Statistics and Research Methods for Psychology

Dolores Frías-Navarro

Universidad de Valencia

M.Dolores.Frias@uv.es

The completion of an introductory course in research methods is a critical step for undergraduate students who will one day need to conduct their own original research. These courses are equally important for students who are not planning to conduct research in the future, because graduates still need to make informed decisions regarding research findings as part of their professional development. The aim of the course is to develop your ability to understand the published research literature, to design and plan research questions with a clear idea of how to test the questions of interest, and to become critical consumers of any sort of statistical information. Your introduction to the computer package SPSS is designed with the goal of making you informed users of the technology.

This course examines basic descriptive and inferential statistics including hypothesis testing for both non-experimental and experimental techniques applicable to the behavioral, social, and medical sciences and to education.

This course offers an in-depth review of some major themes of quantitative research: From research questions to data analyses; experimental and observational studies; the measurement of variables and looking at sample data; the general linear model for the analysis of data; estimation and null hypothesis significance testing; the estimation of effect size; aspects of validity. The social sciences are undergoing fundamental changes with regard to research methodology. These changes include a greatly increased emphasis on the reporting and interpretation of effect sizes and confidence intervals (Statistical Reform). The importance of defining the precise research question of interest and the use of confidence intervals and effect sizes for reporting and interpreting the results will be a theme permeating the course.

Statistics is more than understanding conceptually the types of questions that can be addressed with different methods and interpreting the results of analyses. In order to effectively answer research questions, the use of computer programs is necessary.

Although some analyses can easily be performed by hand when a data set is small, more complex models literally require the use of computer programs. In addition to the methods and techniques discussed in the lecture component, students will be introduced to the statistical program SPSS in the laboratory component so that the methods discussed in the lecture component can be implemented with a computer program.

SPSS is a general statistics program that performs all basic and many advanced analyses. SPSS is extremely easy to use due to the point-and-click nature of the program. The ridged structure imposed by the point-and-click design of the program, however, limits its usefulness for nonstandard and advanced analyses. Nevertheless, SPSS is the most popular statistics program within many domains in the behavioral, educational, and social sciences.

The objectives of the course include reporting on scientific research in writing and oral presentation of its results

Topics: advanced statistical methods including effect size, confidence interval, ANOVA, ANCOVA, randomized block designs, simple repeated measures, factorial designs, and nested designs, multiple regression, and multivariate analysis of variance. Students will become proficient in using a statistical software package to manipulate datasets and perform statistical analyses.

Course goals

1. The aim of this course is to prepare students involved in *research designs and statistical methods* for the social and behavioural sciences.

2. Understand and apply basic research methods in Psychology, including research design, data analysis and interpretation.

3. The ability to understand statistical techniques to analyze experimental data so as to reach objective conclusions based on the obtained data.

4. The ability to understand statistical terms and research reports as found in Psychology.

5. A basic introduction to using software for data summarization and analysis.

6. To equip students with the skills and knowledge necessary to carry out and evaluate psychological research.

7. Understand the role of causality in research design.

Learning Outcomes

Be able to discuss and evaluate critically in research reports, oral presentations, group discussions, and the degree examination the following topics relating to the course syllabus:

1. Have knowledge of the general principles of psychological research and the commonest elementary designs.

2. Be aware of the kinds of approach that are appropriate for different research questions.

3. Be aware of the pitfalls associated with the use of particular research strategies and experimental designs.

4. Knowledge of more sophisticated research strategies and designs.

5. Have a knowledge of the following statistical methods for use in the practical and later in dissertations:

   a. Scientific method. Evidence Based Practice. Research designs principles. Independent and dependent variables. Confounded variable.

   b. Validity. *Statistical Conclusion Validity* (did the treatment covary with the outcome?). *Internal Validity* (did the treatment affect the outcome?). *Construct Validity* (what labels or constructs best represent what we did?). *External Validity* (to what does the effect generalize?). Threats to validity. Methods to determine if data are appropriate for analysis.

   c. Research methods. Randomized experiments Experimental designs. Quasi-experimental designs. Non-experimental design. Observational Studies. Meta-Analysis. Distinguishing Causal from noncausal associations. Selection bias modelling.

   d. Statistical Hypothesis Testing: null hypothesis H_0, alternate hypothesis H_1. Basic concepts of hypothesis testing to include the null hypothesis, statistical

significance and errors in decisions concerning the null hypothesis (*Type I error*, reject the null hypothesis even though it is true, and Type II error (beta error), fail to reject the null hypothesis even though it is false). Statistical summary measures: mean, variance, standard deviation, skewness. The sample size estimation. Value *p* of probability. Significance criterion Alpha. Statistical power. Power of a statistical test (the probability that the test will correctly reject a false null hypothesis, and thus avoid a Type II error) and the Beta Error probability. A hypothesis test tells us the probability of our result (or a more extreme result) occurring, if the null hypothesis is true. If the probability is lower than a pre-specified value (alpha, usually 0.05), it is rejected. The ability to reject the null hypothesis depends upon: 1) **Alpha** (α); usually set to be 0.05, although this is somewhat arbitrary. This is the probability of a Type I error, that is the probability of rejecting the null hypothesis given that that the null hypothesis is true. 2) **Sample size**; a larger sample size leads to more accurate parameter estimates, which leads to a greater ability to find what we were looking for and 3) **Effect Size**; the size of the effect in the population. The bigger it is, the easier it will be to find.

		State of the World	
		H_0 True	H_0 False
Research Findings	Fail to reject the null hypothesis H_0	✓	Type II Error ($p = \beta$)
	Reject the null hypothesis H_0	Type I Error ($p = \alpha$)	✓

e. Measure the size of the treatment effect as a supplement to hypothesis testing. Standardized measures of effect size. Cohen's *d*: mean difference / standard deviation. Parameter non-centrality. A relation among sample size, effect size and non-centrality parameters. Reporting effect size in quantitative research. Meta-analysis: what is meta-analysis, when and why we use meta-analysis, benefits and pitfalls of using meta-analysis, defining a population of

studies and structuring a database, and an introduction to analysis and interpretation. Examples will be drawn from the social sciences

a. Sample size. To ensure that the sample size is big enough, you will need to conduct a power analysis calculation. For any power calculation, you will need to know: What type of test you plan to use, the alpha value or significance level you are using, the expected effect size and the sample size you are planning to use. When these values are entered, a power value between 0 and 1 will be generated. If the power is less than 0.8, you will need to increase your sample size. It is generally accepted that power should be .8 or greater; that is, you should have an 80% or greater chance of finding a statistically significant difference when there is one.

b. One-way between-subjects analysis of variance (ANOVA). Analysis of variance (ANOVA) is a statistical tool used in the comparison of means of a random variable in populations that differ in one or more characteristics (factors), e.g. treatment, age, sex, subject, etc. First, we cover one-way ANOVA, where only one factor is of concern. Depending on the type of the factor, the conclusions pertain to just those factor levels included in the study (fixed factor model), or the conclusions extend to a population of factor levels of which the levels in the study are a random sample (random effects model).

c. One-way within-subject analysis of variance (ANOVA).

d. Factorial between-subjects analysis of variance (ANOVA).In two-way and multi-way ANOVA (populations differ in more than one characteristic), the effects of factors are studied simultaneously to obtain information about the main effects of each of the factors as well as about any special joint effects (factorial design).

e. Factorial within-subject analysis of variance (ANOVA).

f. Nested designs. In nested designs, where each level of a second factor (mostly a random factor) occurs in conjunction with only one level of the first factor, analysis of variance enables us to extract the variability induced by the nested factor from the effects of the main factor. For correct analysis of the

data in multi-way ANOVA, not only the linear model and the type of factor have to be considered but, also, the assumptions that must be satisfied.

g. Mixed design analysis of variance (ANOVA).

h. Analysis of Covariance (ANCOVA).

i. Other F-test: MANOVA, repeated measures designs univariate approach, repeated measures designs multivariate approach.

j. Follow-up tests, such as multiple comparisons. Planned comparisons.

k. The quality of reporting. Reporting guidelines. Improving of quality of reports.

l. Qualitative research methods for psychology.

6. Confidence intervals for effect sizes in Analysis of Variance. Procedures for constructing confidence intervals on contrasts on parameters of a number ANOVA models.

7. including multivariate analysis of variance(MANOVA) models

8. for the analysis of repeated measures data

1. Participants will extend their skill with the statistical computing package SPSS for Windows to the implementation of the techniques described in 5.

2. Be able to plan, conduct, analyse, and report on empirical studies conducted under the supervision of a member of staff.

3. Students acquire the knowledge and skill to evaluate research in applied settings as well as to design studies suitable to different problems and situations, to apply the designs and report the results so they are most useful to clients.

4. Obtain a brief overview of a number of methods that can be used in qualitative analysis of psychological data.

5. In this course we will focus on correct execution of data analysis and understanding the results of this analysis. We will provide insight into the conclusions and pay attention to expressing these conclusions in a correct and understandable way. The different methods will be extensively illustrated with examples from scientific studies in a variety of fields.

Basic Skills of Students

- To develop their capacity for critical thinking.

- To develop skills in presenting an argument and/or data, both orally and in writing.

- To inculcate knowledge about conducting experiments and surveys, and how to interpret their own and other people's data.

COURSE INFORMATION

Teaching strategies

- Active-learning experiences. Student centred instructional methods: problem based learning, discussion in class, group work.

- Lectures, SPSS for Windows practical, experimental practical.

- During the Course, reference will be made to a wide range of material, not all of which can be considered in detail during the lectures. Students are expected to read the textbook chapters and papers that lecturers recommend. Knowledge of such material will be assumed in the written examinations.

Homework

Homework problem will be assigned after each topic and their solutions reviewed in class. As Harris (2001) argues, "true understanding of any statistical technique resides at least as much in the fingertips (be they caressing a pencil or poised over a desk calculator or PC keyboard) as in the cortex" (p. 51).

Although no credit is given for the homework, you will find these problems most helpful in learning the course material and in preparing for examinations.

Laboratory exercises

There will be a series of lab exercises requiring the use of computers.

Exam

The exam consists of 40 items of multiple-choice over material covered in lectures, the text, the power-point and the class.

THE SCIENTIFIC MANUSCRIT: A GUIDE

When you have carried out an experiment or study it is imperative that you record and interpret your data in such a way that the importance of them is successfully communicated to others. It should be possible for a person reading your report to be able to replicate your study (i.e. carry it out exactly as you did) so your description of the experiment needs to be precise and accurate. This is very important in scientific methodology, since it ensures that errors and incorrect conclusions can be identified and rejected. It is essential that the level of the practical report is correctly pitched: the language used should be formal, but you should not assume that the reader has specialist knowledge. If in any doubt, read some journal articles – this is the style that you are trying to achieve. When you are preparing your report, you will need to organise your material and support your statements, where possible, with documented evidence.

All empirical articles follow a similar format, so it is important that you understand and follow this format from the beginning of your Psychology course. Such articles start with a title, then the author's name (or names if there is more than one of them) and affiliation for published papers, followed by an abstract, an introduction, then a method section, including (where appropriate) details of the experimental design, participants, materials and/or equipment and procedure, followed by sections on the results, discussion and conclusion, a reference list and possibly an appendix.

Most of the report should be written mainly in the past tense, as you are describing what has been done and why you have done it. You should try to convey this information as concisely and clearly as you are able. The language that you use should be impersonal ('this was done', rather than 'I did this').

Structure of an article

These are the basic structure guidelines that most journals have:

Title

This should identify the topic of the study in one phrase or sentence.

Abstract

An abstract is a single paragraph. This is a short summary in about 200-400 words of the whole experiment. It includes the area of investigation, why the research was

conducted, how it was conducted, and what the major results, the discussion and the conclusion. Although this is placed at the beginning of the report you will probably find it easier to write it after you have completed the rest of the report. References are typically not cited in the Abstract, since the reader expects a more full discussion in the body of the article.

Introduction

Every scientific report needs an introduction. The length of an introduction depends on the journal and the paper; however, the structure and content should be similar. In the introduction, the author must present the problem his or her research will address, why this problem is significant, and how it applies to the larger field of research. The author must clearly state his or her hypothesis, and quickly summarize the methods used to investigate that hypothesis. The author should address relevant studies by other researchers; however, a full history of the topic is not needed. This introduces the experiment, describes previous similar experiments and their conclusions, and builds up to the purpose of the experiment in terms of an aim. It progresses from the general background towards the specific aims of the experiment you are conducting. You should:

• State the problem to be investigated, in general terms, at the beginning of the introduction.

• Review the literature concerning the practical topic, citing authors by name(s) and year according to the American Psychological Association (APA) style. You will find details of the APA style at the end of these instructions.

• Give a brief overview of how the study will be conducted.

• State the aim of the experiment, including a statement about the expected outcome (the experimental hypothesis or hypotheses).

The introduction should contain all the background information a reader needs to understand the rest of the author's paper. This means that all important concepts should be explained and all important terms defined. The author needs to know who will be reading this paper, and make sure that all the concepts in the paper are accessible to them.

Method

The method should supply enough details so that the experiment could be replicated by another researcher. It should be written in the past tense, noting exactly what was done. The method is organised under a number of subsections:

a. Participants

You should note:

• Details of the participants – the number participating, by gender, their age range and educational level, if known (e.g. 25 first year university Psychology students, 125 males and 110 females, ages 17-32). Also any personal details of the participants which may be relevant to the experiment.

• How the participants were selected (randomly or at a class or by another criterion).

• Whether any of the participants were subsequently excluded and if so, why.

b. Materials

You should note in detail:

• The stimuli used (you may want to include a copy or diagram of these in an appendix).

• How the stimuli were randomised/presented.

• Hat equipment/computer was used.

c. Procedure

You should note:

• The exact instructions to the participants (if these are very long, you could include them in an Appendix).

• Details of what the experimenter did and in what order (e.g. how the equipment was set up).

• How responses were recorded and subsequently scored.

d. Design A description of the plan or design of the experiment.

You should:

• Identify the experimental design (e.g. a within-subjects experiment) and a suitable statistic (e.g. ANOVA).

• Name the variables and identify their type (the independent variable/s (IV's) and dependent variable/s (DV's) or the variables to be correlated.

• Note, if a control group was used, how this was done – what was the basis of the allocation to the experimental and the control group.

• Note that confounding of variables is an important issue to consider in the planning of the study since we need to make sure that the data collected is a true reflection of what is being studied, rather than a reflection of some other factor which has not been taken into account. If we do not control the variables this may produce biased results.

Results

This is the most important part – relaying what actually happened in the experiment. It is a summary of the data collected and includes the outcome of any statistical test used.

The Results section summarises the data collected and the statistical treatment of them. First, state the main results or findings and then report the data in enough detail to support your conclusions. Mention all relevant results, including those that run counter to your hypotheses. Make sure that when you state whether or not your results support your hypotheses. When reporting inferential statistics (e.g. t-tests, F tests, Chi-square), include the magnitude or value of the test, the degrees of freedom, the probability level and the direction of the effect. Be sure to report descriptive statistics, e.g. means or standard deviations.

Report the data clearly and economically. For example, Tables provide exact values and are more than adequate for showing main effects. Figures, on the other hand, best illustrate interactions and general comparisons. Avoid repeating the data in several places, e.g. by presenting both tables and figures and do not use tables for data that can be presented as a few sentences in the text. Label, title and number all tables and graphs (refer to all tables as tables and all graphs, pictures or drawings as figures). Always explain what to look for in tables and figures and use them to supplement the text. Any large tables or long statistical analyses can be included as appendices, suitably annotated and numbered for easy reference. Any material that appears in tables or figures should also be described in text.

Results (e.g. graphs, tables of means, standard deviations, statistical tests (ttests, correlations etc)) should be arranged in the order of the hypotheses, with brief objective comments, but not subjective comments – these should be reserved for the Discussion section. Make sure that you state whether or not your result supports your hypothesis, including the level of statistical significance, plus the critical and observed values, where it has been appropriate to calculate this. Please note that we do not 'prove' hypotheses, or say that they are correct or true. We say that the hypothesis was supported or not supported. Also, it is not appropriate to say that 'unfortunately' a test was non-significant: the language used should always be objective.

No interpretations or conclusions should be drawn. All interpretation and discussion of the results should be saved for the Discussion section.

Discussion

In this section, the author should restate the problem he or she was attempting to address, and summarize how the results have addressed it. This is the place for the consideration of your results and their theoretical or practical implications. The results should be described in relation to other experiments mentioned in the Introduction and in relation to your aims and hypothesis and relevant suggestions made about further studies.

You should:

• Discuss the significance of all the results and interpret their meaning (with reference to the numbered tables and graphs in the Results section or in the Appendices). Comment on any differences and similarities between your findings and those that you have referred to in the literature.

• Summarise what we now know and do not know as a result of this study and make constructive criticisms of the study and suggest what improvements in design could be made if it were to be done again. Consider whether the sample of participants used is an accurate reflection of the general population.

• Suggest what might be done next if the research were to be continued and indicate the form that future experiments might take.

Conclusion

This should be a brief summary of the findings of the study, noting whether they support the hypothesis. The conclusion is less comprehensive than the abstract because it only refers to the results. This should not be a continuation of the discussion and should not refer to the specific statistics of the data. It states the main import of what has been discovered.

References

You should identify the sources you have referred to and list any books and reference material you have used. You should list these alphabetically in the recommended format. The recommended format is the APA format:

- Notice that the name of a journal (but not the paper's title) and the name of a book are highlighted in some way (they can be underlined, boldened or italicised, or a combination as here).

-The year of publication appears after the name(s) of the author(s). If the same author/authors has/have published more than once in a year, use a, b, c etc. after the year (e.g. 1987a, 1987b etc.). The publisher's name is placed after the place of publication separated by a colon.

-If an item appears in an edited book, the editors' initials are located before the surname(s) and (Ed.) or (Eds.) appears after the surname(s). In addition, the name of the book rather than the title of the chapter is highlighted.

-Refer in the text to Frías (2001), Monterde (2003), Frías & Monterde (2006). If there are more than two authors for a single source, refer to all the authors (e.g. Frías, Monterde & Pascual (2006)) the first time and then to Frías et al. (1985) thereafter.

-For citing information found on World Wide Web, use the following format:

StatSoft, Inc. (1999). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB:

http://www.statsoft.com/textbook/stathome.html.

Acknowledgments

An acknowledgements section is not usually required; however, most papers include a paragraph of acknowledgements and thanks for help received on the research or the

paper. In journals where the reviewer's names are revealed, it is considered polite for the author to acknowledge the help of the reviewers. It is customary to acknowledge any special or unusual sources of help you have had in executing or funding your study. You might acknowledge:

• Anybody who has provided you with access to Schools, Clinics etc..

• Your supervisor(s).

• The source of finance if your research has been sponsored.

• A computer, statistical or any other expert who has provided you with specialist help.

Appendices

These are where you may include items such as raw data, actual instructions, further details of stimulus materials, and statistical computations. Other items which might be included are fuller tables of results, analyses, diagrams, computer programs etc. Make sure that appendices are suitably annotated and numbered. Material should not appear in appendices unless it is referred to in the main text of the report.

Readings

American Psychological Association. (1952). Publication Manual of the American Psychological Association (1st ed.). Washington, DC: American Psychological Association.

American Psychological Association (2001). *Publication Manual of the American Psychological Association* (5th ed.). Washington, DC: American Psychological Association.

American Psychological Association (2010). *Publication Manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Bort, HMI; Llobel, JP; Navarro, MDF (2006). Interpretation mistakes in statistical methods: their importance and some recommendations. *Psicothema*, **18**, 848-856. (Monterde-Bort, H., Pascual-Llobell, J., & Frías-Navarro, D. (2006). Interpretation mistakes in statistical methods: Their importance and some recommendations.. Psicothema, 18, 848-856.)

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, **45**, 1304-1312.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, **112**, 155-159.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, **49**, 997-1003.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement*, **61**, 4, 532-574.

Daniel, L. G. (1998a). Statistical significance testing: A historical overview of misuse and misinterpretation with implication for the editorial policies of education journals. *Research in the Schools*, **5**, 2, 23-32.

Everitt, B. S. (2001). Statistics for psychologists: An intermediate course. Mahwah, NJ: Lawrence Erlbaum Associates

Fan, X., & Thompson, B. (2001). Confidence intervals about score reliability coefficients, please: An EPM guidelines editorial. *Educational and Psychological Measurement*, **61**, 4, 515-531.

Freedman, D.A. (1999). From association to causation: Some remarks on the history of statistics. *Statistical Science*, **14**, 243–258.

Harris, R. J. (2001). *A primer of multivariate statistics*. Mahwah, NJ: Erlbaum.

Hays, W. L. *Statistics* (5th ed.). New York, NY: Harcourt Brace College Publishers.

Hess, B., & Olejnik, S. (1997). Top ten reasons why most omnibus ANOVA F-test should be abandoned. *Journal of Vocational Education Research*, **22**, 4, 219-232.

Hinkle, D. E. & Oliver, J. D. (1983). How large should the sample be? A question with no simple answer? or ... *Educational and Psychological Measurement*, **43**, 4, 1051-1060.

Howell, D. C. (2007). Statistical Methods for Psychology (6th ed.). Pacific Grove, CA: Duxbury

Kaufman, A. S. (1998). (ed.). Introduction to the special issue on statistical significance testing. *Research in the Schools*, **5**, 2, 1.

Kelley, K., Maxwell, S. E., & Rausch, J. R. (2003). Obtaining power or obtaining precision: Delineating methods of sample size planning. Evaluation & the Health Professions, **26**, 258-287.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, **56**, 5, 746-759.

Kirk, R.E. (2001). Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement*, **61**, 213-218.

Knapp, T. R. (1998). Comments on the statistical significance testing articles. *Research in the Schools*, **5**, 2, 39-41.

Lazarowitz, R., Hertz-Lazarowitz, R., & Baird, J.H. (1994). Learning science in a cooperative setting: Academic achievement and affective outcomes. *Journal of Research in Science Teaching*, **31**, 1121-1131.

Lazarowitz, R., Hertz-Lazarowitz, R., & Baird, J.H. (1996). Reply: Learning and teaching science in the classroom: New settings and statistical comments. *Journal of Research in Science Teaching*, **33**, 681–683.

Levin, J. R. (1998). What if there were no more bickering about statistical significance tests? *Research in the Schools*, **5**, 2, 43-53.

Maxwell, S. E., Camp, C. J., & Avery, R. D. (1981). Measures of strength of association: A comparative examination. *Journal of Applied Psychology*, **66**, 525-534.

Maxwell, S.E., & Delaney, H.D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). The logic of experimental design. Mahwah, NJ: Erlbaum.

McLean, J. E. & Ernest, J. M. (1998). The role of statistical significance testing in educational research. *Research in the Schools*, **5**, 2, 15-22.

Monterde-i-Bort, H.; Frías-Navarro, D.; Pascual-Llobell, J. (2010). Uses and Abuses of Statistical Significance Tests and Other Statistical Resources: A Comparative Study. *European Journal of Psychology of Education*, 1-19.

Navarro, MDF; Llobell, JP; Perez, JFG (2000). Effect size and statistical significance. *Psicothema*, **12**, 236-240. (Frías-Navarro, D., Pascual Llobell, J. & García-Pérez, J.F. (2000).Effect size and statistical significance. *Psicothema*, *12*, 236-240)

Pascual-Llobell, J.; Pérez-García, J.F. & FrÍas-Navarro, D. (2000). Statistical significance and replicability of the data. *Psicothema*, *12*, 408-412.

Rosenthal, R. (1990). How are we doing in soft psychology? American Psychologist, **45**, 775-776.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper and L. V. Hedges (Eds.). *The handbook of research synthesis*. New York: Russell Sage Foundation.

Shaver, J.P. (1993). What statistical significance testing is, and what it is not. *Journal of Experimental Education*, **61**, 293–316.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education*, **61**, 4, 334-349.

Tatsuoka, M. (1993). Effect size. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 461–479). Hillsdale, NJ: Erlbaum.

Thompson, B. & Synder, P. A. (1997). Statistical significance testing practices in The Journal of Experimental Education. *Journal of Experimental Education*, **66**, 1, 75-79.

Thompson, B. (1998). Statistical significance and effect size reporting: Portrait of a possible future. *Research in the Schools*, **5**, 2, 33-38.

Thompson, B. (1999). Improving research clarity and usefulness with effect size indices as supplements to statistical significance tests. *Exceptional Children*, **65**, 3, 329.

Thompson, B. (1999). Why 'encouraging' effect size reporting is not working: The etiology of researcher resistance to changing practices. *The Journal of Psychology*, **133**, 2, 133-141.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes Educational Researcher, 31, 24-31.

Thompson, B. (Ed.). (1993). Statistical significance testing in contemporary practice: Som proposed alternatives with comments from Journal editors [Special issue]. *Journal of Experimental Education*, **61**, 4.

Tukey, J.W. (1969). Analyzing data: Sanctification or detective work? *American Psychologist*, 24, 83–91.

Vacha-Haase, T. & Nilsson, J. E. (1998, April). Statistical significance reporting: Current trends and uses in MECD. *Measurement and Evaluation in Counseling and Development*, **31**, 1, 46-57.

Wilkerson, M. & Olson, M.R. (1997). Misconceptions about sample size, statistical significance, and treatment effect. *The Journal of Psychology*, **131**, 6, 627-632.

Wilkinson, L., & The Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *The American Psychologist*, **54**, 594–604.

Research Methods in Psychology

Frías-Navarro, D.

University of Valencia (Spain)

As a psychologist, you will read much research and hear many reports. It is crucial that you can distinguish between experimental and nonexperimental research and can recognize the variables in the design. Experimental and nonexperimental research are distinguished by the degree of control that the researcher has over the subjects and conditions in the study. You must distinguish between manipulation and assignment versus observation. An experiment is an investigation in which at least one variable is manipulated by the researcher. You have to know what a variable is: A variable is an aspect of the testing condition that can change or take on different characteristics with changes in conditions. Learn this: You have to know the difference between an independent variable (IV) and dependent variable (DV). IV: The condition or factor manipulated or selected by the experimenter to determine its effect on the subject. DV: Dependent Variable: A measure taken from the subject that reflects the effects of the independent variable (IV). Confounded Variables: A confound occurs when you select an IV and another variable varies with it. What then is the cause of the effect on the DV? Is it the IV or the confound variable? In experiments, you manipulate the IV and measure its impact on the DV. When designing your study, cover as much of the range in your IV as possible. Sometimes we are interested in subject variables - variables that cannot be manipulated by the investigator. We have less confidence in our conclusions when we cannot manipulate the IV. Confounds are other variables associated with the IV that can also explain your DV. Eliminate confounds when possible.

Pooled standard deviation: This is the pooled estimate of standard deviation from both groups, based on the assumption that any difference between their SDs is only due to sampling variation. Pooled standard deviation is a way to find a better estimate of the true standard deviation given several different samples taken in different circumstances where the mean may vary between samples but the true standard deviation (precision) is assumed to remain the same.

Mean Difference: This is simply the difference between the two means.

p Value: Researchers may be interested in the probability of a hypothesis being true. However, the p value does not give this information. The p value gives the probability of observing the sample data or something more extreme, assuming the null hypothesis is true. It is the likelihood of observing the measured, or more extreme, effect observed in the sample data, assuming no effect actually exists. A P-value may be viewed as the probability of obtaining an estimate at least as far from a specified value (most often the null value, i.e., the value of no effect, the so called ''nil hypothesis'') as the estimate we have obtained, if the specified (null or test) value were (note the subjunctive) the true value. In other words, the p-value is a tail area probability based on the observed effect estimate; it is calculated as the probability of an effect estimate as large as or larger than the observed estimate (more extreme in the tails of the distribution), assuming the null hypothesis were true. Miettinen (1985) stated ''The p-value is a function of the data computed under the statistical model that underlies the analysis.''

Effect Size: This is the difference between the two means divided by the pooled estimate of standard deviation. It calibrates the difference between the experimental and control groups (i.e. the effect of the intervention) in terms of the standard deviation. Reporting effect size plays an integral role in educational and psychological research and is required by many journals. Certainly, the best-known measure of effect size is Cohen's d, which represents a substantial improvement over using p values.

Cohen's *d* is defined as the difference between two means divided by the pooled standard deviation for those means. Thus, in the case where both samples are the same size:

$$d = \frac{\text{mean}_1 - \text{mean}_2}{\sqrt{(\text{SD}_1^2 + \text{SD}_2^2)/2}}$$

Where mean i and SD i are the mean and standard deviation for group i, for i = 1, 2.

Different people offer different advice regarding how to interpret the resultant effect size, but the most accepted opinion is that of Cohen (1992) where 0.2 is indicative of a small effect, 0.5 a medium and 0.8 a large effect size.

In the case where both samples don't are the same sizes:

$$\overline{ES} = \frac{\overline{X}_{G1} - \overline{X}_{G2}}{s_{pooled}} \, ,$$

$$s_{pooled} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

This course beginning with an exploration of the foundations of the scientific method. The students will learn the methods of data analysis and the general logic behind statistical inference as well as the use of alternative measures such as effect size.

The Judd and McClelland text presents a process of data analysis where the following simple equation is of prime importance:

DATA = MODEL + ERROR

The meanings of DATA, MODEL, and ERROR are fully discussed in your text. Briefly, DATA are scores you have that you want to statistically evaluate. MODEL is a method of describing the DATA in a more, or less, simple manner. Finally, ERROR is the amount of error you make in representing the DATA with your MODEL.

Effect sizes are quantitative indexes of the relations between variables found in research studies. They can provide a broadly understandable summary of research findings that can be used to compare different studies or summarize results across studies. The effect size is simply a measure of the degree to which the null hypothesis is false. For example, if one group has had an 'experimental' treatment and the other has not (the 'control'), then the Effect Size is a measure of the effectiveness of the treatment.

Thompson (2006) states that "Effect size is a statistic quantifying the extent to which sample statistics diverge from the null hypothesis" (p. 187).

There are many different possible effect sizes, including the difference between treatment and control group means divided by the standard deviation (Cohen's d), the correlation coefficient between the independent variable and the outcome, and the difference in proportions of individuals experiencing a particular outcome.

A standardized effect size is an effect size that describes the size of the effect but that does not depend on any particular measurement scale. The more popular measure of effect size is Cohen's d, which expresses the difference between means in terms of standard deviation units.

$$\text{Effect size} = \mathbf{d} = \frac{(\mu_1 - \mu_2)}{\sigma}$$

Where s = the estimated standard deviation of the population(s).

$$\text{Cohen's } d = \frac{[M_1 - M_2]}{S.D._{pooled}}$$

When the sample N is equal between groups, Cohen's d requires the computation of a pooled standard deviation (SD) taking the form of:

$$S.D._{pooled} = \frac{S.D._1 + S.D._2}{2}$$

When N is not equal between groups, however, a pooled SD weighted by sample size needs to be calculated to obtain Cohen's d using:

$$S.D._{pooled} = \sqrt{\frac{(N_1 - 1)S.D._1^2 + (N_2 - 1)S.D._2^2}{N_1 + N_2 - 2}}$$

The noncentrality parameter is:

$$d = \frac{\mu_1 - \mu_2}{\sigma}$$

$$\delta = d\sqrt{\frac{n}{2}}$$

Noncentrality parameter of the F distribution is the square root.

Rosenthal and Rubin (1979, 1982) have argued that even small effects can be important. Rosenthal has argued in several places that just because an effect size is small doesn't mean that it isn't important.

Jacob Cohen (1988) gives very rough guidelines about estimating effect sizes. Typically, these effect size magnitudes have been interpreted based on rules of thumb suggested by Jacob Cohen (1988), whereby an effect size of about 0.20 is considered "small"; about 0.50 is considered "medium"; and about 0.80 is considered "large". The Cohen guidelines are only broad generalizations, however, covering many types of interventions, target populations, and outcome measures. Nevertheless, it has been standard practice for researchers and policymakers to interpret effect size estimates using these guidelines. The effect sizes should instead be interpreted with respect to empirical benchmarks that are relevant to the intervention, target population, and outcome measure being considered. Thus, it is important to interpret a study's effect size estimate in the context of natural growth for its target population.

Effect Size	*d*	*Percent Overlap*
Small	.20	85
Medium	.50	67
Large	.80	53

To evaluate δ we need to go to tables of power:

POWER: POWER AS A FUNCTION OF δ AND SIGNIFICANCE LEVEL (α)				
	Two-tailed α			
δ	.10	.05	.02	.01
1.00	.26	.17	.09	.06
1.10	.29	.20	.11	.07
1.20	.33	.22	.13	.08
1.30	.37	.26	.15	.10
1.40	.40	.29	.18	.12
1.50	.44	.32	.20	.14
1.60	.48	.36	.23	.17
1.70	.52	.40	.27	.19
1.80	.56	.44	.30	.22
1.90	.60	.48	.34	.25
2.00	.64	.52	.37	28

2.10	.68	.56	.41	.32
2.20	.71	.60	.45	.35
2.30	.74	.63	.49	.39
2.40	.78	.67	.53	.43
2.50	.80	.71	.57	.47
2.60	.83	.74	.61	.51
2.70	.85	.77	.65	.55
2.80	.88	.80	.68	.59
2.90	.90	.83	.72	.63
3.00	.91	.95	75	.66
3.10	.93	.87	.78	.70
3.20	.94	.89	'81	.73
3.30	.95	.91	.84	.77
3.40	.96	.93	.86	.80
3.50	.97	.94	.88	.82
3.60	.98	.95	.90	.85
3.70	.98	.96	.92	.87
3.80	.98	.97	.93	.89
3.90	.99	.97	.94	.91
4.00	.99	.98	.95	.92
4.10	.99	'98	.96	.94
4.20	-	.99	.97	.95
4.30	-	.99	.98	.96
4.40	-	.99	.98	.97
4.50	-	.99	.99	.97
4.60	-	-	.99	.98
4.70	-	-	'99	.98
4.80	-	-	.99	.99
4.90	-	-	-	.99
5.00	-	-	-	.99

Table from Howell, D. C. (1997) Statistical Methods
for Psychology (4th ed.) Belmont, CA: Duxbury.

What should be the role of P-values and confidence intervals in the interpretation of scientific results? This question is not new. Increasing emphasis has been placed on the use of effect size reporting in the analysis of social science data. ***Reform of statistical practice*** in the social and behavioral sciences requires wider use of **confidence intervals** (CIs) and **effect size** measures. For decades, many advocates of statistical reform have recommended CIs as an alternative, or at least as a supplement, to *p* values (Cumming & Finch, 2001). The American Psychological Association's (APA, 2001) Publication Manual now calls CIs "the best reporting strategy" (p. 22). The APA manual stated "*because confidence intervals combine information on location and precision and can often be directly used to infer significance levels, they are, in general, the best reporting strategy. The use of confidence intervals is therefore strongly recommended*" (APA, 2001, p. 22).

Formulas:

$$SS_{between} = \Sigma n_j \left(\overline{X}_j - \overline{X}.. \right)^2$$

$$SS_{error} = \Sigma \left(n_j - 1 \right) s_j^2$$

$$\eta^2 = \frac{\Sigma \left(\overline{Y}_j - \overline{Y}.. \right)^2}{\Sigma \left(Y_{ij} - \overline{Y}.. \right)^2} = \frac{SS_{treatment}}{SS_{total}}$$

But, eta squared is biased because it is optimized for the particular set of data we have. Therefore we need to go to a less biased statistic:

$$\omega^2 = \frac{SS_{treatment} - \left(k - 1 \right) MS_{error}}{SS_{total} + MS_{error}} = \frac{507.84 - (4-1)55.579}{2786.58 + 55.579} = .12$$

ω^2 is always going to be smaller than η^2, but generally it is only a little smaller.

Confidence Interval (CI): For a given estimate, a 95% confidence interval (CI) is the set of all parameter values (i.e., hypotheses) for which $p \geq 0.05$ If the underlying statistical model is correct and there is no bias, the proportion of CIs derived over unlimited repetitions of the study containing the true parameter value is no smaller than 95% and is usually close to that value. This means that a confidence interval produces a move from a single value, or point estimate, to a range of possible effects in the population about which we want to draw conclusions. A confidence interval conveys information about both magnitude of effect and precision and therefore is preferable to a single p-value

References

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.

Hedges, L.V. (1982). Estimation of Effect from a Series of Independent Experiments. *Psychological Bulletin* **92**, 490-499.

Miettinen, O.S. (1985). *Theoretical epidemiology*. Albany: Delmar Publishers.

Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. New York: Guilford.

Uses and misuses of null hypothesis testing

Article: **Monterde-i-Bort, Frías-Navarro, D, & Pascual-Llobell, J. (2010). Uses and abuses of statistical significance tests and other statistical resources: a comparative study. *European Journal of Psychology of Education*, 1-19.**

———————————————————

Statistical significance is too often confused with the idea of the size or power of an effect. Over time, the technical meaning of 'significance' has become confused with the term 'important', and the way it is used often reflects a basic misunderstanding of probability theory (Matthey, 1998).

When something is reported to be statistically 'significant' it is often understood to mean that this is an indication of importance when often the true difference between means or amount of variance accounted for is of minimal practical significance (McCartney & Rosenthal, 2000). Thus the problem with null hypothesis significance testing is that something can be significant statistically yet be quite trivial in absolute terms. The probability value is crucially affected by the size of the sample and even very small differences between means can be statistically significant if enough data is gathered. Campbell (2005) indicated "For example, consider a hypothetical weight reduction program. With 20 clients in a control and treatment group, a mean difference in weight between the groups of 0.1kg is both statistically nonsignificant and obviously too small a difference to support the treatment program. But, if you keep the mean difference the same (0.1kg) and increase the numbers per group to 200 the difference becomes statistically significant at $p < .05$. With $N = 2000$ the difference becomes significant with $p < .0001$. The mean difference between the groups is still only 0.1 units but it is clear that by simply increasing N it is possible to create statistical significance.

When we estimate a parameter such as the relative risk o the mean, each possible value of that parameter is the expected value under some hypothesis, and each hypothesis has a P-value.

What we call "the" P-value is the P-value for the null hypothesis. Approximately, each P-value is the probability of obtaining an estimate at least as far from a specified value as the estimate we have obtained, if that specified value were the true value. It follows that no P-value, for the null hypothesis or any other, is the probability that the specified hypothesis is true. As an obvious example, the hypothesis corresponding to the point estimate has a (two-sided) P-value of 1.0. However, we do not treat our point estimates as absolutely certain to be true. Neither is the point estimate, in general, the most probable value. For a given estimate, the 95% confidence interval is the set of all parameter values for which P $ 0.05. For the value at each limit of a 95% confidence interval, P 5 0.05 (two-sided). Thus, if either of the 95% confidence limits for a relative risk estimate equals 1.0 (the null value of this parameter), we can infer that the null P-value is 0.05. From this link between confidence intervals and P-values, it follows that a 95% confidence interval is not a range of values within which the unknown true value lies with 95% probability.

The null hypothesis or any other hypothesis can be highly probable even though its P-value is less than 0.05. The null hypothesis or any other hypothesis can have a low probability even though its P-value is greater than 0.05. A relative risk can be highly probable even though it lies outside a 95% confidence interval. A relative risk can be highly improbable even though it lies inside a 95% confidence interval.

The indispensable role of hypotheses in the computation of P-values and confidence intervals, with each hypothesis assigning a probability to each estimate we might possibly obtain, means that these measures are not the descriptive statistics they are sometimes said to be.11 P-values and confidence intervals are inferential statistics, but the flow of the inference is a deductive flow, in which hypotheses confer probability "down" to estimates.

The only way we can determine the probability of the null hypothesis, or a range of values within which the true value lies with a given level of probability, is by using Bayesian methods.

It has become increasingly clear that the null P-value (hereafter called "the" P-value) does not do a very good job of the task for which it was originally intended: to quantify the statistical evidence against the null hypothesis (Poole, 2001).

One minus the Type I error rate is the specificity of a significance test: the probability of not declaring "significance" when the null hypothesis is true. One minus the Type II error rate is the test's power or sensitivity: the probability of declaring "significance" when the alternative hypothesis is true. The interest many methodologist express in how low the P-value is, if it is lower than 0.05,18 raises still other questions. How much evidence against the null hypothesis do we have when P 5 0.04, or when P 5 0.001?

Statisticians who have examined these questions in detail have found, under widely ranging conditions, that P-values on the order of 0.05, 0.01, and even lower provide much less evidence against the null hypothesis than they appear to provide at face value. As a general matter, P-values in the vicinity of 0.05 provide almost no evidence against the null hypothesis at all. P 5 0.04, for instance, is typically found to be almost equally probable under the null and alternative hypotheses.

Bibliography

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, *10*, 389-396.

Campbell, A. (2005). It's a Difference, but Is It Significant? *ANZJFT*, 26, 227–229.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use and calculation of confidence intervals based on central and noncentral distributions. *Educational and Psychological Measurement*, *61*, 530-572.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals, and how to read pictures of data. *American Psychologist*, *60*, 170–180.

Cumming, G., Williams, J., & Fidler, F. (2004). Replication and researchers' understanding of confidence intervals and standard error bars. *Understanding Statistics*, *3*, 299–311.

Denis, D. (2003). Alternatives to Null Hypothesis Significance Testing, *Theory & Science*, *4*, 1.

Fan, X. (2001). Statistical Significance and Effect Size in Education Research: Two Sides of a Coin. *Journal of Educational Research*, *94*, 275–282.

Fidler, F. (2005). From statistical significance to effect estimation: Statistical reform in psychology, medicine and ecology. Department of History and Philosophy of Science. University of Melbourne. Retreived from http://www.botany.unimelb.edu.au/envisci/docs/fidler/fidlerphd_aug06.pdf

Fidler, F., & Thompson, B. (2001). Computing correct confidence intervals for ANOVA fixed and random effect sizes. *Educational and Psychological Measurement*, *61*, 575-604.

Jacobson, N. S. & Revenstorf, D. (1988). Statistics for Assessing the Clinical Significance of Psychotherapy Techniques: Issues, Problems, and New Developments. *Behavioral Assessment*, *10*, 133–145.

Jacobson, N. S. & Truax, P. (1991). Clinical Significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research. *Journal of Consulting and Clinical Psychology*, *59*, 12-19.

Kazdin, A. E. (1999). The Meanings and Measurement of Clinical Significance. *Journal of Consulting and Clinical Psychology*, *67*, 332–339.

Matthey, S. (1998). p < .05. But Is It Clinically Significant? Practical Examples for Clinicians. *Behaviour Change*, *15*, 140–146.

McCartney, K., & Rosenthal, R. (2000). Effect Size, Practical Importance, and Social Policy for Children. *Child Development*, *71*, 173–180.

Poole, Ch. (2001). Low P-Values or Narrow Confidence Intervals: Which Are More Durable? *Epidemiology*, *12*, 291-294.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher*, *25*, 26–30.

Thompson, B. (2002). 'Statistical', 'Practical', and 'Clinical': How Many Kinds of Significance do Counselors Need to Consider? *Journal of Counseling and Development*, *80*, 64–71.

Zakzanis, K. K. (2001). Statistics to Tell the Truth, the Whole Truth, and Nothing but the Truth: Formulae, Illustrative Numerical Examples, and Heuristic Interpretation of Effect Size Analyses for Neuropsychological Researchers. *Archives of Clinical Neuropsychology*, *16*, 652–667.
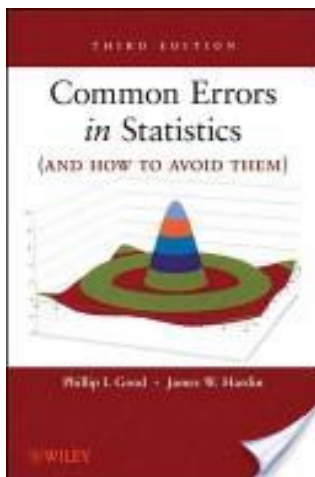
Visualizing Statistical Concepts: use computer graphics to understand statistical concepts

http://www.du.edu/psychology/methods/concepts/#Randomization

Simple Interactive Statistical Analysis
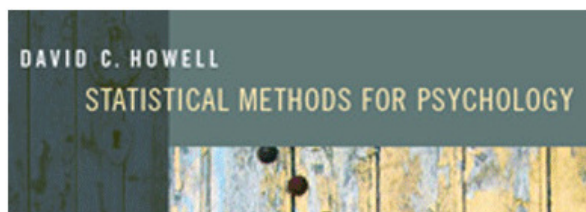
http://www.quantitativeskills.com/sisa/

Books

Good, P.I., & Hardin, J.W. (2009). Common Errors in Statistics (and How to Avoid Them) (Third Edition) . John Wiley and Sons. Introduction
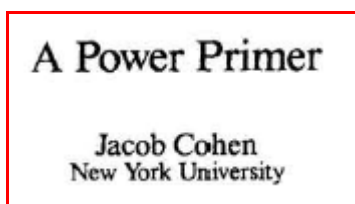
Web

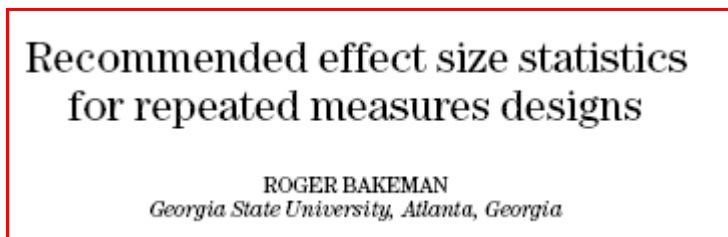http://www.uvm.edu/~dhowell/methods7/

- http://www.psych.utoronto.ca/courses/c1/c1.html
- http://www.psych.utoronto.ca/courses/c1/statstoc.htm

➢ **Visualizing Statistical Concepts:**

http://www.du.edu/psychology/methods/concepts/

Materials

A Power Primer

Jacob Cohen
New York University

http://www.uv.es/friasnav/Cohen1992

Recommended effect size statistics
for repeated measures designs

ROGER BAKEMAN
Georgia State University, Atlanta, Georgia

http://www.uv.es/friasnav/Bakeman2005

Statistics calculators:

http://www.danielsoper.com/statcalc/

Reporting Statistics in Psychology:

http://huberb.people.cofc.edu/Guide/Reporting_Statistics%20in%20Psychology.pdfs

Publications of **Frías-Navarro D.** about "Practice-Based Evidence"

Frías, M.D. y Pascual, J. (2003). Psicología Clínica Basada en Pruebas: efecto del tratamiento. *Papeles del Psicólogo*, 24, 85.

Frías, M.D., Pascual, J. y García, J.F.(2002). Concepto y método de la psicología basada en la evidencia. III Congreso Virtual de Psiquiatría. Interpsiquis 1-28 Febrero. Psiquiatria.com. Presentada en la mesa redonda de *Psiquiatría Basada en la Evidencia*.

Frías, M.D., Pascual, J. y Monterde, H. (2004). Impacto de la búsqueda de la evidencia en la formación universitaria del psicólogo. V Congreso Virtual de Psiquiatría. Interpsiquis 1-28 Febrero. Psiquiatria.com. Presentada en la mesa redonda de *Psiquiatría Basada en la Evidencia*.

Frías, M.D., Pascual, J. y Monterde, H. (2006). La mejor evidencia de la investigación psicológica clínica. VII Congreso Virtual de Psiquiatría. Interpsiquis 1-28 Febrero. Psiquiatria.com. Presentada en la mesa redonda de *Salud Mental Basada en la Evidencia*.

Monterde-Bort, H., Pascual Llobell, J. y Frías Navarro, M.D. (2006). Errores de interpretación de los métodos estadísticos: importancia y recomendaciones. *Psicothema*, **18**, 4, 848-856, 2006.

Pascual J., Frías M.D. y Monterde, H. (2004). Tratamientos psicológicos con apoyo empírico y práctica basada en la evidencia. *Papeles del Psicólogo*, **87**.

STUDY DESIGNS

➢ **COHORT STUDY DESIGN:** an observational research design which begins when a group of people (a cohort) initially free of disease or other outcome, are classified according to a given exposure, and then followed up over time.
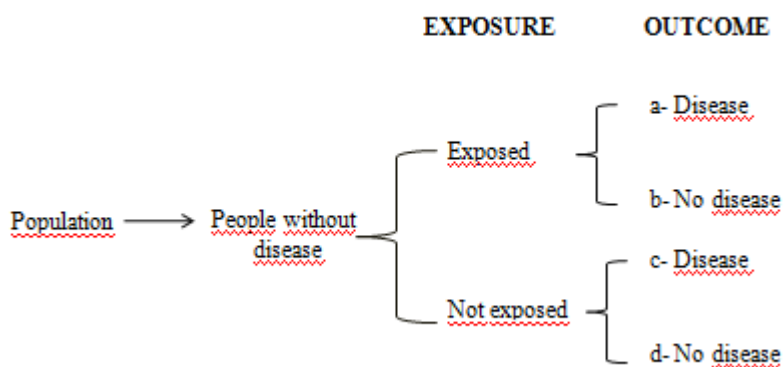
Cohort structure

A cohort study is useful for estimating the risk of disease or other outcome, the incidence rate and/or relative risks.

|	Case (Number)	Non-cases (Unnecessary if T known)	Total Exposure
Exposed	**A**	**B**	T. Exposed
Not Exposed	**C**	**D**	T. Non Exposed
	T. Cases	T. Non-cases	TOTAL

Types of cohort: **prospective** and **retrospective**. The simplest cohort design is prospective, i.e., following a group forward in time, but a cohort study can also be 'retrospective'. In general, the descriptor, 'prospective' or 'retrospective' indicates when the cohort is identified relative to the initiation of the study.

Prospective cohort (*concurrent*; *longitudinal study*). An investigator identifies the study population at the beginning of the study and accompanies the subjects through time. In a prospective study, the investigator begins the study at the

same time as the first determination of exposure status of the cohort. When proposing a prospective cohort study, the investigator first identifies the characteristics of the group of people he/she wishes to study. The investigator then determines the present case status of individuals, selecting only non-cases to follow forward in time. Exposure status is determined at the beginning of the study.

Retrospective cohort study (*historical cohort*; *non-concurrent prospective cohort*). An investigator accesses a historical roster of all exposed and non-exposed persons and then determines their current case/non-case status. The investigator initiates the study when the disease is already established in the cohort of individuals, long after the measurement of exposure. Doing a retrospective cohort study requires sound data on exposure status for both cases and non-cases at a designated earlier time point. Potential problems with the retrospective cohort approach include selection bias and misclassification bias because of the retrospective nature of the study.

➢ **CASE-CONTROL STUDY:** A case-control study is distinguished by the fact that subjects are selected on the basis of whether or not they have the disease (or other outcome) of interest. Cases (those with disease) are then compared to controls (those without disease) in terms of their history of exposure to a hypothesized causal factor.
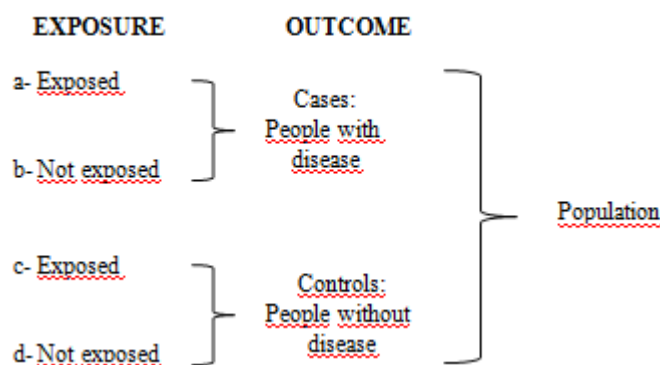
Case-control structure

> **RANDOMISED CONTROLLED TRIAL (RCT):** an experimental study design in which subjects are randomly allocated to groups which either do (treatment group) or do not (control group) receive a therapeutic or preventive intervention being evaluated.
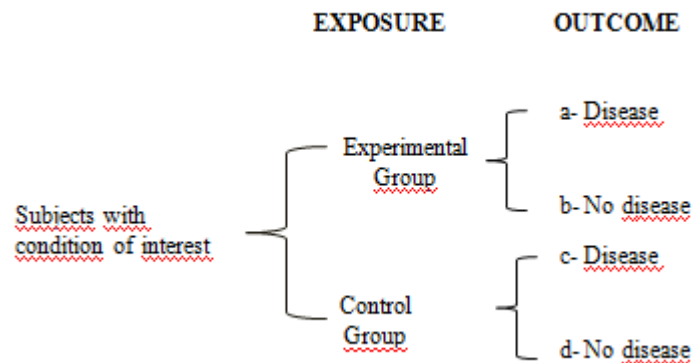
Randomised Controlled Trial structure

- A RCT provides the best chance of obtaining strong evidence of a cause and effect.
- Allows standardization of the eligibility criteria, treatment (or other intervention) and outcome assessment.
- Allows the use of statistical methods which have few inbuilt assumptions.
- Expensive to undertake in terms of time, money and people.
- May be unethical for certain research questions.
- May be unsuitable because of the lack of cooperation of subjects or rarity of outcome.