PSYCHOLOGY Live **SERIES**

USER'S GUIDE

Inferential Statistics

This programme is the third in a series of four on **Research Methods and Statistics**.

The guide is written for teachers and will be particularly helpful for those new to psychology teaching.

It is designed to be read before viewing and an overview of the content and structure of the video/DVD is given to assist with planning and lesson preparation.

Running time: 33 minutes

We hope you find the video/DVD a useful teaching tool.

Aim of the Series

Many students of Social Sciences approach ‘anything to do with Maths’ with trepidation. The aim of this series is to try to nurse them gently through the process of preparation, analysis and testing in experiments while teaching them all the basic terms and techniques required by the examination boards.

Content

The first programme now known as ‘the spaghetti bolognaise one’ takes a relatively light-hearted introductory look at design and methods. The second one uses a musical theme to explain how to use descriptive statistics to illustrate quantitative data. It is expected that students will have a basic knowledge of these topics before watching this video/DVD.

This third programme looks at probability and testing using a magic and circus skills theme. It is an introduction to inferential statistics and aims to show students that statistics is important but can be fun and within their capabilities.

Amy is back as presenter/narrator and she, with the help of Matt the Magician, guides us through the need to make probability statements and set significance levels. We are then joined by a team of students and we use data based on their circus skills to discuss choice of test. We go step-by-step through the signed ranks test together and then describe how other tests of difference look at data.

There is live (and lively!) footage throughout with supporting graphics, animations, discussion points and questions for the viewers. Test of correlation will be considered in the fourth video in the series on qualitative research methods.

Writing materials may be useful for the Sign Test as well as a statistics textbook so that students can practise working with critical values tables. We have reproduced the more detailed charts here to prevent students copying off the screen rather than concentrating on what is being said!

Running times (where relevant) given in pink.

00.00 Introduction

The opening sequence of circus skills footage and narration recaps where students will hopefully now be up to - that is, understanding how to develop a hypothesis, design an experiment, consider ethics, prepare materials, select a sample, gather results and present some descriptive data. “Now, on to the statistical calculation and interpretation.”

01.49 Probability

Introduces an analogy between statistics and magic - both full of uncertainties. The importance of being as certain as we can that we have an effect.

Discussion Points where the video could be paused

Answers are given and the Chart of Probabilities gradually builds up.
What can we be 100% certain about?
What is the probability a coin will fall down as heads or tails?
What is the chance or probability of throwing a six on a die?
What has 0% chance of happening?
Why do we need to know how probable an event is?

06.43 Setting significance levels

Chart of Ratio/Percentage statements then introducing expression in decimals.
Type 1/Type 11 errors.
Lead into need to test for differences.

12.20 Tests of Difference

Revision of Levels of Measurement
Looking at power
Assumptions of a *t*-test
Designs: non-parametric/parametric
Choosing an appropriate test
Chart of Tests

19.48 The Sign Test

Example data from students who have taken a circus skills course. Can the participants juggle more balls or not more balls after the course?

4 Quick Questions are posed

Is our hypothesis going to be one or two tailed?
Is the design repeated or independent measures?
What Level of Measurement did we collect?
Based on that information, which would be the most appropriate test? (decision chart shown on screen)

The calculations are worked through on screen, the test statistic looked up on a table of critical values and a probability statement made.

24.25 Other tests of difference are considered with example data

They are not fully calculated but the importance of the careful choice of test and understanding how the test examines the data are stressed. The limitations of tests are not discussed in detail.
Simple chi-squared, Wilcoxon, Mann Whitney and the *t*-tests.

Tables used in the programme (reproduced below)

Tables of Probability (3 versions)
Test of Differences Chart (builds up gradually in five stages)
2 x 2 chi-squared Table of Observations

Table of Probabilities (1)

Birth	Certain	p = 100%
Heads/tails	50:50	p = 50%
Throwing a six	1 in 6	p = 16.66%
Living on the sun	Impossible	p = 0%

Table of Probabilities (2)

Birth	Certain	p = 100%
Heads/tails	50:50	p = 50%
Throwing a six	1 in 6	p = 16.66%
Minimum level	1 in 20	p = 5%
Living on the sun	Impossible	p = 0%

Expressions of Probability

P as ratio	**P as percent**	**P as decimal**	
1:10	10%	0.1	Very lenient
1:20	5%	0.05	Usual in social science
1:100	1%	0.01	More stringent
1:1000	0.1%	0.001	Very stringent

power

Level of Measurement	Repeated Measures design	Independent Measures design
Non-parametric nominal	Sign Test **S**	Chi-squared χ^2
ordinal	Wilcoxon **T**	Mann-Whitney **U**
Parametric Interval/ratio	Related t-test	Unrelated t-test

Degrees of Freedom (df)

In this programme we use degrees of freedom for the chi-squared test but say that we will explain them in the User's Guide. Students should know that they will appear in other tests (parametric ones) in a slightly different form but they still have much the same meaning. In the chi-squared 2 x 2 test we calculate the df by subtracting one from the number of rows (2-1) and one from the number of columns (2-1). We then look the answer up on the critical values table for chi-squared which is when we need to use the degrees of freedom.

The reason for them is quite difficult to explain to students and many textbooks do not include an explanation.

Basically they are concerned with how much our data is able to vary. If you have two possible choices you have one degree of freedom to choose one or the other; once you have chosen one, you have no freedom of choice - no degree of freedom.

Applying that to a chi-squared test, if we collected 32 observations we would expect some to fall into one of each of four categories if they had an equal chance of doing so. It is unlikely that exactly 8 would fall into each cell and equally unlikely that they would all fall into one.

When we have the total for one we have a better idea because there are less scores left to fit into the other three cells. But, it can still vary in any way - the next cell may contain all the remaining scores or none or anything in between. Once we know the totals for two cells the degrees of freedom reduce again as we only have a certain number of scores left. We then only have one degree of freedom left. Once we have the third total, there is no degree of freedom as all remaining observations have to go into the final empty cell.

The chi-squared test observations we used in our programme can be seen below. Obviously with a variable such as male and female there is less freedom to vary and this provides an opportunity for discussing df further if there are students who 'need to know'.

Degrees of Freedom = no of rows minus1 x no of columns minus 1

(2-1) x (2-1) = 1df

Results of a study observing whether males and females would take a card on request from a magician in the street

χ^2	Males	Females	Total
Took card	10	19	**29**
Did not take card	13	8	**21**
Total	**23**	**27**	**50**

Keeping up to date

It can be difficult to keep abreast of new ideas and developments in psychology when there is so much change in many areas of research methods and statistics. Much of this is at the 'academic researcher level' but filters down to the classroom teacher eventually through papers presented at conferences and published in journals or simplified for us even later in textbooks and magazines. When this happens the exam boards tend to gradually adopt the new ideas. Does this sound familiar? Paradigm shifts?

Although we shouldn't confuse a global paradigm shift with a theoretical approach, the ideas and change of thought in statistics gradually filtering down to us in recent decades reminds one of the Popperian version of the inductive method.

Kuhn (1962) considered psychology to be in a state of pre-science, others consider we are in a state of normal science having gone through various paradigm shifts. A more recent view is that we have several paradigms going on concurrently.

Let us consider some of the debates going on at the moment in research methods and statistics.

Yates Correction

We were all taught to use Yates Correction to adjust for over-simplification in a simple 2 x 2 chi-squared test - for example, classifying human behaviour into just two categories such as male and female, may exaggerate differences or associations especially when the expected frequencies are small. However, there seems to be a widely held opinion that Yates correction is no longer essential. Some popular textbooks ignore it, some mention that it is rarely used these days and others who calculate it. Howitt and Cramer (2003) call it a 'slightly outmoded statistical procedure' and conclude that 'Really it is a matter of personal choice as far as convention goes'.

Expected frequency cell numbers in chi-squared

This is another issue that has changed since many of us studied for our psychology degrees. Remember 'the result of a chi-squared test is not reliable if 20% of the expected frequencies are less than 5'? Well, researchers now think that the test can tolerate expected frequencies as low as one or two in one or two cells. Coolican (1994) gives more detail but his advice is to use large samples (observed frequencies of at least 50) so that the expected frequencies are more likely to be above five and the risk of a Type 1 error reduced.

Setting significance levels

It has always been the practice to set the significance level before you conduct a study, usually at a 5% level. However there is a line of thought that asks why shouldn't we report that we not only reached the 5% level of significance but also the 1% level. This approach is accepted by exam boards and most research papers seem to accept it also.

Abuse of parametric tests

Hayes (2000) highlights the controversy over fulfilling the criteria for using parametric tests. We take for granted the idea of the sample having been drawn from a normal distribution and with equal variances rather than investigating it. This is not always easy to do with the small samples that students collect and the reduction in pieces of coursework required (remember the days of 10 pieces submitted?). This may be a controversy which will hit all textbooks soon and eventually change requirements from exam boards. We will see!

Plastic interval scales

This issue has been highly controversial for much longer and all textbooks cover it. This is the question of what is equal-interval data? Can psychometric scales be treated as interval data and therefore use parametric tests?

Many people would argue, as we do in this video/DVD, that intervals on a scale such as IQ, even though they have been standardised, cannot have the exact intervals of precise measurements such as minutes or centimetres. Fife-Schaw (1995) suggests that the problem can be resolved by analysing the data twice on this type of data using an ordinal test as well as a test suitable for equal-interval data. If the results are very different, accept the ordinal one for safety. This is a difficult one for students and hopefully they take the advice of their teachers before performing inappropriate tests. They should be treating all attitude, opinions or ratings as ordinal and be very cautious when treating 'words spelt correctly' type of scores as interval. They can ask the question about memory (and allocating scores for attractiveness etc) - is it equally difficult to improve your score from remembering 3 to 6 words as it is to remember 7 to 10 words? If the answer is 'No' then stick with ordinal tests.

Null Hypothesis Significance Testing (NHST)

You will be aware that we are still significance testing the null hypothesis in this programme. You may not be aware that this has been a raging debate amongst academics globally for decades although we did broach this in our previous programme's User Guide and suggest further reading. It is right that statistical approaches should be questioned and debated by researchers but recent articles have addressed the issue of how null hypotheses are tested for A Level. It is therefore a good time, as change comes ever nearer, to consider the issues.

Brief Overview of the Debate

Criticisms

Significance tests tell us the probability of getting a result as extreme as we have, given the null hypothesis but really **we want to know the opposite** – if the null hypothesis could be true, given the result we have just got.

The null hypothesis is always false – the means of the populations from which we drew our sample will never be exactly equal or the correlation exactly zero.

There is too little margin for error. eg if p = 0.85 we may reject the null hypothesis; if it was 0.84 we may accept it (known as **the true / false dichotomy**). Thus we inappropriately stress making a decision rather than inference.

It does not tell us how big the difference or how strongly related – we need to measure the **size of the effect**.

Studies that do not achieve a statistically significant result are less likely to be published and this leads to **over-estimation of an effect** if it is only those that have shown an effect that are available.

It **emphasises the power of large sample studies** thus losing the valuable contribution of small sample studies. Several studies together may provide more evidence about a particular phenomenon but individually have failed to reach statistical significance.

It is a **universally misunderstood concept** – while this is not an argument against significance tests, it is an argument against their continued use.

Alternatives to NHST

Many critics of significance testing have suggested alternatives, some common ones that we have found on websites being:

Report **parameter estimates with confidence intervals** in the form of standardized effect sizes where appropriate.

Enhance the quality of data gathered.

Use more qualitative data with largely graphical techniques to look more closely at the data.

Replicate results using analysis techniques which test for internal reliability.

Use meta-analysis to pool the results of multiple studies to give an overall estimate of effect size and a more useful understanding of relationships among different variables.

Improve the clarity and understanding of the language used. Get rid of the ambiguity and different forms of expression eg ‘significant’ widely used and abused.

Summary

There is no globally accepted conclusion to the debate as yet. We visited numerous university and college websites to access the current usage of NHST at undergraduate level and found that more than half who had posted any recent student papers still describe the traditional method.

So, it may be best not to burn our boats and confuse our students at this stage when they may go on to study at an institution that still uses NHST. While boards still examine the concept of hypotheses there is no reason to change before you have got to grips with it yourself.

Statisticians have been arguing about this since at least 1925, so there is no rush.

Instead you could:

- ✓ follow the debate on the web, in journals, papers, etc;
- ✓ when you are confident slowly introduce it to your students as a discussion point;
- ✓ await the time when all universities have abandoned NHST;
- ✓ await the time when there is clear guidance about what replaces it;
- ✓ await the time when the A Level textbooks have included it;
- ✓ and await direction from the exam boards.

Bear in mind the story of a country which decided to change the side of the road that cars were driven on. To avoid total chaos all cars were told to change sides one day and all buses and trucks a day later!

References

Coolican, H. (1994) *Research Methods and Statistics in Psychology* 2nd Ed. Hodder and Stoughton

Fife-Schaw, C. (1995) cited in Hayes, N. (2000)

Hayes, N. (2000) *Doing Psychological Research – gathering and analysing data* Open University Press

Howitt, D. and Cramer, D. (2003) *An Introduction to Statistics in Psychology* Prentice Hall

Kuhn, T. (1962) *The Structure of Scientific Revolutions* University of Chicago

Searle, A. (1999) *Introducing Research and Data in Psychology* Routledge

Teaching books, resources and materials

Teachers often struggle with the Methods and Statistics part of the course or, at least, with the teaching of it and with finding stimulating resources for the students.

There are now a large number of textbooks specialising in this topic as well as excellent chapters in general textbooks.

We give a few personal favourites below:

Clegg, Frances (1982) *Simple Statistics* Cambridge
An oldie, but has helped generations overcome their fears and may well be in its umpteenth edition.

Coolican, Hugh (2004) *Introduction to Research Methods and Statistics in Psychology* 4th Edition Hodder Arnold
The new edition of a modern favourite which covers all syllabus content.

Howitt, Dennis and Cramer, Duncan (2000) *First Steps in Research and Statistics* Routledge
Clear explanations and graphics. Very beginner-friendly.

Two other very useful books for research studies are:

Flanagan Cara (1998) *Practicals for Psychology* Routledge
Ideal for teachers to find ideas for practicals. Students can select from 20 practical reports and learn how to design, conduct and write up own report. Full of useful teaching points and advice for students including ethical issues, questionnaires and examiners' comments.

Banyard, Philip and Grayson, Andrew (1996) *Introducing Psychological Research* Macmillan
This is a marvellous handbook for teachers (may have run into further editions by now!). It contains over 60 detailed summaries of well-known research studies covering most syllabus topic areas. Questions for students with suggested answers.

Recommended websites

www.uniview.co.uk
a large collection of psychology videos, DVDs, posters, X-psyting extras, etc

www.theATP.org
the Association for the Teaching of Psychology home website

www.psyonline.edgehill.ac.uk
an excellent site designed for AQA (A) specifically but useful for all students – with a staff chat room!

www.s-cool.co.uk
revision guide covering most topics for psychology students

www.bps.org.uk/publications/rd.cfm
access to the free BPS Research Digest which will be e-mailed to you fortnightly; articles specially written with A Level students in mind

www.holah.karoo.net
information, fun activities, links – excellent for staff and students alike

The Association for the Teaching of Psychology

The ATP has highly experienced teachers and examiners ready to give advice and assistance, especially for new teachers of this topic. They can recommend textbooks and resources that will get you started.

ATP Helpline: Dorothy Coombs
work: dorothy@pursglove.ac.uk 01287 280800
home: dorothycoombs@24whinchat.freeserve.co.uk 01287 636502

New teachers of this topic are well advised to get in touch with ATP:

Association of Teachers of Psychology
c/o The British Psychological Society
St Andrew's House
48 Princess Road East
Leicester
LE1 7DR

Annual Conference - The ATP holds an excellent conference for members each July. As well as lectures and workshops, there is an opportunity to meet the examiners and to browse all the latest books and resources.

Materials, Videos, DVDs

Other titles available in the Psychology Live Series include:

- ✓ The Study of Attention
- ✓ The Study of Memory
- ✓ Perception: the theories
- ✓ Classical and Operant Conditioning
- ✓ Further Approaches to Learning
- ✓ Cognitive Development
- ✓ Language Development
- ✓ Evolution by Natural Selection
- ✓ Reductionism
- ✓ Introduction to Designing Experiments
- ✓ Organising Quantitative Data

Uniview has a large collection of resources for most areas of psychology, biology and sociology. See the catalogue online or contact us to request a copy.

Uniview Worldwide Ltd
PO Box 20
Hoylake
Wirral
CH48 7HY
Tel: 0151 625 3453
Fax: 0151 625 3707
www.uniview.co.uk
sales@uniview.co.uk

Uniview
WORLDWIDE