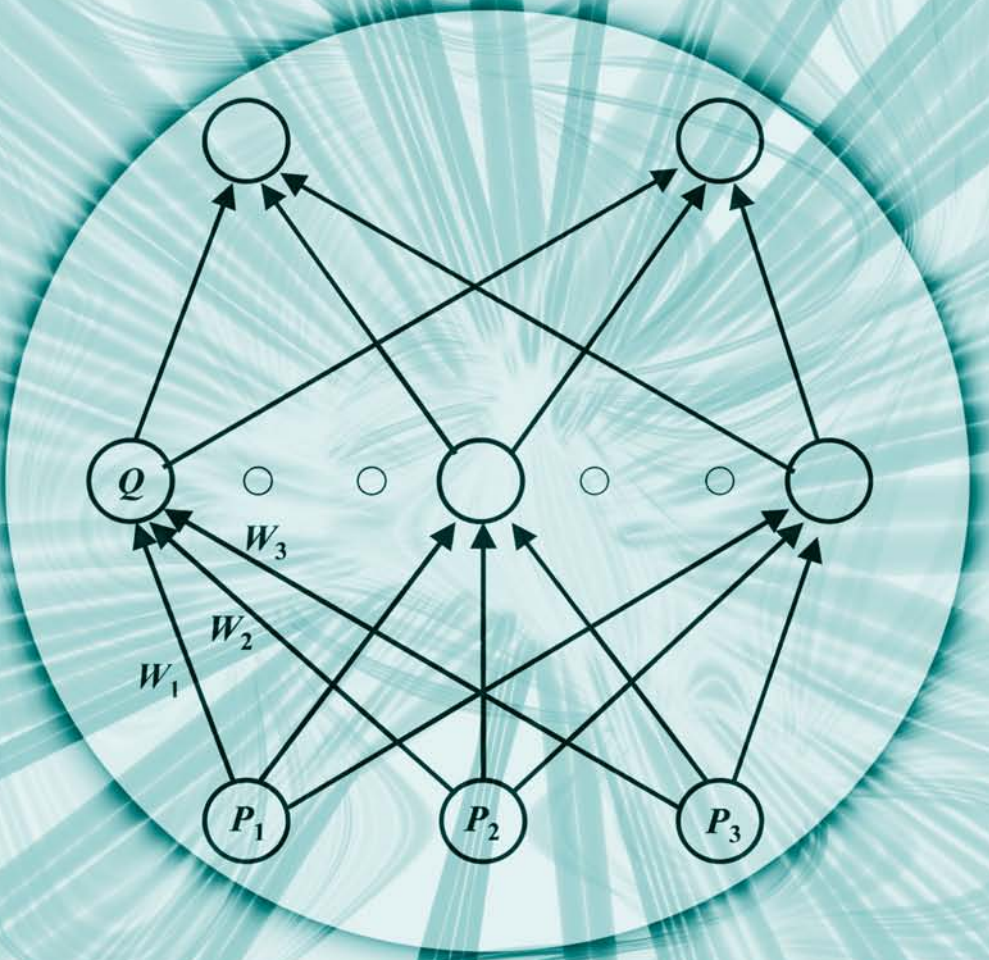
N. Anthony Armstrong

Pharmaceutical Experimental Design and Interpretation

Second Edition



CRC Taylor & Francis
Taylor & Francis Group

Pharmaceutical Experimental Design and Interpretation

Second Edition

Pharmaceutical Experimental Design and Interpretation

Second Edition

N. Anthony Armstrong
University of Cardiff
UK

Taylor & Francis
Taylor & Francis Group
Boca Raton London New York

A CRC title, part of the Taylor & Francis imprint, a member of the Taylor & Francis Group, the academic division of T&F Informa plc.

Published in 2006 by
CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-10: 0-415-29901-2 (Hardcover)
International Standard Book Number-13: 978-0-415-29901-5 (Hardcover)
Library of Congress Card Number 2005024462

Library of Congress Cataloging-in-Publication Data

Armstrong, N. A. (N. Anthony)
Pharmaceutical experimental design and interpretation / N. Anthony Armstrong. -- 2nd ed.
p. ; cm.
Includes bibliographical references and index.
ISBN-10: 0-415-29901-2 (Hardcover : alk. paper)
ISBN-13: 978-0-415-29901-5 (Hardcover : alk. paper)
1. Drugs--Research--Methodology. 2. Experimental design. I. Title.
[DNLM: 1. Drug Design. 2. Data Interpretation, Statistical. 3. Research Design. QV 744 A737p 2006]

RM301.25.A76 2006
615'.19'072--dc22 2005024462

T&F informa
Taylor & Francis Group
is the Academic Division of T&F Informa plc.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Author

Norman Anthony Armstrong graduated B.Pharm. and Ph.D. from London University. After some years in the pharmaceutical industry, Dr. Armstrong joined the Welsh School of Pharmacy, Cardiff University, U.K., where he became senior lecturer in pharmaceutical technology. He retired from that position in 2002.

Dr. Armstrong is a fellow of the Royal Pharmaceutical Society of Great Britain and is the author of over 150 scientific papers, reviews, and books.

Table of Contents

Dedication

for

Kenneth Charles James, 1926–1997

The first two editions of this book were written in collaboration with Dr. Kenneth Charles James, reader in pharmaceutics at the Welsh School of Pharmacy, Cardiff University. Sadly, just as the second edition was being completed, Ken's health deteriorated and he died shortly after its publication.

This edition is therefore dedicated to the memory of Ken James, mentor, colleague, and friend.

1 Introduction to Experimental Design

1.1 THE EXPERIMENTAL PROCESS

Experimentation is expensive in terms of time, work force, and resources. It is therefore reasonable to ask whether experimentation can be made more efficient, thereby reducing expenditure of time and money.

Scientific principles of experimental design have been available for some time now. Much of the work originated with Sir Ronald Fisher and Professor Frank Yates, who worked together at Rothamsted Agricultural Research, U.K.[1] The principles that they and others devised have found application in many areas, but it is surprising how little these principles have been used in pharmaceutical systems. The reasons for this neglect are a matter of speculation, but there is no doubt that principles of experimental design do have a widespread applicability to the solution of pharmaceutical problems.

Experimentation may be defined as the investigation of a defined area with a firm objective, using appropriate tools and drawing conclusions that are justified by the experimental data so obtained. Most experiments consist of measuring the effect that one or more factors have on the outcome of the experiment. The factors are the independent variables, and the outcome is the response or dependent variable.

The overall experimental process may be divided into the following stages:

1. Statement of the problem. What is the experiment supposed to achieve? What is its objective?
2. Choice of factors to be investigated, and the levels of those factors that are to be used.
3. Selection of a suitable response. This may be defined in Stage 1, statement of the problem. If so, then we must be sure that the measurement of the chosen response contributes to achieving the objective. The proposed methods of measuring the response and their accuracy must also be considered at this stage.
4. Choice of the experimental design. This is often a balance between cost and statistical validity. The more an experiment is replicated, the greater the reliability of the results. However, replication increases cost, and the experimenter must therefore consider what is an acceptable degree of uncertainty. This in turn is governed by the number of replicates that can be afforded. Inextricably linked with this stage is selection of the method to be used to analyze data.

5. Performance of the experiment: the data collection process.
6. Data analysis.
7. Drawing conclusions.

The steps in the process may be illustrated using a simple example that is developed further in Chapter 4. Gebre-Mariam et al.[2] investigated the relationship between the composition of mixtures of glycerol and water and the viscosity of those mixtures, as part of a study of diffusion through gels.

Thus, the objective (Stage 1) was to establish the dependence of the viscosity of glycerol–water mixtures on their composition. The factor to be investigated (Stage 2) was composition of the mixture up to a maximum of about 40% w/w glycerol. The response (Stage 3) was the viscosity of the liquids, measured by an appropriately accurate method, in this case a U-tube viscometer. Because only one factor was to be investigated, any other factor that might influence the response had to be eliminated or kept constant. Temperature was an obvious example in this case.

At the outset, it was not known whether the relationship would be rectilinear or curvilinear. Furthermore, results were to be fitted to a model equation, and for both these reasons, an adequate number of data points had to be obtained. Five concentrations of glycerol were selected, covering the desired range (Stage 4). This was expected to be the minimum number that would enable a valid regression analysis to be performed. Many data points could have been used, thereby improving the reliability of any relationship, but of course this would have involved additional work.

The experiments were then carried out (Stage 5), the data was subjected to regression analysis (Stage 6), and the relationship between composition and viscosity was established (Stage 7).

Thus, the experimental design and the method to be used to analyze the data are selected before the experiment is carried out. Conclusions that can be drawn from the data depend, to a large extent, on the manner in which the data were collected. Oftentimes, the objective of the experiment is imperfectly defined, the experiment is then carried out, and only after these are methods of data analysis considered. It is then discovered that the experimental design is deficient and has provided insufficient or inappropriate data for the most effective form of analysis to be carried out. Thus, the term experimental design must include not only the proposed experimental methodology, but also the methods whereby the data from the experiments is to be analyzed. The importance of considering both parts of this definition together cannot be overemphasized.

1.2 COMPUTERS AND EXPERIMENTAL DESIGN

A point that must be considered at this stage is the availability of computing facilities such as mainframes, personal computers (PCs), and even a pocket calculator. The advantages of the computer are obvious. The chore of repetitive calculation has been removed as well as an undeniable disincentive to use statistical methods. However,

using a computer can cause two related problems. The first is absolute reliance on the computer — if the computer says so, it must be so. The second is the assumption that the computer can take unreliable data or data from a badly designed experiment and somehow transform them into a result which can be relied upon. The computer jargon GIGO — garbage in, garbage out — is just as appropriate to problems of experimental design as to other areas in which computers are used.

It is undeniable that access to a computer is invaluable. Many readers will have access to a mainframe computer equipped with comprehensive statistical packages including SPSS® (McGraw-Hill, New York, NY, USA), SAS® (SAS Institute, Cary, NC, USA), and MINITAB® (Minitab, State College, PA, USA). Bohidar[3] has described the application of SAS to problems of pharmaceutical formulation.

MINITAB contains many features that are relevant to experimental design. In addition to useful statistical techniques, it includes programs for determinant analysis and principal component analysis (Chapter 5). The commands FFDESIGN and PBDESIGN generate fractional factorial designs and Plackett–Burman designs respectively for a specified number of experimental factors (Chapter 6). Randomization of the order in which the experiments are to be performed can also be carried out. The command FFACTORIAL analyzes data from experiments based on these designs, and facilities for drawing contour plots from the data are also available (Chapters 7 and 8). Details are given in Ryan and Joiner.[4]

However, a desktop computer will suffice for many of the calculations described in this book, because many statistical packages for PCs are now commercially available. Spreadsheet packages such as Lotus 1-2-3® (Lotus Development Corporation, Cambridge, MA, USA) and Excel® (Microsoft Corporation, Redmond, WA, USA) are of great value for these calculations.[5] The latter is used extensively in this book.

Several software packages specifically intended for experimental design and optimization purposes are also available. One example is the RS/Discover® suite of programs from BBN Software Products Corporation (Cambridge, MA, USA). The menu-driven program in this package prompts the user to specify the independent variables, together with their units, the ranges of values for the variables, and the required degree of precision and to indicate whether the value of a given variable can be easily altered. The program then produces a worksheet that gives the design of the experiment (full factorial, central composite, etc.) and the values of the independent variables for each experiment. The experiments are usually given in random order, except in those cases where a particular experimental variable cannot be easily altered in value. In such cases, the experiments are grouped so that the time taken to alter that variable is minimized. After the experiments are carried out, the responses are added to the worksheet. Data can then be analyzed and fitted to models and contour plots, and response surfaces can be produced. Applications of this package have been reported by McGurk et al.[6]

The Design-Ease® and Design-Expert® packages offered by Stat-Ease (Minneapolis, MN, USA) provide facilities for the design and analysis of factorial experiments. The programs generate worksheets of experiments in random order or in blocks for experiments involving process variables or mixtures and, from the results, can produce a statistical analysis and three-dimensional response surface and contour graphs.

Similar programs include ECHIP® (Expert on a Chip, Hockessin, DE, USA), which has been reviewed by Dobberstein et al.,[7] CHEOPS® (Chemical Operations by Simplex, Elsevier Scientific Software, Amsterdam, The Netherlands), Statgraphics Plus® (Statgraphics, Rockville, MD, USA), and CODEX® (Chemometrical Optimisation and Design for Experimenters, AP Scientific Services, Stockholm, Sweden).

1.3 OVERVIEW OF EXPERIMENTAL DESIGN AND INTERPRETATION

This is not a textbook on statistics. However, some statistical knowledge is essential if the full power of techniques in experimental design is to be appreciated. Neither is this a compendium of methods of experimental design. Rather, it discusses methods that are of value in the design of experiments and in the interpretation of results obtained from them.

The literature in this area is considerable, and for readers wishing to develop their knowledge of a particular technique, references to further reading are given at the end of each chapter. Moreover, statistical textbooks and some general texts on experimental design are given at the end of this chapter.

Many experiments consist in acquiring groups of data points, each group having been subjected to a different treatment, and methods for evaluating data from such experiments are included in Chapter 2. Essentially, these methods are based on establishing whether the mean values of the various groups differ significantly. When there are only two groups of data, Student's *t*-test is usually applied, but for three or more groups, analysis of variance is the method of choice. The latter also forms the basis of many of the methods of experimental design described in later chapters.

For Student's *t*-test and analysis of variance to be applicable, the data should, strictly speaking, be normally distributed about the mean and must have true numerical values. Such tests cannot be applied to adjectival information or when data have been assigned to numbered but arbitrarily designated categories. In such cases, nonparametric methods come into their own. These methods do not depend for their validity on a normal or Gaussian distribution, and "adjectival" data can be assessed using them. However, such methods depend on the presence of an adequate number of data points to facilitate comparison, and hence the degree of replication in the experiment must be appropriate if such methods are to be used. Nonparametric methods involve either paired data, where each subject acts as its own control, or unpaired data. Both are discussed in Chapter 3.

Having obtained raw data from the experiment, one has to decide on how best to use them. The decision may be simple; for example, all that is required is a mean value and standard deviation or the plot of one value against another, which gives a perfect straight line. Usually, more is required, in which case the statistical method that is most appropriate to the problem must be chosen.

An obvious example involves a series of pairs of results where it is required to know whether they are related, and if so how. A simple example could be the variation of the weights of a collection of laboratory animals with their heights. A plot of height (h) against weight (w) drawn on a graph paper may not give a definite answer,

because the points could be such that it is not clear whether or not the results are scattered around a straight line. The probability that the results are so related is given by regression analysis, together with the value of the line in predicting unknown results. Alternatively, the relationship may be curved but fits a quadratic equation.

If the results are not related, a third property, for example, age (A), may make an important contribution. It is not possible to plot a graph in this situation, although one could construct a three-dimensional model.

It is not possible to visually represent equations with more than three variables, but such higher relationships can be expressed by an equation. Thus, for example, if the variation of animals' weights (w) with height, age (A), and waist circumference (c) is examined, a relationship of the form shown in (1.1) can be devised:

$$w = b_0 + b_1 h + b_2 A + b_3 c \tag{1.1}$$

in which b_0, b_1, b_2, and b_3 are constants and can be derived by regression analysis. A minimum of four sets of data (because there are four variables) would be required to derive such an equation, and a perfect relationship would result. For a reliable relationship, a minimum of five sets of data for each unknown, giving a minimum of 20 sets of results, are necessary.

Other relationships can be detected, either by trial and error or by suspected relationships, derived theoretically or found for similar systems in the literature; for example, logarithmic (1.2), ternary (1.3), or square root (1.4). Some examples are given in the book, and methods for calculating them and evaluating their reliability are described.

$$y = b_0 + b_1 \log x \tag{1.2}$$

$$y = b_0 + b_1 + b_2 x^2 + b_3 x^3 \tag{1.3}$$

$$y = b_0 + b_1 x^{1/2} \tag{1.4}$$

Regression analysis looks for relationships between a dependent variable and one or more independent variables. This method of analysis is called a univariate method. Multivariate methods look for relationships between several variables, considering them collectively. These data are often presented in the form of a matrix, an example of which follows:

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \\ d_1 & d_2 & d_3 & d_4 \end{bmatrix} \tag{1.5}$$

Each column represents a property of the materials under examination. For example, 1 could represent tablet weight, 2 disintegration time, 3 crushing strength, and 4 moisture content. Each row represents a combination of the properties of one example, in this case the properties of a different tablet formulation. To work

with these, one must have a knowledge of matrices and their manipulation, which differs from basic algebraic methods. The basic matrix algebra necessary to understand this section is given in Appendix 2, followed by examples of their use.

When a series of results is presented, the individual results can frequently be arranged into unrelated groups, within which the results are related. This is called cluster analysis. Alternatively, the validity of preconceived classifications can be examined by discrimination analysis.

Relationships within sets of results can often be detected and used to simplify data. Thus, the number of rows shown in (1.5) could possibly be reduced to three or even less by principal components analysis and the columns reduced in a similar manner by factor analysis. Cluster, discrimination, principal components, and factor analysis are all described in Chapter 5.

Experimental programs can, if not efficiently designed, consume much time, materials, and labor, and hence, it is essential that programs be designed in the most cost-effective manner. In Chapter 6, the principles of factorial design are discussed. Factorial design, when allied to statistical techniques such as analysis of variance, is a powerful tool for gaining the maximum amount of information from a limited number of experiments.

Factorial design involves the variation of two or more experimental variables or factors in a planned manner, and the factors are investigated at two or more levels. The technique establishes the relative order of importance of the factors and can also indicate whether factors interact and whether such interactions are significant.

Even so, full factorial designs involving several factors at three or even more levels can demand considerable resources. Therefore, methods by which the number of experiments can be reduced in factorial designs are also explored. The potential hazards of using such limited designs are also discussed.

Having determined which factors and interactions make a significant contribution to the response, one can use the same experiments to predict the response for combinations of factors that have not been studied experimentally. The prediction is carried out by deriving a mathematical model relating the factors to the response. The construction of the model equation and establishing its validity draw heavily on correlation and regression techniques described in Chapter 4.

Once the model is established, it can be used to construct contour plots. These plots are diagrams of the value of the response in terms of the values of the experimental variables. The model can also be used to derive the response surface. This is usually a three-dimensional diagram, with the response plotted on the vertical axis and two factors forming the horizontal axes. Such diagrams are invaluable in visualizing relationships between independent and dependent variables and also in assessing the robustness of the response. Both are described in Chapter 7.

Many pharmaceutical formulations and processes lend themselves to optimization procedures, whereby the best possible result is sought, given a series of limits or constraints. Thus, the best possible solution is not necessarily a maximum (or minimum) value, but is rather a compromise, taking many factors into account. There are two principal methods of optimization. One is model-dependent optimization, in which a group of experiments is carried out and the results are then fitted to an equation (the model). Such techniques are discussed in Chapter 8.

Model-dependent methods require that a series of experiments should be carried out and the results assessed only when the whole series has been completed. Methods by which the results of only a few experiments govern the conditions of further experiments are sequential or model independent, and the results are examined continuously as they become available. No attempt is made to express results in a model equation. Such methods are described in Chapter 9, which also includes a comparison between model-dependent and model-independent techniques.

Many pharmaceutical formulations involve mixtures of several ingredients, the total mass or volume of which is fixed. The composition of a fixed-volume injection or the contents of a hard-shell capsule are good examples. Here, if the proportion of one ingredient is changed, then the proportion of at least one of the others must also change. Such mixtures are amenable to the principles of experimental design, the applications of which are described in Chapter 10.

In the final chapter, the use of artificial neural networks in pharmaceutical experimental design is considered (Chapter 11). Artificial neural networks are machines that learn from experience, in a similar manner to the brain. Their underlying function is to identify patterns, that is, to recognize the relationship between input data and the corresponding response. These relationships are then applied in a predictive manner.

Each chapter is illustrated by a number of worked examples. Their selection has sometimes caused problems. Inevitably the author has tended to select examples which he has found of value, and which are therefore in fields in which he is personally interested. However he accepts that there are many other areas of pharmaceutical science that could have been explored. Therefore, many of the chapters end with a bibliography that indicates those areas where a particular technique has been used, and the reader is referred to the original articles.

The appendices of the book contain material to which reference may be required, but which would be intrusive if it was contained in the main body itself. Tabulated statistical data (e.g., values of Student's *t*-test, *F*-test, and correlation coefficients at given significance levels) has been reduced to a minimum and only includes material that is needed in the worked examples used in the book. Complete tables are readily available elsewhere.

Useful Statistical Texts

Bolton, S. and Bon, C., *Pharmaceutical Statistics: Practical and Clinical Applications*, 4th ed., Marcel Dekker, New York, 2004.

Clarke, G. M. and Cooke, D. A., *A Basic Course in Statistics*, 4th ed., Arnold, London, 1998.

Jones, D. S., *Pharmaceutical Statistics*, Pharmaceutical Press, London, 2002.

Useful General Texts on Experimental Design

Anderson, V. L. and McLean, R. A., *Design of Experiments: A Realistic Approach*, Marcel Dekker, New York, 1974.

Box, G. E. P., Hunter, W. G., and Hunter, J. S., *Statistics for Experimenters: Introduction to Design, Data Analysis and Model Building*, Wiley, New York, 1978.

Cornell, J. A., *Experiments with Mixtures*, 3rd ed., Wiley, New York, 2002.

Fisher, R. A. and Yates, F., *The Design of Experiments*, 8th ed., Oliver & Boyd, Edinburgh, 1966.

Hicks, C. R. and Turner, K. V., *Fundamental Concepts in the Design of Experiments*, 5th ed., Oxford University Press, Oxford, 1999.
Montgomery, D. C., *Design and Analysis of Experiments*, 5th ed., Wiley, New York, 2001.
Strange, R. S., Introduction to experiment design for chemists, *J. Chem. Educ*., 67, 113, 1990.

REFERENCES

1. Fisher, R. A., *The Design of Experiments*, Oliver & Boyd, London, 1926.
2. Gebre-Mariam, T. et al., The use of electron spin resonance to measure microviscosity, *J. Pharm. Pharmacol*., 43, 510, 1991.
3. Bohidar, N. R., Pharmaceutical formulation optimization using SAS, *Drug Dev. Ind. Pharm*., 17, 421, 1991.
4. Ryan, B. F. and Joiner, B. L., *Minitab Handbook*, 4th ed., Duxbury Press, Pacific Grove, 2001.
5. Dranchuk, J., *Excel for Windows Spreadsheet Databases*, Wiley, New York, 1994.
6. McGurk, J. G., Storey, R., and Lendrem, D. W., Computer-aided process optimisation, *J. Pharm. Pharmacol*., 41, 128P, 1989.
7. Dobberstein, R. H. et al., Computer-assisted experimental design in pharmaceutical formulation, *Pharm. Technol*., 3, 84, 1994.

2 Comparison of Mean Values

2.1 INTRODUCTION

A common feature of many experimental programs is to obtain groups of data under two or more sets of experimental conditions. The question then arises: Has the change in experimental conditions affected the data? The question may be rephrased to a more precise form: Do the means of each group differ significantly or are all groups really taken from the same population, the change in experimental conditions having had no significant effect? A variety of experimental techniques exist to answer this question. Hence, it is all too easy to select an inappropriate technique, with misleading results.

For selecting the correct procedure, further questions must be asked:

1. Are the data truly numerical? Some data are purely nominal, in that they are given a name, for example, male or female, black or white. Such data, especially if they are to be processed by a computer, are often given a numerical value, for example, male = 0, female = 1, but these are labels, not actual numbers. Data can also be ordinal, in that they are ranked. For example, five children can be ranked in order of increasing height, with the value of 1 assigned to the shortest child and 5 to the tallest. These are not truly numerical values, in that the series does not represent a scale with equal intervals. Thus, there is no suggestion that the difference in height between numbers 1 and 2 is the same as that between 2 and 3. If, however, the actual heights of the children had been used, then these are truly numerical data and can be used in the tests described below.
2. Are there more than two sets of data?
3. Are the data normally distributed?
4. Are there many data points in each group (more than 30)?
5. If there are only two sets of data, do these sets represent the total population or do they represent samples drawn from a larger population? Do we know the variance of the whole population? Examples of the former could be sets of examination results, when the performance of every candidate is known. Also, in a long-running industrial process, where many batches have been made under identical conditions, the pooled variance of all the batches will be very close to or even equal to the variance of the total population or universe.
6. Are the data paired or unpaired?

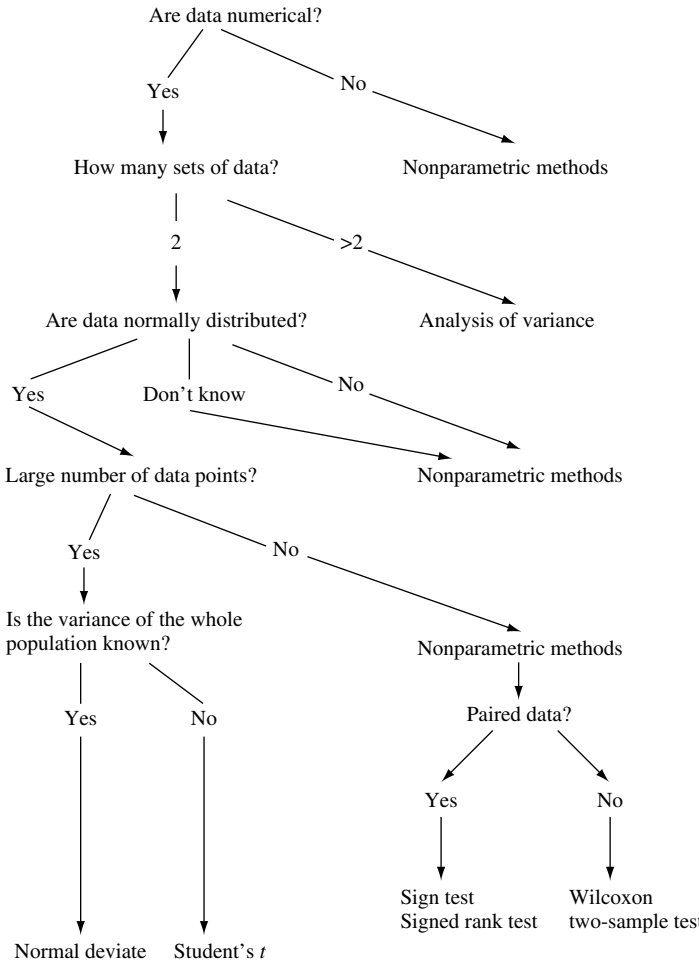


FIGURE 2.1 Chart to help select the correct statistical test for comparison of the means of groups of data.

Figure 2.1 shows Questions 1 to 6 in a diagrammatic form.

The available procedures can best be illustrated by examples which, though apparently straightforward, will serve as media through which several aspects of experimental design can be explored.

2.2 COMPARISON OF MEANS WHEN THE VARIANCE OF THE WHOLE POPULATION IS KNOWN

Twenty university students are taught a given subject in two groups of ten (Groups A and B), each group having its own tutor. At the end of the course, all 20 students take the same examination, the results of which are shown in Table 2.1.

TABLE 2.1
Marks Obtained by Two Groups of Ten Students (%)

	Group A	Group B
	70	66
	60	56
	59	55
	56	53
	56	48
	54	45
	52	45
	51	44
	44	42
	44	38
n	10	10
Mean	54.6	49.2
Variance	53.4	61.8
Standard deviation	7.3	7.9

The means differ by over 5% on marks of about 50%, which seems quite large. On the other hand, the values of the standard deviations show that there is considerable scatter around each mean. The university is concerned by the difference in mean marks between the two groups and wishes to assess whether this difference is statistically significant.

Figure 2.1 shows that use of the normal deviate is an appropriate test, because the data relate to the whole of the population and not just to samples. The procedure is to use the normal deviate to construct confidence intervals for the means. The confidence interval for Group A is given by (2.1):

$$\text{confidence interval} = x_{\text{mA}} \pm \left(\frac{Z_P \sigma}{\sqrt{n_{\text{A}}}} \right) \tag{2.1}$$

where

x_{mA} = mean of Group A
σ = standard deviation of Group A
n_{A} = number of observations in Group A
P = required level of probability
Z_P = normal deviate corresponding to the $(P+1)/2$ percentile of the cumulative standard normal distribution.

Thus, a key point to be decided is the required level of probability, as this governs the value of Z_P, and this decision must be taken before the calculation can be made. In most physicochemical experiments, a significance level of 0.05 is selected, which means that there is a 1 in 20 chance of the wrong inference being made.

Table A1.1 in Appendix 1 summarizes a selection of values of the standard normal variable. The value to choose is that corresponding to $(P+1)/2$, which in this case is 0.975. Therefore, $Z=1.96$.
Hence, the confidence interval for the mean of Group A is

$$54.6 \pm \frac{1.96 \times 7.3}{\sqrt{10}} = 54.6 \pm 4.5 = 50.1 \text{ to } 59.1$$

The mean of Group B falls outside this range; therefore, it can be concluded that, at this level of significance, there is a difference between the means. It may be, however, that the university foresees serious consequences if a significant difference between the performances of the two groups is established. It therefore decides to choose a significance level of 0.01, so that there is now only a 1 in 100 chance of an incorrect inference being made. The chosen value of the standard normal deviate now corresponds to $(P+1)/2=0.995$. Therefore, the tabulated value of Z is now 2.58. Substituting this into (2.1) gives the confidence interval for the mean of Group A as 48.6 to 60.6. The mean of Group B lies within this range, and it could be claimed that there is no significant difference between the means. Thus, whether a significant difference exists depends on the level of significance that is chosen. This, in turn, is selected with the consequences of drawing the wrong conclusion firmly in mind.

2.3 COMPARISON OF TWO MEANS WHEN THE VARIANCE OF THE WHOLE POPULATION IS NOT KNOWN

In the previous example, every member of the population (all 20 students) was tested. In many cases, however, this is not feasible. The total population may be too high for it all to be tested or the testing may be destructive. In such cases, the variance must be estimated from data obtained from samples.

As an example, consider the following situation. Hard-shell capsules are filled with a mixture of active ingredients and diluents (Formulation A). A new formulation is devised (Formulation B) which, it is believed, will alter the disintegration times of the capsules. The objective of the experiment is therefore to establish whether a significant difference exists between the mean disintegration times of the two formulations. The capsules are subjected to the disintegration test of the European Pharmacopoeia, and the results are given in Table 2.2.

Figure 2.1 indicates that the appropriate test in this case is Student's *t*-test. There are two formulae that can be used for calculating *t*. The first is (2.2):

$$t = \frac{x_{mA} - x_{mB}}{\sqrt{s_p^2 \left(1/n_A + 1/n_B\right)}} \tag{2.2}$$

where

x_{mA} and x_{mB} = means from Formulations A and B, respectively
n_A and n_B = the number of data points in each group

TABLE 2.2
Disintegration Time (Minutes) of Hard-Shell Capsules Containing Two Formulations, A and B

	Formulation A	Formulation B
	11.1	9.2
	10.3	10.3
	13.0	11.2
	14.3	11.3
	11.2	10.5
	14.7	9.5
n	6	6
Mean	12.43	10.33
Variance	3.36	0.74
Standard deviation	1.83	0.86

S_p^2 = the pooled variance, which is, in turn, given by (2.3)

$$s_p^{\,2} = \frac{(n_A - 1)s_A^{\,2} + (n_B - 1)s_B^{\,2}}{(n_A + n_B - 2)} \tag{2.3}$$

where

S_A^2 and S_B^2 = the variances of the data from Formulations A and B, respectively.

Alternatively, t can be calculated from (2.4)

$$t = \frac{x_{mA} - x_{mB}}{\sqrt{\left(s_A^{\,2}/n_A + s_B^{\,2}/n_B\right)}} \tag{2.4}$$

Equation (2.2) is used when the variances of the two sets of data do not differ considerably. A ratio between the variances of less than 3 is a good rule of thumb. If the variances differ by more than this, (2.4) is used instead. Use of (2.4) gives a more conservative estimate of significance than (2.2), even when both samples have similar variances. For the data shown in Table 2.2, the ratio of the variances is 4.8 (3.4:0.7); therefore, (2.4) is used to calculate t.

Statistical tests such as Student's t involve comparison of a value of t calculated from the data with a tabulated value. If the calculated value exceeds the tabulated value, then a significant difference between the means of the two groups has been detected. Tabulated values of t are shown in Table A1.2 in Appendix 1. Before the correct tabulated value can be selected, two items of information are required, which are in turn dependent on the design of the experiment. The first is the required level of significance, that is, the required value of P