errors from the first edition. Also seven new chapters (chapters 8, 10, 13, 15-18) have been added. The principles of regression analysis and its resemblance to analysis of variance was missing in the first edition, and have been described in chapter 8. Chapter 10 assesses curvilinear regression. Chapter 13 describes the statistical analyses of crossover data with binary response. The latest developments including statistical analyses of genetic data and quality-of-life data have been described in chapters 15 and 16. Emphasis is given in chapters 17 and 18 to the limitations of statistics to assess non-normal data, and to the similarities between commonly-used statistical tests. Finally, additional tables including the Mann-Whitney and Wilcoxon rank sum tables have been added in the Appendix.

December 2001, Dordrecht, Amsterdam, Delft

PREFACE TO THE THIRD EDITION

The previous two editions of this book, rather than having been comprehensive, concentrated on the most relevant aspects of statistical analysis. Although well-received by students, clinicians, and researchers, these editions did not answer all of their questions. This called for a third, more comprehensive, rewrite. In this third edition the 18 chapters from the previous edition have been revised, updated, and provided with a conclusions section summarizing the main points. The formulas have been re-edited using the Formula-Editor from Windows XP 2004 for enhanced clarity. Thirteen new chapters (chapters 8-10, 14, 15, 17, 21, 25-29, 31) have been added. The chapters 8-10 give methods to assess the problems of multiple testing and data testing closer to expectation than compatible with random. The chapters 14 and 15 review regression models using an exponential rather than linear relationship including logistic, Cox, and Markow models. Chapter 17 reviews important interaction effects in clinical trials and provides methods for their analysis. In chapter 21 study designs appropriate for medicines from one class are discussed. The chapters 25-29 review respectively (1) methods to evaluate the presence of randomness in the data, (2) methods to assess variabilities in the data, (3) methods to test reproducibility in the data, (4) methods to assess accuracy of diagnostic tests, and (5) methods to assess random rather than fixed treatment effects. Finally, chapter 31 reviews methods to minimize the dilemma between sponsored research and scientific independence. This updated and extended edition has been written to serve as a more complete guide and reference-text to students, physicians, and investigators, and, at the same time, preserves the common sense approach to statistical problem-solving of the previous editions.

August 2005, Dordrecht, Amsterdam, Delft

FOREWORD

In clinical medicine appropriate statistics has become indispensable to evaluate treatment effects. Randomized controlled trials are currently the only trials that truly provide evidence-based medicine. Evidence based medicine has become crucial to optimal treatment of patients. We can define randomized controlled trials by using Christopher J. Bulpitt's definition "a carefully and ethically designed experiment which includes the provision of adequate and appropriate controls by a process of randomization, so that precisely framed questions can be answered". The answers given by randomized controlled trials constitute at present the way how patients should be clinically managed. In the setup of such randomized trial one of the most important issues is the statistical basis. The randomized trial will never work when the statistical grounds and analyses have not been clearly defined beforehand. All endpoints should be clearly defined in order to perform appropriate power calculations. Based on these power calculations the exact number of available patients can be calculated in order to have a sufficient quantity of individuals to have the predefined questions answered. Therefore, every clinical physician should be capable to understand the statistical basis of well performed clinical trials. It is therefore a great pleasure that Drs. T. J. Cleophas, A.H. Zwinderman, and T.F. Cleophas have published a book on statistical analysis of clinical trials. The book entitled "Statistics Applied to Clinical Trials" is clearly written and makes complex issues in statistical analysis transparant. Apart from providing the classical issues in statistical analysis, the authors also address novel issues such as interim analyses, sequential analyses, and meta-analyses. The book is composed of 18 chapters, which are nicely structured. The authors have deepened our insight in the applications of statistical analysis of clinical trials. We would like to congratulate the editors on this achievement and hope that many readers will enjoy reading this intriguing book.

E.E. van der Wall, MD, PhD, Professor of Cardiology, President Netherlands Association of Cardiology, Leiden, Netherlands

CHAPTER 1

HYPOTHESES, DATA, STRATIFICATION

1. GENERAL CONSIDERATIONS

Over the past decades the randomized clinical trial has entered an era of continuous improvement and has gradually become accepted as the most effective way of determining the relative efficacy and toxicity of new drug therapies. This book is mainly involved in the methods of prospective randomized clinical trials of new drugs. Other methods for assessment including open-evaluation-studies, cohort- and case-control studies, although sometimes used, e.g., for pilot studies and for the evaluation of long term drug-effects, are excluded in this course. Traditionally, clinical drug trials are divided into IV phases (from phase I for initial testing to phase IV after release for general use), but scientific rules governing different phases are very much the same, and can thus be discussed simultaneously.

A. CLEARLY DEFINED HYPOTHESES

Hypotheses must be tested prospectively with hard data, and against placebo or known forms of therapies that are in place and considered to be effective. Uncontrolled studies won't succeed to give a definitive answer if they are ever so clever. Uncontrolled studies while of value in the absence of scientific controlled studies, their conclusions represent merely suggestions and hypotheses. The scientific method requires to look at some controls to characterize the defined population.

B. VALID DESIGNS

Any research but certainly industrially sponsored drug research where sponsors benefit from favorable results, benefits from valid designs. A valid study means a study unlikely to be biased, or unlikely to include systematic errors. The most dangerous error in clinical trials are systematic errors otherwise called biases. Validity is the most important thing for doers of clinical trials to check. Trials should be made independent, objective, balanced, blinded, controlled, with objective measurements, with adequate sample sizes to test the expected treatment effects, with random assignment of patients.

C. EXPLICIT DESCRIPTION OF METHODS

Explicit description of the methods should include description of the recruitment procedures, method of randomization of the patients, prior statements about the methods of assessments of generating and analysis of the data and the statistical methods used, accurate ethics including written informed consent.

D. UNIFORM DATA ANALYSIS

Uniform and appropriate data analysis generally starts with plots or tables of actual data. Statistics then comes in to test primary hypotheses primarily. Data that do not answer prior hypotheses may be tested for robustness or sensitivity, otherwise called precision of point estimates e.g., dependent upon numbers of outliers. The results of studies with many outliers and thus little precision should be interpreted with caution. It is common practice for studies to test multiple measurements for the purpose of answering one single question. E.g., the benefit to health of a new drug may be estimated by mortality in addition to various morbidity variables, and there is nothing wrong with that practice. We should not make any formal correction for multiple comparisons of this kind of data. Instead, we should informally integrate all the data before reaching conclusions, and look for the trends without judging one or two low P-values among otherwise high P-values as proof.

However, subgroup analyses involving post-hoc comparisons by dividing the data into groups with different ages, prior conditions, gender etc can easily generate hundreds of P-values. If investigators test many different hypotheses, they are apt to find significant differences at least 5% of the time. To make sense of these kinds of results, we need to consider the Bonferroni inequality, which will be emphasized in the chapters 7 and 8, and states that if k statistical tests are performed with the cut-off level for a test statistic, for example t or F, at the α level, the likelihood for observing a value of the test statistic exceeding the cutoff level is no greater than k times α. For example, if we wish to do three comparisons with t-tests while keeping the probability of making a mistake less than 5%, we have to use instead of $\alpha = 5\%$ in this case $\alpha = 5/3\% = 1.6\%$. With many more tests, analyses soon lose any sensitivity and do hardly prove anything anymore. Nonetheless a limited number of post-hoc analyses, particularly if a plausible theory is underlying, can be useful in generating hypotheses for future studies.

2. TWO MAIN HYPOTHESES IN DRUG TRIALS: EFFICACY AND SAFETY

Drug trials are mainly for addressing the efficacy as well as the safety of the drugs to be tested in them. For analyzing efficacy data formal statistical techniques are normally used. Basically, the null hypothesis of no treatment effect is tested, and is rejected when difference from zero is significant. For such purpose a great variety of statistical significance tests has been developed, all of whom report P values, and compute confidence intervals to estimate the magnitude of the treatment effect. The appropriate test depends upon the type of data and will be discussed in the next chapter. Of safety data, such as adverse events, data are mostly collected with the hope of demonstrating that the test treatment is not different from control. This concept is based upon a different hypothesis from that proposed for efficacy data, where the very objective is generally to show that there actually is a difference between test and control. Because the objective of collecting safety data is thus different, the approach to analysis must be likewise different. In particular, it may be less appropriate to use statistical significance tests to analyze the latter data. A significance test is a tool that can help to establish whether a difference between treatments is likely to be real. It cannot be used to demonstrate that two treatments are similar in their effects. In addition, safety data, more frequently than efficacy data, consist of proportions and percentages rather than continuous data as will be discussed in the next section. Usually, the best approach to analysis of these kinds of data is to present suitable summary statistics, together with confidence intervals. In the case of adverse event data, the rate of occurrence of each distinct adverse event on each treatment group should be reported, together with confidence intervals for the difference between the rates of occurrence on the different treatments. An alternative would be to present risk ratios or relative risks of occurrence, with confidence intervals for the relative risk. Chapter 3 mainly addresses the analyses of these kinds of data.

Other aspects of assessing similarity rather than difference between treatments will be discussed separately in chapter 6 where the theory, equations, and assessments are given for demonstrating statistical equivalence.

3. DIFFERENT TYPES OF DATA: CONTINUOUS DATA

The first step, before any analysis or plotting of data can be performed, is to decide what kind of data we have. Usually data are continuous, e.g., blood pressures, heart rates etc. But regularly proportions or percentages are used for the assessment of part of the data. The next few lines will address how we can summarize and characterize these two different approaches to the data.

Samples of **continuous data** are characterized by:

$$\textbf{Mean} = \frac{\mathbf{\Sigma x}}{\mathbf{n}} = \bar{x} \, ,$$

where Σ is the summation, x are the individual data, n is the total number of data.

$$\textbf{Variance between the data} = \frac{\sum (x-\bar{x})^2}{n-1}$$

$$\textbf{Standard deviation (SD)} = \sqrt{(Variance)}$$

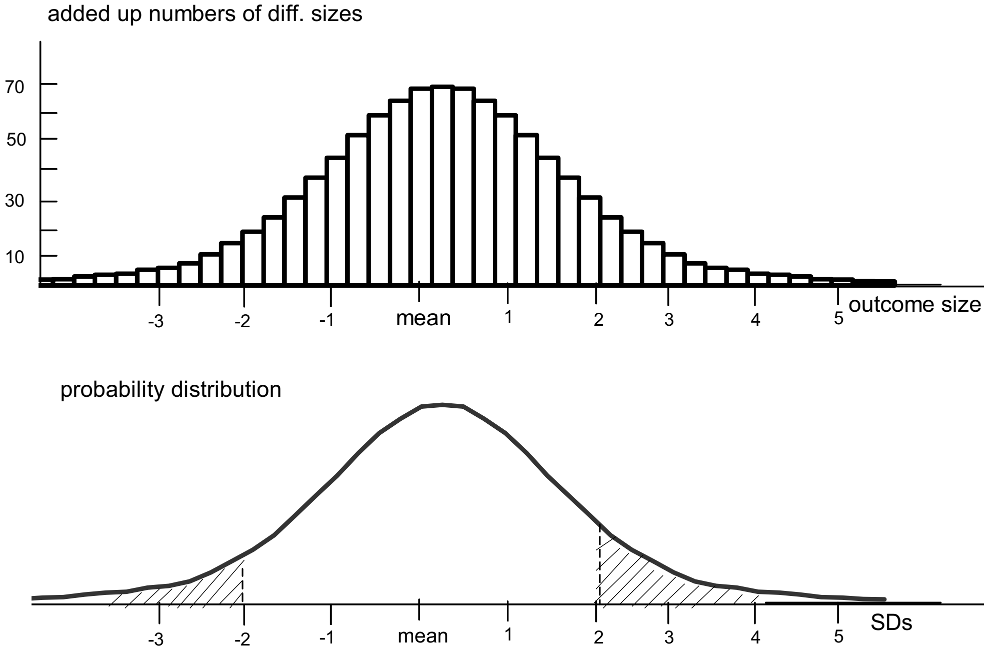


Figure 1. Histogram and Gaussian curve representation of data.

Continuous data can be plotted in the form of a histogram (Figure 1 upper graph). On the x-axis, frequently called z-axis in statistics, it has individual data. On the y-axis it has “how often”. For example, the mean value is observed most frequently, while the bars on either side gradually grow shorter. This graph adequately represents the data. It is, however, not adequate for statistical analyses. Figure 1 lower graph pictures a Gaussian curve, otherwise called normal (distribution) curve. On the x-axis we have, again, the individual data, expressed either in absolute data or in SDs distant from the mean. On the y-axis the bars have been replaced with a continuous line. It is now impossible to determine from the graph how many patients had a particular outcome. Instead, important inferences can be made. For example, the total area under the curve (AUC) represents 100% of the data, AUC left from mean represents 50% of the data, left from -1 SDs it has 15.87% of the data, left from -2SDs it has 2.5% of the data. This graph is better for statistical purposes but not yet good enough.

Figure 2 gives two Gaussian curves, a narrow and a wide one. Both are based on the same data, but with different meaning. The wide one summarizes the data of our trial. The narrow one summarizes the mean of many trials similar to our trial. We will not try to make you understand why this is so. Still, it is easy to conceive that the distribution of all means of many similar trials is narrower and has fewer outliers than the distribution of the actual data from our trial, and that it will center around the mean of our trial because our trial is assumed to be representative for the entire population. You may find it hard to believe but the narrow curve with standard errors of the mean (SEMs) or simply SEs on the x-axis can be effectively used for testing important statistical hypotheses, like (1) no difference between new and standard treatment, (2) a real difference, (3) the new treatment is better than the standard treatment, (4) the two treatments are equivalent. Thus, mean ± 2 SDs (or more precisely 1.96 SDs) represents the AUC of the wide distribution, otherwise called the 95% confidence interval of the data, which means that 95% of the data of the sample are within. The SEM-curve (narrow one) is narrower than the SD-curve (wide one) because SEM = SD/ $\sqrt{n}$ with n = sample size. Mean ± 2 SEMs (or more precisely 1.96 SEMs) represents 95% of the means of many trials similar to our trial.

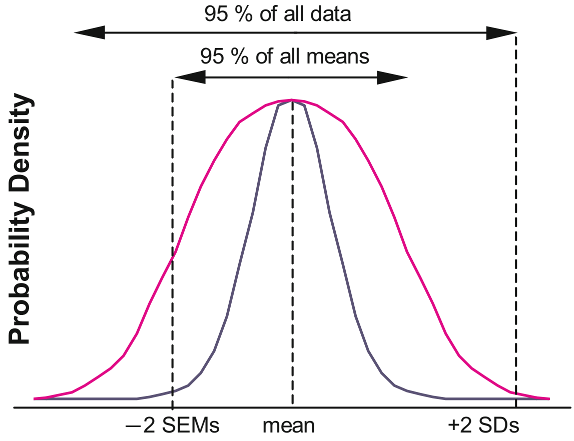


Figure 2. Two examples of normal distributions.

$$\textbf{SEM= SD / } \sum n$$

As the size of SEM in the graph is about 1/3 times SD, the size of each sample is here about n = 10. The area under the narrow curve represents 100% of the sample means we would obtain, while the area under the curve of the wide graph represents 100% of all of the data of the samples.

Why is this SEM approach so important in statistics. Statistics makes use of mean values and their standard error to test the null hypotheses of finding no difference from zero in your sample. When we reject a null hypothesis at P<0.05, it literally means that there is < 5% chance that the mean value of our sample crosses the area of the null hypothesis where we say there is no difference. It does not mean that many individual data may not go beyond that boundary. So, actually it is just a matter of agreement. But it works well.

So remember:
Mean ± 2 SDs covers an area under the curve including 95% of the data of the given sample.
Mean ± 2 SEMs covers an area under curve including 95% of the means of many samples, and is, otherwise, called the 95% confidence interval (CI).

In statistical analysis we often compare different samples by taking their sums or differences. Again, this text is not intended to explain the procedures entirely. One more thing to accept unexplainedly is the following. The distributions of the sums as well as those of the difference of samples are again normal distributions and can be characterized by:

$$\text{Sum: } \text{mean}_1 + \text{mean}_2 \pm \sqrt{(\text{SD}_1^2 + \text{SD}_2^2)}$$
$$\text{Difference: } \text{mean}_1 - \text{mean}_2 \pm \sqrt{(\text{SD}_1^2 + \text{SD}_2^2)}$$

$$\text{SEM}_{\text{sum}} = \sqrt{(\text{SD}_1^2/\text{n}_1 + \text{SD}_2^2/\text{n}_2)}$$

$$\text{SEM}_{\text{difference}} = \quad \text{''}$$

Note: If the standard deviations are very different in size, then a more adequate calculation of the pooled SEM is given on page 22.
Sometimes we have paired data where two experiments are performed in one subject or in two members of one family. The variances with paired data are usually smaller than with unpaired because of the positive correlation between two observations in one subject (those who respond well the first time are more likely to do so the second). This phenomenon translates in a slightly modified calculation of variance parameters.

$$\text{SD}_{\text{paired sum}} = \sqrt{(\text{SD}_1^2 + \text{SD}_2^2 + 2\,\text{r}\,\text{SD}_1 \cdot \text{SD}_2)}$$
$$\text{SD}_{\text{paired differrence}} = \sqrt{(\text{SD}_1^2 + \text{SD}_2^2 - 2\,\text{r}\,\text{SD}_1 \cdot \text{SD}_2)}$$

Where r = correlation coefficient, a term that will be explained soon.

Note that SEM does not directly quantify variability in a population. A small SEM can be mainly due to a large sample size rather than tight data.

With small samples the distribution of the means does not exactly follow a Gaussian distribution. But rather a t-distribution, 95% confidence intervals cannot be characterized as the area under the curve between mean ± 2 SEMs but instead the area under curve is substantially wider and is characterized as mean ± t.SEMs where t is close to 2 with large samples but 2.5-3 with samples as small as 5-10. The appropriate t for any sample size is given in the t-table.

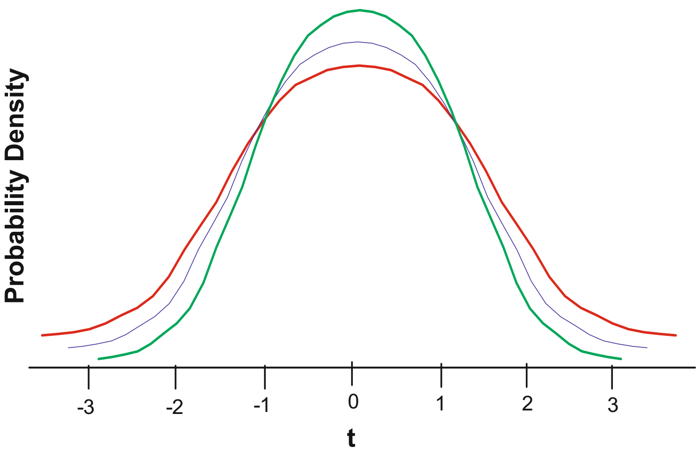


Figure 3. Family of t-distributions: with n = 5 the distribution is wide, with n = 10 and n = 1000 this is increasingly less so.

Figure 3 shows that the t-distribution is wider than the Gaussian distribution with small samples. Mean ± t.SEMs presents the 95% confidence intervals of the means that many similar samples would produce.

Statistics is frequently used to compare more than 2 samples of data. To estimate whether differences between samples are true or just chance we first assess variances in the data between groups and within groups.

Group	sample size	mean	SD
Group 1	n_1	mean_1	SD_1
Group 2	n_2	mean_2	SD_2
Group 3	n_3	mean_3	SD_3

This procedure may seem somewhat awkward in the beginning but in the next two chapters we will observe that variances, which are no less than estimates of noise in the data, are effectively used to test the probabilities of true differences between, e.g., different pharmaceutical compounds. The above data are summarized underneath.

Between-group variance:

Sum of squares$_{\text{between}}$ = SS$_{\text{between}}$ = n_1 (mean$_1$ – overall mean)^2 + n_2(mean$_2$ – overall mean)^2 + n_3 (mean$_3$ – overall mean)^2

Within-group variance:

Sum of squares$_{\text{within}}$ = SS$_{\text{within}}$ = $(n_1 - 1)\ SD_1^2 + (n_2-1)\ SD_2^2 + (n_3-1)\ SD_3^2$

The ratio of the sum of squares between-group/sum of squares within group (after proper adjustment for the sample sizes or degrees of freedom, a term which will be explained later on) is called the big F and determines whether variances between the sample means is larger than expected from the variability within the samples. If so, we reject the null hypothesis of no difference between the samples. With two samples the square root of big F, which actually is the test statistic of analysis of variance (ANOVA), is equal to the t of the famous t-test, which will further be explained in chapter 2. These 10 or so lines already brought us very close to what is currently considered the heart of statistics, namely ANOVA (analysis of variance).

4. DIFFERENT TYPES OF DATA: PROPORTIONS, PERCENTAGES AND CONTINGENCY TABLES

Instead of continuous data, data may also be of a discrete character where two or more alternatives are possible, and, generally, the frequencies of occurrence of each of these possibilities are calculated. The simplest and commonest type of such data are the binary data (yes/no etc). Such data are frequently assessed as proportions or percentages, and follow a socalled binomial distribution. If $0.1<$ proportion (p) <0.9 the binomial distribution becomes very close to the normal distribution. If $p<0.1$, the data will follow a skewed distribution, otherwise called Poisson distribution. Proportional data can be conveniently laid-out as contingency tables. The simplest contingency table looks like this:

	numbers of subjects with side Effect	numbers of subjects without side effect
Test treatment ($group_1$)	a	b
Control treatment ($group_2$)	c	d

The proportion of subjects who had a side effect in $group_1$ (or the risk (**R**) or probability of having an effect):

$$\mathbf{P = a/(a+b)}\text{ , in group}_2\ \mathbf{p = c/(c+d)},$$

The ratios **a/(a+b)** and **c/(c+d)** are called **risk ratios (RRs)**

Note that the terms proportion, risk and probability are frequently used in statistical procedures but that they basically mean the same.

Another approach is the **odds** approach **a/b** and **c/d** are odds and their ratio is the **odds ratio (OR)**.
In clinical trials we use ORs as surrogate RRs, because here a/(a+b) is simply nonsense. For example:

	treatment-group		control-group		entire-population
sleepiness	32	a	4	b	4000
no sleepiness	24	c	52	d	52000

We assume that the control group is just a sample from the entire population but that the ratio b/d is that of the entire population. So, suppose 4 = 4000 and 52 = 52000, then we can approximate $\frac{a/(a+b)}{c/(c+d)} = \frac{a/b}{c/d}$ = RR of the entire population.

Proportions can also be expressed as percentages:

p.100% = a/(a+b). (100%) etc

Just as with continuous data we can calculate SDs and SEMs and 95% confidence intervals of rates (or numbers, or scores) and of proportions or percentages.

SD of number n = $\sqrt{n}$

SD of difference between two numbers n_1 and n_2 = $(n_1 - n_2)/\sqrt{(n_1 + n_2)}$

$$\text{SD proportion} = \sqrt{p(1-p)}$$
$$\text{SEM proportion} = \sqrt{p(1-p)/n}$$

We assume that the distribution of proportions of many samples follows a normal distribution (in this case called the **z**-distribution) with 95% confidence intervals between:

$$p \pm 2\sqrt{p\,(1-p)/n}$$

a formula looking very similar to the 95% CI intervals formula for continuous data

$$\text{mean} \pm 2\sqrt{SD^2\ /n}$$

Differences and sums of the SDs and SEMs of proportions can be calculated similarly to those of continuous data:

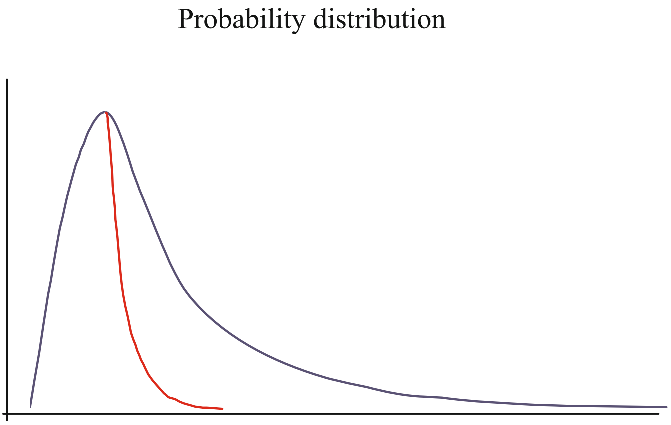
$$\text{SEM}_{\text{of differences}} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

with 95% CI intervals : $p_1 - p_2 \pm 2.$ SEMs

More often than with continuous data, proportions of different samples are assessed for their ratios rather than difference or sum. Calculating the 95% CI intervals of it is not simple. The problem is that the ratios of many samples do not follow a normal distribution, and are extremely skewed. It can never be less than 0 but can get very high. However, the logarithm of the relative risk is approximately symmetrical. Katz's method takes advantage of this symmetry:

$$\text{95\% CI of log RR} = \quad \log RR \ \pm 2\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}$$

This equation calculates the CIs of the logarithm of the RR. Take the antilogarithm (10^x) to determine the 95% CIs of the RR.



Probability distribution

Figure 4. Ratios of proportions unlike continuous data usually do not follow a normal but a skewed distribution (values vary from 0 to ∞). Transformation into the logarithms provides approximately symmetric distributions (thin curve).

Figure 4 shows the distribution of RRs and the distribution of the logarithms of the RRs, and illustrates that the transformation from skewed data into their logarithms is a useful method to obtain an approximately symmetrical distribution, that can be analyzed according to the usual approach of SDs, SEMs and CIs.

5. DIFFERENT TYPES OF DATA: CORRELATION COEFFICIENT

The SD and SEM of paired data includes a term called r as described above. For the calculation of r, otherwise called R, we have to take into account that paired comparisons, e.g., those of two drugs tested in one subject generally have a different variance from those of comparison of two drugs in two different subjects. This is so, because between subject variability of symptoms is eliminated and because the chance of a subject responding beneficially the first time is more likely to respond beneficially the second time as well. We say there is generally a positive correlation between the response of one subject to two treatments.

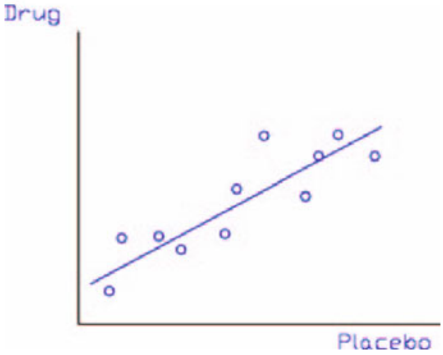


Figure 5. A positive correlation between the response of one subject to two treatments.

Figure 5 gives an example of this phenomenon. X-variables, e.g., blood pressures after the administration of compound 1 or placebo, y-variables blood pressures after the administration of compound 2 or test-treatment.

The SDs and SEMs of the paired sums or differences of the x- and y-variables are relevant to estimate variances in the data and are just as those of continuous data needed before any statistical test can be performed. They can be calculated according to:

$$\mathrm{SD}_{\text{paired sum}} = \sqrt{(\mathrm{SD}_1^2 + \mathrm{SD}_2^2 + 2\,\mathrm{r}\,\mathrm{SD}_1 \cdot \mathrm{SD}_2)}$$

$$\mathrm{SD}_{\text{paired differrence}} = \sqrt{(\mathrm{SD}_1^2 + \mathrm{SD}_2^2 - 2\,\mathrm{r}\,\mathrm{SD}_1 \cdot \mathrm{SD}_2)}$$

where r = correlation coefficient, a term that will be explained soon.

Likewise:

$$\mathrm{SEM}_{\text{paired sum}} = \sqrt{(\mathrm{SD}_1^2 + \mathrm{SD}_2^2 + 2\,\mathrm{r}\,\mathrm{SD}_1 \cdot \mathrm{SD}_2)\,/\mathrm{n}}$$

$$\mathrm{SEM}_{\text{paired differrence}} = \sqrt{(\mathrm{SD}_1^2 + \mathrm{SD}_2^2 - 2\,\mathrm{r}\,\mathrm{SD}_1 \cdot \mathrm{SD}_2)\,/\mathrm{n}}$$

where $n = n_1 = n_2$

and that:

$$\mathrm{r} = \frac{\sum (\mathrm{x} - \bar{\mathrm{x}})(\mathrm{y} - \bar{\mathrm{y}})}{\sqrt{\sum (\mathrm{x} - \bar{\mathrm{x}})^2 \sum (\mathrm{y} - \bar{\mathrm{y}})^2}}$$

r is between –1 and +1, and with unpaired data r = 0 and the SD and SEM formulas reduce accordingly (as described above). Figure 5 also shows a line, called the regression line, which presents the best-fit summary of the data, and is the calculated method that minimizes the squares of the distances from the line.

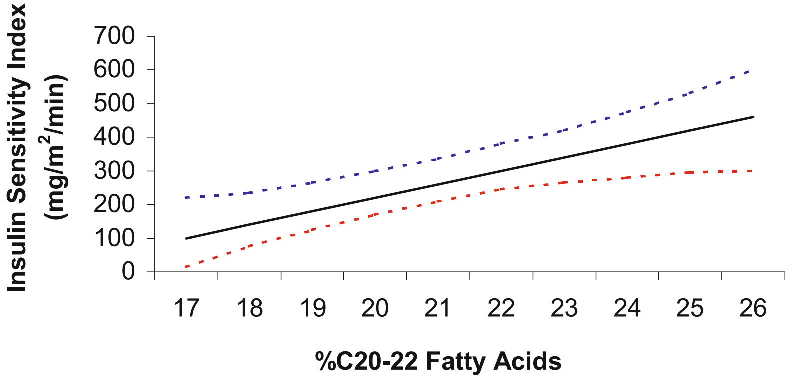


Figure 6. Example of a linear regression line of 2 paired variables (x- and y-values), the regression line provides the best fit line. The dotted curves are 95% CIs that are curved, although we do not allow for a nonlinear relationship between x and y variables.

The 95% CIs of a regression line can be calculated and is drawn as area between the dotted lines in Figure 6. It is remarkable that the borders of the straight regression line are curved although we do not allow for a nonlinear relationship between the x-axis and y-axis variables. More details on regression analysis will be given in chapters 2 and 3.

In the above few lines we described continuous normally distributed or t-distributed data, and rates and their proportions or percentages. We did not yet address data ordered as ranks. This is a special method to transform skewed data into a approximately normal distribution, and is in that sense comparable with logarithmic transformation of relative risks (RRs). In chapter 3 the tests involving this method will be explained.

6. STRATIFICATION ISSUES

When published, a randomized parallel-group drug trial essentially includes a table listing all of the factors, otherwise called baseline characteristics, known possibly to influence outcome. E.g., in case of heart disease these will probably include apart from age and gender, the prevalence in each group of diabetes, hypertension, cholesterol levels, smoking history. If such factors are similar in the two groups, then we can go on to attribute any difference in outcome to the effect of test-treatment over reference-treatment. If not, we have a problem. Attempts are made to retrieve the situation by multivariate analysis allocating part of the differences in outcome to the differences in the groups, but there is always an air of uncertainty about the validity of the overall conclusions in such a trial. This issue is discussed and methods are explained in chapter 8. Here we discuss ways to avoid this problem. Ways to do so, are stratification of the analysis and minimization of imbalance between treatment groups, which are both techniques not well-known. Stratification of the analysis means that relatively homogeneous subgroups are analyzed separately. The limitation of this approach is that it can not account for more than two, maybe three, variables and that thus major covariates may be missed. Minimization can manage more factors. The investigators first classify patients according to the factors they would like to see equally presented in the two groups, then randomly assign treatment so that predetermined approximately fixed proportions of patients from each stratum receive each treatment. With this method the group allocation does not rely solely on chance but is designed to reduce any difference in the distribution of unsuspected contributing determinants of outcome so that any treatment difference can now be attributed to the treatment comparison itself. A good example of this method can be found in a study by Kallis et al.[1] The authors stratified in a study of aspirin versus placebo before coronary artery surgery the groups according to age, gender, left ventricular function, and number of coronary arteries affected. Any other prognostic factors other than treatment can be chosen. If the treatments are given in a double-blind fashion, minimization influences the composition of the two groups but does not influence the chance of one group entering in a particular treatment arm rather than the other.

There is an additional argument in favor of stratification/minimization that counts even if the risk of significant asymmetries in the treatment groups is small. Some prognostic factors have a particularly large effect on the outcome of a trial. Even small and statistically insignificant imbalances in the treatment groups may then bias the results. E.g., in a study of two treatment modalities for pneumonia[2] including 54 patients, 10 patients took prior antibiotic in the treatment group and 5 did in the control group. Even though the difference between 5/27 and 10/27 is not statistically significant, the validity of this trial was being challenged, and the results were eventually not accepted.

7. RANDOMIZED VERSUS HISTORICAL CONTROLS

A randomized clinical trial is frequently used in drug research. However, there is considerable opposition to the use of this design. One major concern is the ethical problem of allowing a random event to determine a patient's treatment. Freirich[3] argued that a comparative trial which shows major differences between two treatments is a bad trial because half of the patients have received an inferior treatment. On the other hand in a prospective trial randomly assigning treatments avoids many potential biases. Of more concern is the trial in which a new treatment is compared to an old treatment when there is information about the efficacy of the old treatment through historical data. In this situation use of the historical data for comparisons with data from the new treatment will shorten the length of the study because all patients can be assigned to the new treatment. The current availability of multivariate statistical procedures which can adjust the comparison of two treatments for differing presence of other prognostic factors in the two treatment arms, has made the use of historical controls more appealing. This has made randomization less necessary as a mechanism for ensuring comparability of the treatment arms. The weak point in this approach is the absolute faith one has to place in the multivariate model. Also, some confounding variables e.g., time effects, simply can not be adjusted, and remain unknown. Despite the ethical argument in favor of historical controls we must therefore emphasize the potentially misleading aspects of trials using historical controls.

8. FACTORIAL DESIGNS

The majority of drug trials are designed to answer a single question. However, in practice many diseases require a combination of more than one treatment modalities. E.g., beta-blockers are effective for stable angina pectoris but beta-blockers plus calcium channel blockers or beta-blockers plus calcium channel blockers plus nitrates are better (Table 1). Not addressing more than one treatment modality in a trial is an unnecessary restriction on the design of the trial because the assessment of two or more modalities in on a trial pose no major mathematical problems.

Table 1. The factorial design for angina pectoris patients treated with calcium channel blockers with or without beta-blockers

	Calcium channel blocker	no calcium channel blocker
Beta-blocker	regimen I	regimen II
No beta-blocker	regimen III	regimen I

We will not describe the analytical details of such a design but researchers should not be reluctant to consider designs of such types. This is particularly so, when the recruitment of large samples causes difficulties.

9. CONCLUSIONS

What you should know after reading this chapter:

1. Scientific rules governing controlled clinical trials include prior hypotheses, valid designs, strict description of the methods, uniform data analysis.
2. Efficacy data and safety data often involve respectively continuous and proportional data.
3. How to calculate standard deviations and standard errors of the data.