Learning to use statistical tests in psychology

SECOND EDITION

Judith Greene
and
Manuela d'Oliveira

Open University Press
Buckingham · Philadelphia

Open University Press
Celtic Court
22 Ballmoor
Buckingham
MK18 1XW

e-mail: enquiries@openup.co.uk
world wide web: http://www.openup.co.uk

and
325 Chestnut Street
Philadelphia, PA 19106, USA

First published 1982
Reprinted 1983, 1985, 1987, 1988, 1989, 1990, 1992, 1993, 1994, 1995, 1996, 1998

First published in this second edition, 1999

A catalogue record of this book is available from the British Library

ISBN 0 335 20378 7 (hb) 0 335 20377 9 (pb)

Library of Congress Cataloging-in-Publication Data
Greene, Judith.
Learning to use statistical tests in psychology/Judith Greene and Manuela d'Oliveira. – 2nd ed.
p. cm. – (Open guides to psychology)
Includes bibliographical references and index.
ISBN 0-335-20378-7 (hardcover). – ISBN 0-335-20377-9 (pbk.)
1. Research methods. 2. Psychology – Statistical methods.
I. D'Oliveira, Manuela. II. Title. III. Series.
BF39.G725 1999
150′.7′27–dc21 99-13094
CIP

Typeset by Graphicraft Ltd, Hong Kong
Printed in Great Britain by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk

Contents

Preface to the second edition

There have been an enormous number of textbooks which have claimed to present statistics in a simple way. Despite this, many psychology students still find the whole statistical business something of a mystery.

How does this book differ from these other attempts?

We believe that virtually all books on statistics feel obliged to start with the mathematical principles underlying probability distributions, samples and populations, and statistical testing. But, however simply these are presented, in our view they obscure the basic reasons why psychologists use statistical tests.

So we had better come clean straight away. This book sets out to achieve one single aim. This is to enable students to select appropriate statistical tests to evaluate the significance of data obtained from psychological experiments. In other words, this book is concerned with *inferential statistics* as used in psychological experimental studies.

We have concentrated on this to the exclusion of much else. Topics like descriptive statistics and the basic principles of probability are well covered in other statistical texts. Moreover, we will be concentrating on psychological *experiments* rather than other types of psychological investigation. There is nothing here about the use of surveys, observational techniques or psychometric tests of intelligence and personality. All we have included is the battery of statistical tests which are usually introduced to psychology students as part of their undergraduate laboratory course. We hope that, by aiming at a single target, we will maximize our chances of scoring a bull's-eye.

While this is definitely a 'beginners' book, it takes students from the simplest non-parametric tests, like the Wilcoxon test, through to complex analysis of variance designs. The principle is the same throughout: always to give the rationale for using appropriate statistical analyses for particular experimental designs. It is certainly our expectation that anyone who has mastered the why and how of the statistical tests given in this book will be in a position to understand the basic principles of statistical tests as presented in advanced textbooks of psychological statistics.

Our belief is that, with the aid of this book, students will feel comfortable about the basis for selecting and applying all kinds of statistical tests. We hope that teachers will wish to use a book which frees students from much of the panic usually associated with statistics. That way they should be in a far more receptive state to learn.

The major change in the second edition is that the book is now organized into three parts. Part I contains a general introduction to the principles of research and design. Part II presents all the information required for non-parametric tests. It is not until Part III that multivariable designs are introduced to prepare students for analysis of variance.

We hope this reorganization will have several advantages. Both students and teachers will gain from the clear division between non-parametric and parametric tests. The new approach embraces the principle of introducing more complex designs on a 'need to know' basis, providing room for a more extended treatment of concepts students find difficult, like degrees of freedom and interactions.

The second innovation is to introduce students to the use of computerized statistics packages. Although many readers will have access to computer programs, we still provide step-by-step instructions for those who do not. One danger is that when students can simply press buttons on a computer they can lose sight of the purpose of statistical analysis. We have concentrated on ensuring that students understand the inputs and can interpret the outputs of programs.

This book does not attempt to provide complete instructions for logging into particular packages, all of which differ slightly. A brief account is given in Appendix 1 about one of the most commonly used computer packages and the appropriate terminology. The intention is that students will have no problem in adapting easily to whichever programs are available in a particular institution.

Note about subjects and participants

Traditionally psychologists used the word 'subjects' to describe the people taking part in experiments in order to distinguish them from 'objects'. More recently, it has been agreed by the British Psychological Society that a better term would be 'participants' in order to remind researchers that people are participating in their experiments. However, it is not common to use this term when describing experimental designs and statistics. For instance, 'within-participants' and 'between-participants' would sound quite odd, rather than the usual 'between-subjects' and 'within-subjects'.

The usage in this book refers to the *people* taking part in experiments and the differences between people as a source of variability. But, when the discussion from Chapter 2 onwards moves to types of experimental designs and the consequences of using same and different subjects, 'subjects' is used in its technical sense (as in the Decision Charts).

Study guide for students

The aim of this book is to explain the rationale for using statistical tests to evaluate the results of psychological experiments. The problem is that in psychology you have to carry out these experiments on *human beings*, often other students. Unlike most physical objects, human beings are unique, each interpreting and performing whatever task you set them in a slightly different way.

You will find that the data and observations obtained from the people doing a psychological experiment are often extremely varied and that many of the things which influence their behaviour may have nothing to do with the experiment. It is for this reason that you have to sort out whether experimental results are really significant. And, as you will see, this is just what statistical tests enable you to do.

You will probably be relieved to hear that the chapters which introduce the basic rationale for statistics and summarize all you need to know in order to select an appropriate statistical test are the shortest chapters in the book. These are aspects of using statistical tests that students often find rather puzzling, but we hope that these chapters will clear up all your worries.

Other chapters in the book present statistical tests, explaining the rationale for each one, taking you 'step by step' through any necessary calculations and giving precise instructions about how to use the statistical tables in Appendix 2. For more complex types of statistical analysis you will be introduced to the latest types of computer programs for carrying out numerical calculations.

One essential feature of this book is the *questions* which occur throughout the text. It is not enough to read the summaries presented in the *progress boxes*. The only way you can make sure that you understand the context of each section is to attempt the questions *before* looking up the answers at the back! It is equally important to work your way through the step-by-step instructions given for each statistical test. Otherwise, you will never gain the confidence which comes from fully understanding the rationale for a statistical test as you successfully complete the necessary arithmetical calculations.

Let us end by making some important, and, we hope, encouraging points. The first thing to grasp is that statistical tests are not magic formulae to be turned to, desperately wondering which on earth to choose. They simply follow as a natural result of the kind of experiment you have chosen to do. What makes most people give up all hope of mastering statistics is the thought that they will find themselves presented with a huge amount of numerical data without the foggiest idea of how to deal with them. But this is quite the wrong way to go about things. The important thing is to decide what experiment you want to carry out. You will find that such a decision immediately narrows your possible choice of statistical tests to only one or two, and that there are good reasons for selecting one or the other.

With statistical tests selection is all; the actual calculations are quite easy once you have understood the reasons for doing them. The aim has been to introduce the principles of using statistical tests *without referring to any mathematical concepts*. And, in order to do the calculations for the tests themselves, you will only need to know how to add, subtract, multiply, divide and square numbers. With modern pocket calculators and computer programs this really should be child's play.

Good luck – and if, in spite of everything, you do find yourself getting disheartened by statistics, turn back and reread this study guide.

Prologue

The way we propose to tackle the daunting task of introducing statistics ‘without tears’ is to concentrate on what statistical tests are *for*. Once you have grasped this we hope the rest will fall into place.

The first thing to consider is *why* psychologists carry out experiments. The simple answer is that they want to develop and test *theories* about human nature and experience.

Let us start by supposing that you as a psychologist have a theory about how children acquire reading skills. On the basis of this you have developed a new reading scheme which consists of a set of picture/sentence cards. Now you want to carry out some research to demonstrate whether your theory works or not. Filled with enthusiasm, you find a friendly school which allows you to give your reading scheme to a group of children and to measure their reading scores at the end of term. What might be the reaction of a sceptical teacher?

Sceptical teacher: How do I know the children’s scores were any better after the reading scheme than they were before?

You as a researcher: Well, I measured their reading scores *before* as well as *after* the reading scheme. Their scores after the scheme were higher, showing an improvement in reading.

Sceptic: Can you be sure that their scores wouldn’t have gone up anyway without the scheme? After all, the children were 3 months older by the time you tested them the second time.

Researcher: I compared the reading scores of the children who were given the reading scheme with another group of children who didn’t have the scheme. The children who were given the scheme improved more.

Sceptic: How do you know that the children you gave the reading scheme to weren’t better at reading anyway? Or perhaps they were worse at reading in the first place and so had more room for improvement. Or perhaps the first lot were all girls who tend to learn to read more quickly.

Researcher: I tried to match my two groups of children as carefully as possible for all relevant factors, e.g. sex, intelligence, initial reading skills. As it wouldn't have been possible to match the children on all possible characteristics, I otherwise allocated them randomly to the two groups. So any differences between the groups ought to be due to my reading scheme rather than to any other factor.

Sceptic: How can you be sure that the teacher who administered the reading scheme wasn't particularly enthusiastic about it and expected improved reading scores? The other group may have got the same old bored and discouraged teacher and that's why their scores didn't go up – nothing to do with your reading scheme.

Researcher: I was particularly careful that the same teacher was responsible for teaching both groups and that he/she was given something interesting to do even with the other group of children.

Sceptic: But if you only used one teacher at one school how do I know that your reading scheme would help children in other schools?

Researcher: What I really meant to say was that I carried out the research in several schools, all in different areas and with different kinds of children.

Sceptic: How did you manage to standardize conditions in all these different schools, or did you let things just happen?

Researcher: I made up some instructions telling teachers how to administer the reading tests, how many weeks to operate the scheme, how large the classes should be, and so on.

Sceptic: The more I hear you talking the more I wonder about all the variability introduced by using different children, different teachers, different schools. Individuals vary so much in their performance, even from day to day. How can you be certain that the improvement in reading scores that you attribute to your reading scheme is big enough to count as a real difference between the reading scheme group and the no reading scheme group? Perhaps the results of your experiment were all due to chance fluctuations in performance.

Researcher: Ah well, I went off and did a statistical test which told me that the difference in reading scores between the two groups was unlikely to have occurred simply by chance. It was a big enough and a consistent enough difference to count as a real difference between the two groups of children.

Sceptic: Now I come to think of it, I am not really interested in overall differences between children who were given the reading scheme and those who were not. What I want to know is whether it is only children who are fairly good at spelling who benefit from the scheme, or whether it is particularly helpful for backward spellers.

Researcher: Why didn't you say so before? I could have measured the children's initial spelling scores to see whether it was children with less good or better spelling scores who were most likely to show improvement after the reading scheme.

Sceptic: But, even if you found that better spelling scores were associated with improved reading scores, mightn't this be due to some quite different factor? Children who enjoy school lessons may be more likely both to be good spellers and to benefit from any new teaching scheme? So it wouldn't be spelling ability as such which was responsible for these children showing more improvement.

Researcher: It certainly is a problem to discover exactly what lies behind an association between spelling and reading scores. That is why it would be a good idea to have equal numbers of less good and good spellers, and children who enjoy and who hate school lessons in my experiment. That way I might be able to find out whether spelling abilities or attitudes to school have an effect on whether children benefit from the reading scheme.

In this Prologue we have raised some possible objections to experimental research. A lot of them stem from the fact that psychologists study people rather than physical objects. The trouble with people from a psychologist's point of view is that they differ from each other in so many unpredictable ways. Of course, from the point of view of the people themselves, they are pleased that there is so much variability in people's behaviour.

So what is the unfortunate researcher to do? Obviously it is not possible to look at all possible factors which might affect the way children learn to read. Sometimes it is appropriate for psychologists to carry out an exploratory investigation in which they can observe as much as they can about ongoing behaviour. This can be a useful stage in developing a theory about what might be the most important factors affecting a particular type of behaviour. However, at some point a researcher will want to test out a theory. In order to do this, the researcher will make a prediction about the kind of behaviour which would be expected to occur if the theory is true.

Part I Introduction

1 Introduction to experimental research

1.1 The experimental hypothesis

The aim of psychological research is to test psychological theories. Researchers make predictions which would arise from a theory and design an experiment to see whether the results support the prediction or not. This may be clearer from an example. Suppose the theory is that words are remembered as whole words rather than as groups of letters. The prediction would be that all words are equally easy to remember, long words like *catapult* being as easy as *cat*, rare words like *anteater* as easy as common words like *elephant*. The test would be to carry out an experiment to find out whether people actually find all these words equally easy to recall.

In experimental research a prediction is formulated as an experimental hypothesis. It is called a hypothesis because it is not known until after the experiment has been done whether the prediction is supported or not. So the prediction is only a hypothesis, i.e. something which is proposed and has to be tested.

An **experimental hypothesis** makes a *prediction* about the effects of one or more events on people's behaviour. For instance, a researcher may predict that being given a reading scheme will improve reading skills. The aim of an experiment is to test whether the experimental hypothesis is supported by the behaviour of the people who are taking part in the experiment.

1.2 The null hypothesis

One especially important point to note is that in order to test an experimental hypothesis it must be possible for the predicted effects to occur or *not* to occur. If it were always certain that the predicted behaviour would occur there would be no point in doing an experiment. In our example, it must be possible for the results of the reading scheme either to *support* the experimental hypothesis (i.e. that children given the reading scheme will obtain higher reading scores) or *not* to support the hypothesis (i.e. that there will be no differences in reading scores between the children whether they are given the reading scheme or not).

This is a basic rule of experimental research. An experimental hypothesis is always tested against a **null hypothesis**, which states that an experimenter will *not* find the experimental results he/she expects. According to the null hypothesis, any results found in an experiment are due to random fluctuations in people's performance rather than the predicted effects the experimenter is interested in.

The next question is how to set about showing whether there is indeed a predicted relationship between two events (i.e. whether the reading scheme versus no reading scheme has any effect on children's reading scores). In other words, how can we rule out the null hypothesis?

1.3 Independent and dependent variables

An experimental hypothesis predicts a relationship between two events. One would be the event of being exposed to a reading scheme and the other the event of measuring children's reading scores. In research, these events are usually called **variables** because the events can vary. This applies to the variable which the researcher manipulates, for example, presenting a reading scheme or not presenting a reading scheme. It also applies to the scores which the subjects produce. Obviously children's reading scores vary over a wide range – in fact it would be strange if reading scores were all the same.

We should now like to introduce the terminology of independent variables and dependent variables. The variable which the researcher manipulates is known as the **independent variable**. This is because the experimental conditions to test this variable are set up *independently* before the experiment even begins. The second variable, representing children's scores, is known as the **dependent variable**. This is because the reading scores are *dependent* on the way in which the experimenter manipulates the independent variable of the reading scheme.

The most common method of testing variables in a psychological experiment is for the experimenter to manipulate an *independent variable* to see whether it has an effect on the *dependent variable*. In the reading scheme experiment the researcher manipulates one variable (reading scheme or no reading scheme) to see what effect this will have on another variable (children's reading scores). If children who are given the reading scheme show more improvement in reading scores than children given no reading scheme, then the experimenter could claim that the relationship between the two variables was in the direction predicted by the experimental hypothesis, namely that the reading scheme would improve children's reading skills.

The basic experimental design is to allocate people to **experimental conditions** representing different conditions for the independent variable (reading scheme versus no reading scheme). In this way a comparison can be made between the experimental conditions to see what effect manipulating the reading scheme variable has on the dependent variable of reading scores.

? **Question 1** Suppose a psychologist designs an experiment to test an experimental hypothesis that people will take less time to read a text with illustrations than the same text without illustrations.

(a) What is the independent variable?

(b) What is the dependent variable?

(c) What is the experimental hypothesis?

(d) What is the null hypothesis?

(Answers on p. 176)

1.4 Controlling irrelevant variables

It is important to grasp that the point of designing an experiment is to demonstrate that the results are due to the independent variable selected by the researcher *and to nothing else*. Any other variable that may be affecting subjects' behaviour is considered undesirable. Such variables are called **irrelevant variables** because they are irrelevant to the main purpose of an experiment. For instance, suppose a researcher is interested in investigating the effect of a reading scheme. From this point of view, the motivation of teachers is an irrelevant variable which might be affecting children's performance in unpredictable ways.

The fact that teachers who present the reading scheme may be more motivated and enthusiastic would still result in the predicted effect that the children with the reading scheme would show improved performance. But the result might be due to the irrelevant variable that the children had a more interesting teacher. The experimenter intended to test the effects of the reading scheme, not the irrelevant variable of enthusiastic versus boring teachers.

It is possible for an experimenter to take action about an irrelevant variable as long as he/she realizes that it could be a possible alternative explanation of the children's behaviour. For instance, the researcher could make sure that equally good teachers were selected for both the reading scheme and the no reading scheme groups and that the teacher gave an interesting class to the no reading scheme group. In this way, the potentially irrelevant variable of enthusiastic versus boring teachers would have been systematically eliminated.

Good experimenters attempt to develop **standardized experimental procedures** by specifying the instructions to be given to subjects, the location and the timing of the experiment. The written and spoken materials to be used in the reading scheme and the measure of reading scores will all have been decided in advance. The objective is to eliminate as many irrelevant variables as possible.

Another type of irrelevant variable is caused by individual variability which affects people's performance. Some of the children may have parents who don't care about school, others may believe that their reading performance is already excellent, others may be daydreaming or looking out of the window. It would be very difficult for the experimenter to make allowance for all these kinds of irrelevant variables. It might be possible to take into account parents' attitudes to school as an extra independent variable. But things would get out of hand if any systematic attempt was made to control irrelevant variables like daydreaming. The researcher has to live with this kind of individual variability.

The truth of the matter is that it is impossible to control all irrelevant variables which may influence people's behaviour. It is for this reason that researchers employ statistical tests to test the effects of predicted independent variables against all other potential irrelevant variables.

Progress box one

Relations between variables

- An *experimental hypothesis* predicts a relationship between variables.
- *Variables* are any characteristics which vary in an experimental situation.
- The experimenter manipulates *independent variables* and predicts their effects on *dependent variables*.
- The experimenter sets up *experimental conditions* in which an independent variable is varied (e.g. reading scheme versus no reading scheme). The scores represent *differences* in the dependent variable (e.g. reading scores) between the people allocated to the experimental conditions.

- The *null hypothesis* states that the scores resulting from an experiment are not due to the effects of an independent variable as predicted by the experimental hypothesis. Instead people's scores are influenced by the effects of *irrelevant variables*.
- Experimenters attempt to *standardize* experimental procedures in order to eliminate irrelevant variables.
- It is impossible to *eliminate* all individual variability.