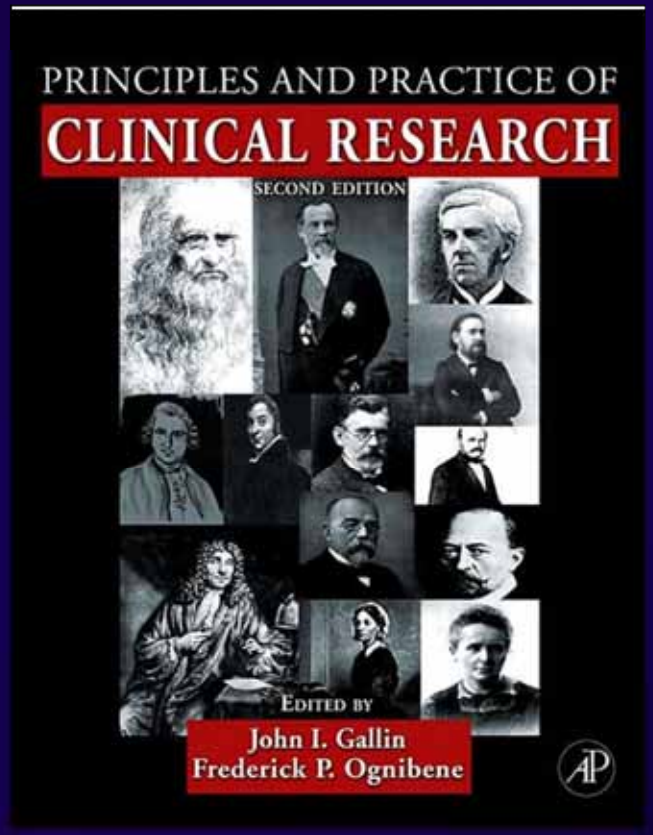
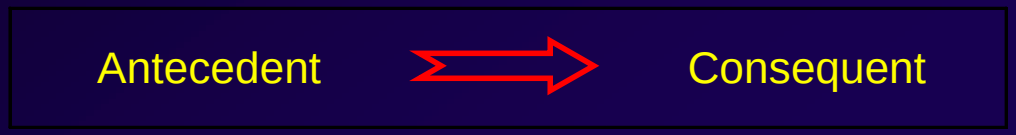
I. Group Studies of Treatment (adapted from Gallin & Ognibene, 2007)

Gallin, J. I., Ognibene, F. P. (2007). *Principles and Practices of Clinical Research*, 2nd Ed. Academic Press.



A. A straightforward challenge in logical conclusion



Exposure	Event	Good Thing or Bad Thing
Risk	Outcome	
Predictor Variable	Observed Variable	
Independent Variable	Dependent Variable	

Events, e.g.,

1. Death
2. Disease state
3. Successful treatment outcome
4. Return to work
5. Discharge to independent living

B. Ecological or correlational designs

The units of measure are population values (not observations of individuals)

C. Observational designs

1. Cross Section

   a. Select a sample

   b. Measure both the antecedent and consequent

   c. Examine the linkage

   d. Note: A cross section design produces an estimate of prevalence re. antecedents

   e. Note: The results are mostly descriptive and have value for generating hypotheses.

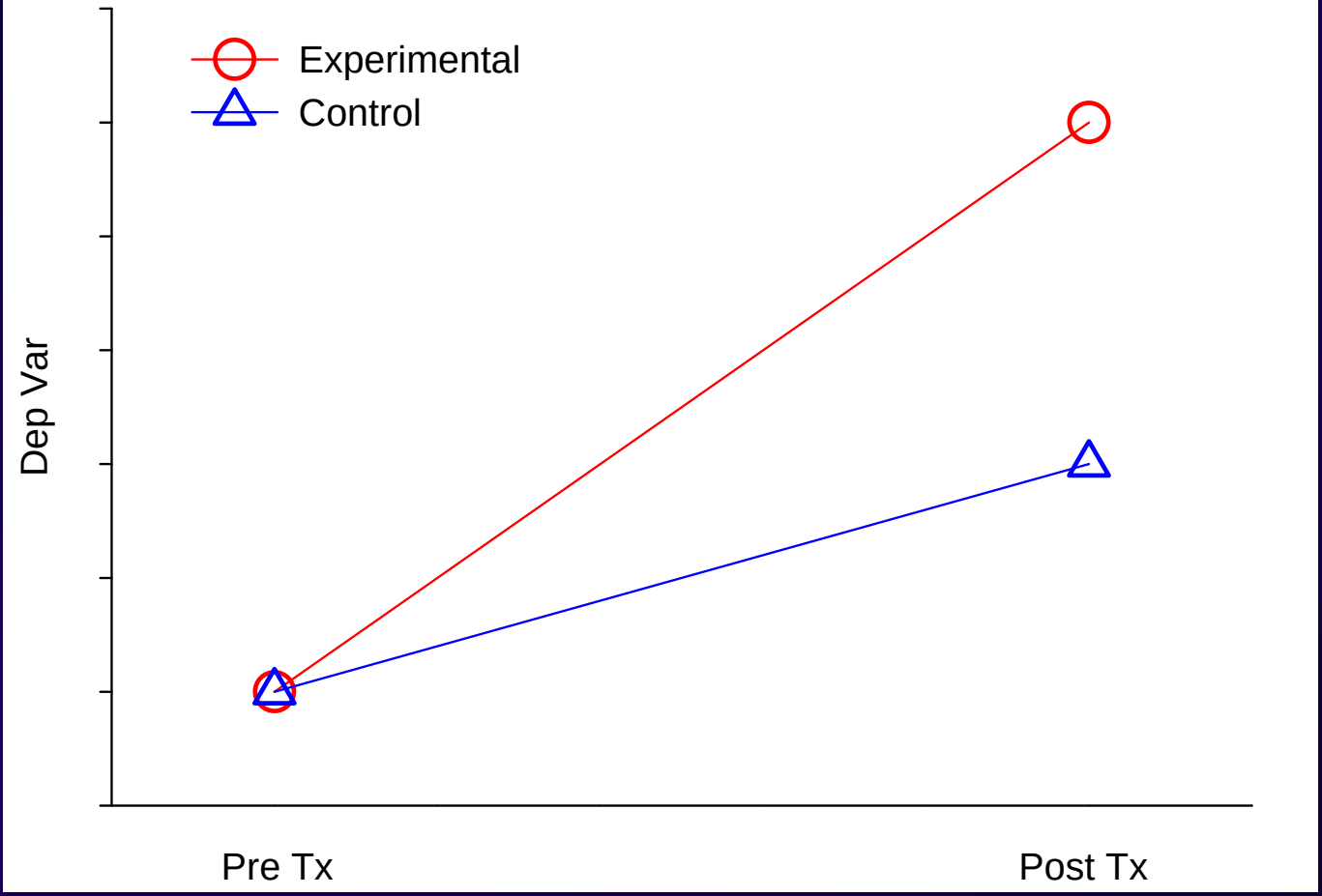
2. Case Control (retrospective)
    a. Select a sample of cases on the basis of the consequent
    b. Select a sample of controls on the basis of the consequent
    c. Look backwards in time to documented antecedents for explanations
    d. Examine the linkage (often through an odds ratio)
    e. Note: A prospective variant is termed a Case Control Crossover

3. Cohort (prospective)
    a. Enlist the cooperation of a cohort of participants and measure the antecedent
    b. Follow members of the cohort forward in time and then measure the consequent
    c. Examine the linkage (often through a relative risk ratio)
    d. Note: A retrospective variant is possible
    e. Note: A cohort design produces an estimate of incidence

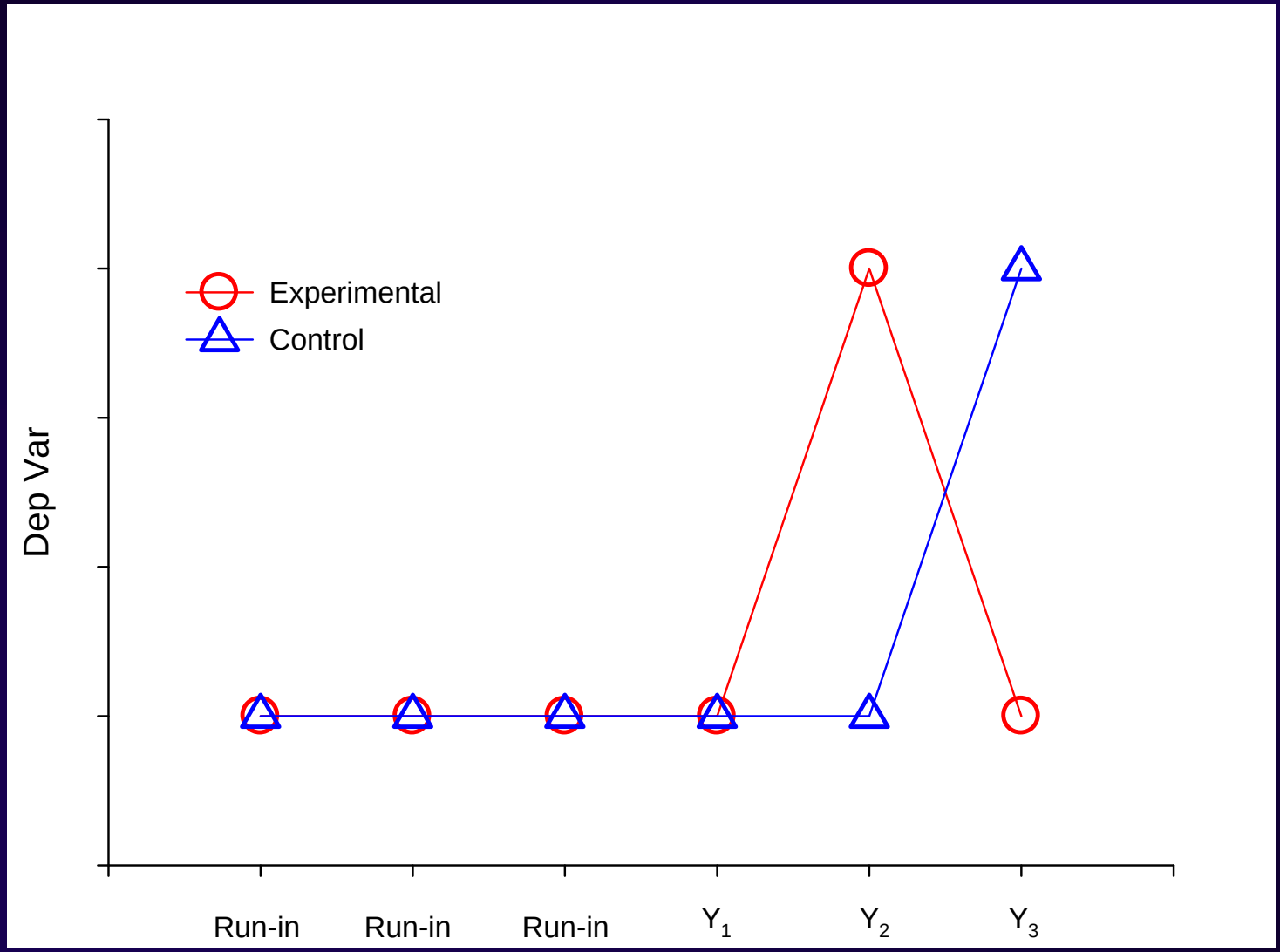
D. Causal Inference Studies: Controlled Trials

1. Parallel groups
    a. Sample participants
    b. Allocate participants to arms
    c. Make baseline observations
    d. Implement protocols making intermediate observations
    e. Conclude protocols and make post observations
    f. Perhaps later, make follow up observations

	Pre	Post
Control		
Experimental		


Experimental
Control
Dep Var
Pre Tx
Post Tx

2. Cross Over
    a. Sample participants
    b. Allocate participants to arms
    c. Make run-in observations
    d. Make pre-period-1 observations
    e. Make period-1/period-2 cross-over observations
    f. Make post-period-2 observations


Experimental
Control
Dep Var
Run-in
Run-in
Run-in
Y1
Y2
Y3

II. How Do I Establish Just What Constitutes an Important Finding and How Many Participants Do I need to Detect It?

A. Premise

1. The role of the binary choice between [$p \leq \alpha$] and [$p > \alpha$], is necessary for deciding the tenability of a null hypothesis (statistical significance).

2. Rejecting a false null hypothesis is wholly insufficient for deciding the meaningfulness of an outcome (clinical significance).

3. Setting $\alpha=0.05$ is a choice based largely in a rigid ritual rather than critical thought. However, a long history has brought us to this point.

4. What is needed to assess meaningfulness are point and interval estimates of effect size.

5. However, graduating effect size as small (d=0.20), medium (d=0.50), and large (d=0.80) flirts with becoming a rigid and meaningless ritual.

6. The value of an estimate of effect size produced through a new experiment is found in its relationship to the estimates of effect size produced in the studies that justified the new experiment.

   That is, just as the justification of an experiment is found in a focused set of existing studies, so too is the meaning of a new result uncovered in its relationship to the corresponding body of existing results.

   All interpretations of effect size are local.

7. The width of a confidence interval about an estimate of effect size is a measure of experimental precision.

   As error variance in a study increases, so does the width of the confidence interval about the estimate of effect size produced by that study.

B. Bruce Thompson figured this out quite a while ago.

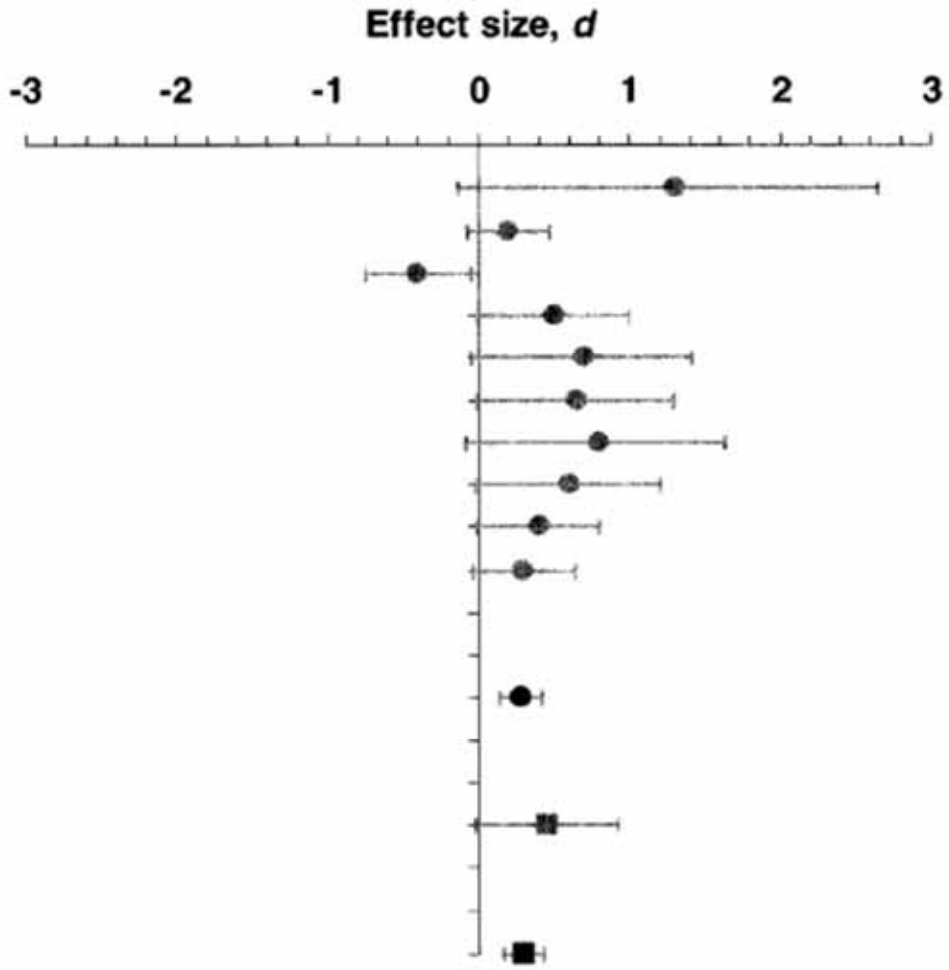
Thompson, B. (2002). What Future Quantitative Social Science Research Could Look Like: Confidence Intervals for Effect Sizes, *Educational Researcher, 31*, 25-32.

Table 1. ***Practical and Statistical Significance Statistics for 10 Previous Studies and One New Study***

	Results					
Study	*d*	*n*	*t* calc	*p* calc	Decision	CI for *d*
Prior Literature						
1	1.30	4	2.600	0.080	NS	−0.13 to 2.64
2	0.20	53	1.456	0.151	NS	−0.07 to 0.47
3	−0.40	34	−2.332	0.026	*	−0.75 to −0.05
4	0.50	17	2.062	0.056	NS	−0.01 to 1.00
5	0.70	9	2.100	0.069	NS	−0.05 to 1.42
6	0.65	11	2.156	0.056	NS	−0.02 to 1.29
7	0.80	7	2.117	0.079	NS	−0.09 to 1.64
8	0.60	12	2.078	0.062	NS	−0.03 to 1.21
9	0.40	25	2.000	0.057	NS	−0.01 to 0.80
10	0.30	35	1.775	0.085	NS	−0.04 to 0.64
Past research, pooled	*0.278*	*207*	*4.000*	*0.00009*	***	*0.14 to 0.42*
Current study						
11	0.45	19	1.962	0.065	NS	−0.03 to 0.92
Past and current, pooled	*0.292*	*226*	*4.390*	*0.00002*	***	*0.16 to 0.42*

Note. Pooled results are presented in italics. For the one-group case $t = d$ (square root of n). Exact p calculated values can be found in Excel by using the function "TDIST(t, $n-1$, 2)". The weighted average effect size can be computed, for example for the 10 prior studies, as [(1.30 × 4) + (0.20 × 53) + ...(0.30 × 35)] / [4 + 53 + ...35] = [5.2 + 10.6 + ...10.5] / 207 = 57.45 / 207 = .278. calc = calculated.

* $p < .05$ *** $p < .001$



Note. The top 10 effect size intervals are for the studies from the prior literature, with the *d*'s represented as gray circles within the intervals. The pooled weighted average *d* (i.e., .278) is presented next, as a darker oval. The result in the present study (*d* = .45) is then presented as a gray square. Finally, the weighted average *d* across all 11 studies is presented last as a dark square.

noteworthy and then simply ignore the study. But once reporting effect sizes becomes normatively standard practice, at least all authors and readers will then be in a position to evaluate how replicable or stable are the effects within a given area of inquiry.

Three Recommendations for Practice

Several recommendations for practice are suggested here:

1. *Report and explicitly interpret effect sizes in the context of effect sizes from prior related studies and not by invoking rigid benchmarks.* The potential benefits of reporting and interpreting an effect size (e.g., Cohen's *d*, Glass' delta, η^2, or adjusted R^2) arise not from interpreting effects against benchmarks, but rather by comparing effect sizes directly with the effects reported in related prior studies (Wilkinson & APA Task Force, 1999, p. 599). The overly rigid use of fixed benchmarks for small, medium, and large effects fails to consider the possibility that small, replicable effects involving important outcomes can be noteworthy, or that large effects involving trivial outcomes may not be particularly noteworthy.

C. Minimal Clinically Important Difference (MCID)

Man-Son-Hing, et al. (2002) advanced the notion that not every statistically significant difference (proportion, correlation, etc.) is important.

Although the units-of-measure for Man-Son-Hing, et al. were descriptive statistics (rather than estimates of effect size), they also understood that all interpretations of experimental results are local.

On the basis of existing literature, a researcher must determine a criterion that a new result must exceed to be considered clinically significant: MCID

Adapting Man-Son-Hing, et al. by making the leap from mean differences to differences in effect sizes renders MCID practicable.

1. Three different examples of MCID

   a. No intervention is available for a certain debilitating condition.

      Any improvement, no matter how small relative to a no-treatment control, represents an important advancement in managing the condition.

      In this case, obtaining a value of say $d \geq .10$ could very well constitute an important difference.

b. An intervention protocol is broadly recognized as a clinical standard for care and is known to effect a level of change corresponding to an average effect size of d = .80 (i.e., an average effect size in comparison with no-treatment control studies).

A new technology is introduced as an alternate form of care but only at substantial cost in making the change from one technology to another.

The cost is deemed worthwhile if the new technology improves outcomes by at least 25%.

All other things remaining constant, an outcome of $d \geq .20$ is an important one in an ANCOVA of data obtained through a parallel-groups design contrasting the new technology and the old technology.

c. Consider the same situation but one in which the new technology achieves the same level of change as the old technology but at a substantially faster rate and substantially reduced cost.

In this case, d = 0.00 is an important outcome using the same research design.

That is, the new technology achieves the same outcome as the standard but in less time and at less cost. The analysis in this case would be supplemented with equivalency testing.

d. A new treatment protocol will be considered an important advancement if if produces an estimate of effect size that exceeds the average effect size of the treatment studies testing competing protocols.

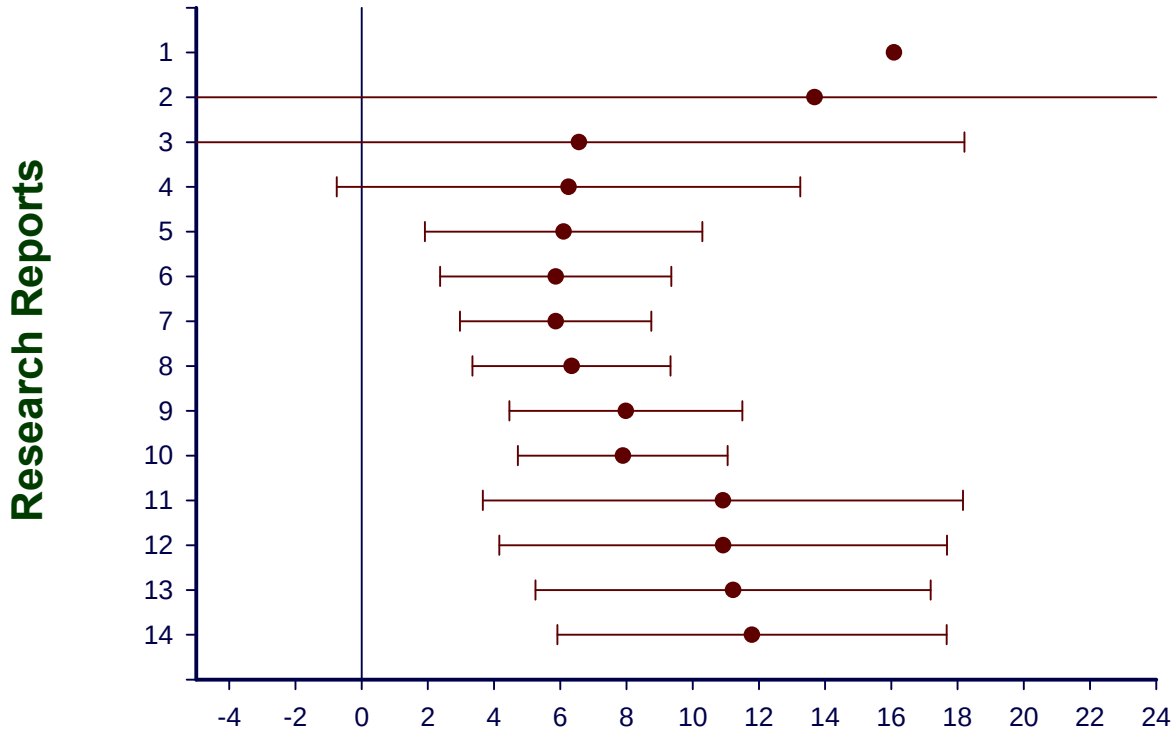
e. That same new treatment will be considered very important if it produces and estimate of effect size that equals or exceeds the upper boundary of the confidence interval about that average effect size.

Single-Subject Data: Direct-Treatment Effects

Study	Class	Phase	Obs.	d	Treatment
1	3	1	16	16.08	Auxiliary 'Is' training
2	3	1	10	9.85	Syntax stim.
3	3	1	103	4.76	Spoken + written modalities stim.
4	3	1	12	2.99	Syntax stim.
5	3	1	83	5.83	Wh interrogative training
6	3	2	17	2.75	LST
7	3	1	25	5.86	LST
8	3	2	18	13.42	Syntax Stim. & PACE
9	3	2	77	14.01	LST
10	3	2	23	6.54	LST
11	3	2	39	40.64	LST
12	3	1	9	11.59	LST
13	2	2	67	13.11	LST
14	3	2	18	27.73	LST

Single-Subject Direct Treatment Effects

Outcome: Syntax



Average of Effect Size with .95 CI
(Progressive Cumulative Average)

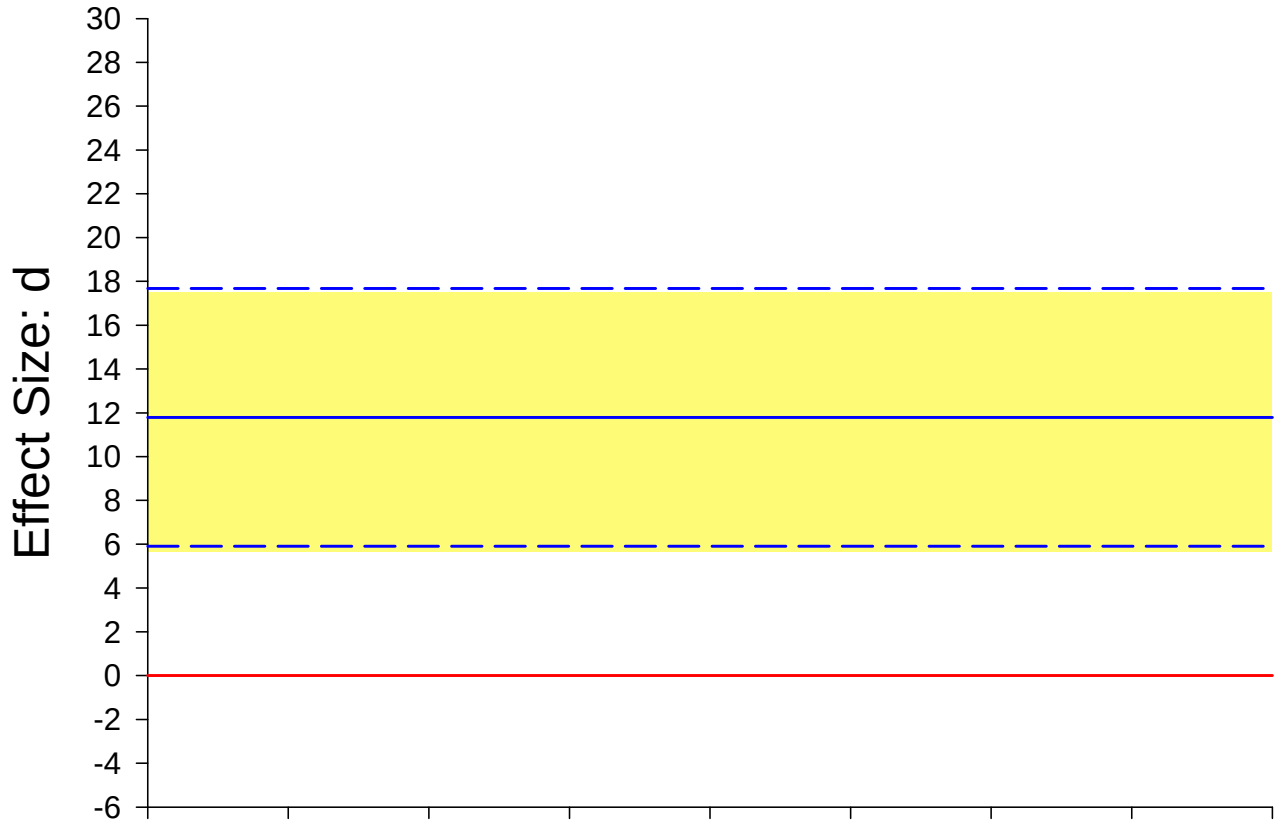
The weighted mean of these effects is 11.79.

A confidence interval for that mean value with probability set at .95 (i.e., $CI_{.95}$) equals ±5.88.

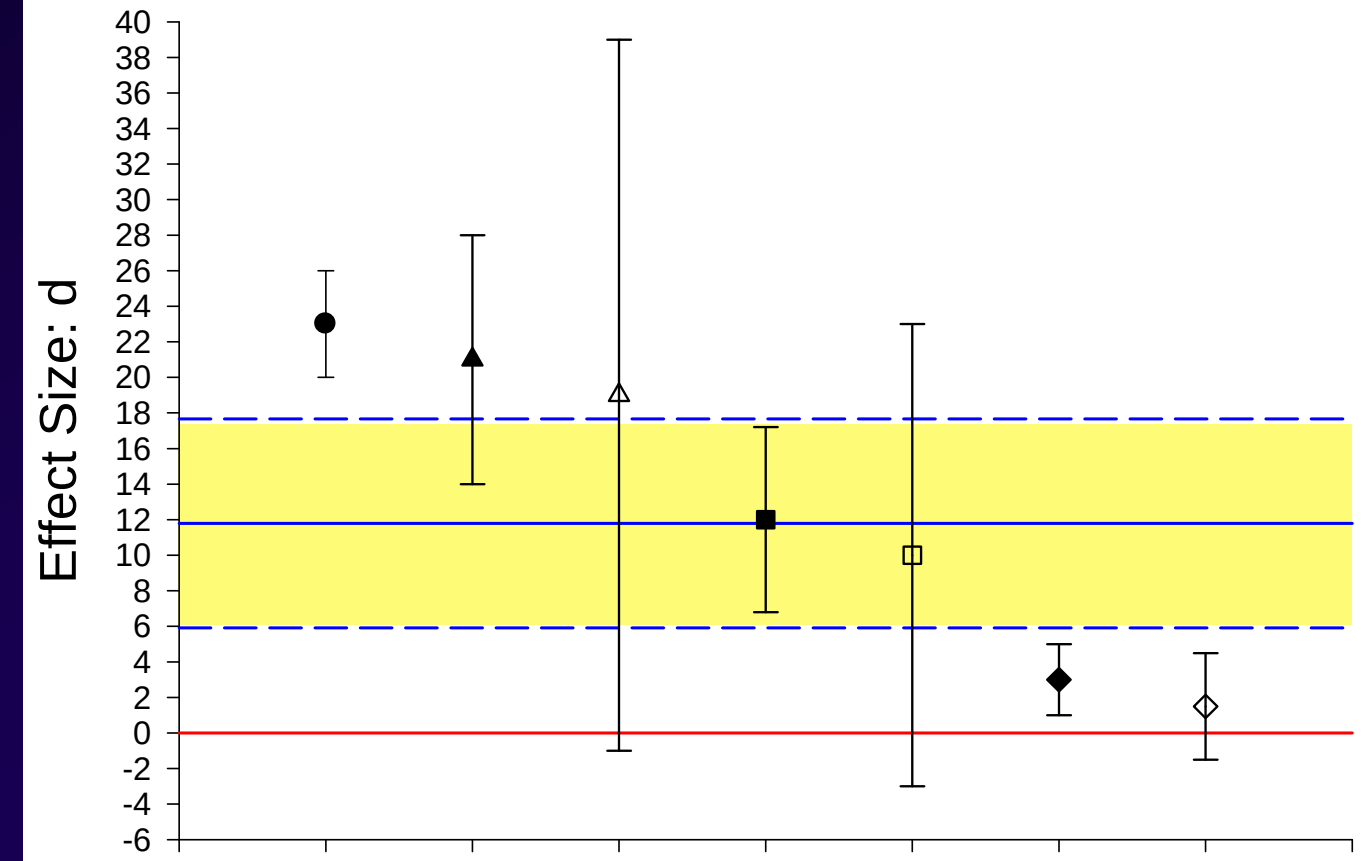
d		
Lower Limit	Mean	Upper Limit
5.91	11.79	17.67

Reasonably, we could set the size of a small effect at d=5.91, a medium effect at d=11.79, and a large effect at d=17.67.

$CI_{.95}$ Interval of Effect Size for Single-Subject Studies of Syntax Improvement Treatments



Possible Outcomes and Clinical Significance



How do I obtain values for this mini meta-analysis?

If a meta-analysis has been published in your target literature, you're golden.

If not, work with your statistician to obtain what you need.

III. Types of Effect Size

A. d

B. r

C. Odd Ratio

D. Relative Risk

E. Risk Reduction

F. NNT

IV. A Priori Statistical Power Analysis

The following four terms are algebraically linked.

A. Effect size

B. Type I error tolerance

C. Statistical power

D. Sample size (n)

Knowing the values of any three allows us to solve for the value of the fourth.

V. Obtaining and Reporting Estimates of Effect Size Obtained Through Your Study

A. Four benefits realized through reporting estimates of ES

1. Decreased reliance on, or misuse of, statistical significance

2. Meaningful interpretations observed results in the context of previous research through empirical, objective, and transparent means

3. Increased precision in designing experiments

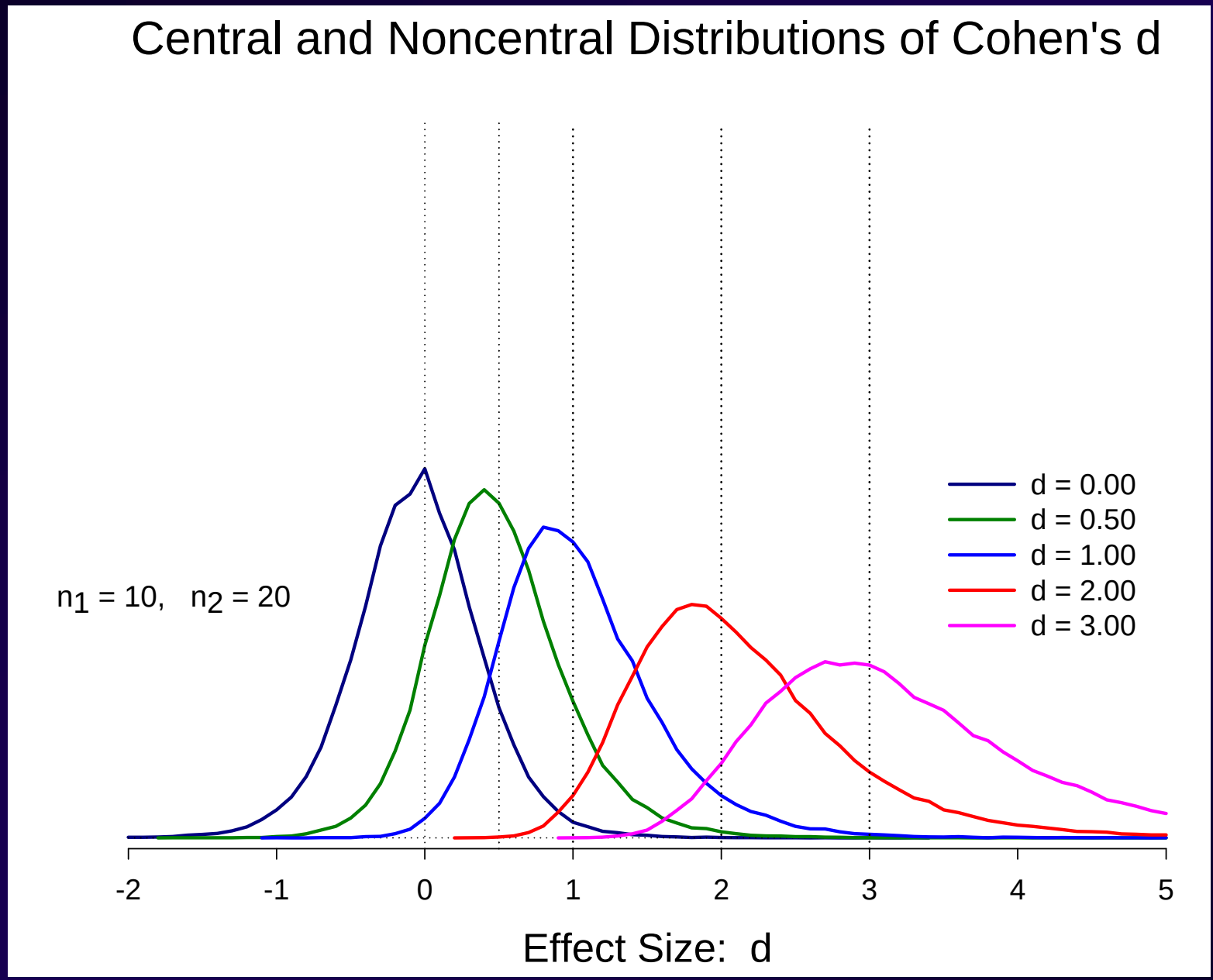
4. Direct support for eventual meta-analyses of clinical research.

B. In the course of the past 10 years, statisticians have made available a powerful tool for assessing a literature base, designing experiments, and interpreting results: noncentral confidence intervals (CI) for point estimates of effect size.

1. Because a non-zero estimate of effect size characterizes a departure from a null hypothesis, the sampling distribution forming the mathematical basis for a confidence interval is a noncentral distribution.

   Bird (2002), Cumming & Finch (2001), Fidler & Thompson (2001), Robey (2005) and Smithson (2001) constitute central readings

2. The mathematics of finding a point on a noncentral distribution are exceptionally complex.


Central and Noncentral Distributions of Cohen's d
$n_1 = 10, \quad n_2 = 20$
d = 0.00
d = 0.50
d = 1.00
d = 2.00
d = 3.00
-2
-1
0
1
2
3
4
5
Effect Size: d

3. Through advances in software applications, recently, statisticians have made noncentral distributions accessible for practitioners.

Confidence Intervals - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://psychology3.anu.edu.au/people/smithson/details/CIstuff/CI.html

Most Visited Getting Started Latest Headlines

Principles and Practice of Clinical Resea... | Confidence Intervals

Scripts and Software for Noncentral Confidence Interval and Power Calculations

This webpage contains download links for scripts that enable you to compute confidence intervals and power based on the noncentral t, F, and Chi-square distributions. So far, these are available in SPSS, SPlus, and R. I will update this webpage with links to other freeware and/or scripts as they become available and known to me.

A general (but not expensive!) reference on confidence intervals that covers noncentral confidence interval estimation, as well as the more well-known confidence intervals, is:

Smithson, M.J. (2003). Confidence Intervals.
Quantitative Applications in the Social Sciences Series, No. 140.
Thousand Oaks, CA: Sage.

SPSS Scripts

There is a PDF file based on a short workshop I've given on noncentral confic ...cludes instructions on how to use these scripts. This will have to do as an instruction manual until I write a more thorough-going version.

Download PDF file.

Noncentral F files	Noncentral t files	Noncentral Chi-square files
Download NoncF3.sps	Download NoncT2.sps	Download NoncChi.sps
Download NoncF.sav	Download NoncT.sav	Download NoncChi.sav
Download F2R2.sps	Download T2D.sps	

SAS Scripts

These scripts come with a "Readme" PDF file. Users not familiar with noncentral confidence intervals may also wish to download the workshop instructional PDF file mentioned above.

Download Readme file.

Noncentral F file	Noncentral t files	Noncentral Chi-square files
Download NoncF2.sas	Download Nonct.sas	Download Noncchi.sas
	Download t2d1samp.sas	Download CramersV.sas
	Download t2d2samp.sas	

SPlus and R Scripts

These scripts come with a "Readme" PDF file. Users not familiar with noncentral confidence intervals may also wish to download the workshop instructional PDF file mentioned above.

Done

ESCI - Exploratory Software for Confidence Intervals - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.latrobe.edu.au/psy/esci/index.html

Most Visited Getting Started Latest Headlines

Principles and Practice of Clinical Resea... | Confidence Intervals | ESCI - Exploratory Software for C...

Skip to content Site map SEARCH

La Trobe Home | About La Trobe | Courses | Campuses | Faculties | Research | Contacts

esci *Exploratory Software for Confidence Intervals*

ESCI HOME | DOWNLOADS | STATISTICS REFORM | FAQ | CONTACT ESCI

*ESCI (Pronounced "esky") is a set of **interactive simulations** that run under Microsoft Excel*

With ESCI you can:

- explore many **Confidence Interval** (CI) concepts
- calculate and **display CIs** for your own data, for some simple designs
- calculate CIs for Cohen's **standardised effect size *d***
- explore **noncentral *t*** distributions and their role in **statistical power**
- use CIs for simple **meta-analysis**, using original or standardised units
- explore all these concepts via vivid interactive **graphical simulations**

Professor Geoff Cumming
School of Psychological Science,
La Trobe University,
Bundoora, Melbourne,
Australia

New ESCI modules

The following ESCI modules go with published articles. (There is information about these and other articles, including abstracts and/or downloads of the articles themselves, at my School page.) These ESCI modules are free downloads, for non-commercial use, and were developed in Microsoft Excel 2003 or XP. (They may also run in earlier versions of Excel.) Sorry, no Mac versions yet.

ESCI JPP (downloads) is a module that allows you to calculate and display CIs for a wide variety of measures and designs. It goes with this article:
Finch, S., & Cumming, G. (2009). Putting research in context: Understanding confidence intervals from one or more studies. *Journal of Pediatric Psychology*. doi: 10.1093/jpepsy/jsn118

ESCI PPS p intervals (downloads) is a module that allows you to explore three of the figures in:
Cumming, G. (2008). Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. Perspectives on Psychological Science, 3, 286-300.

NOTE: Excel 2007 The above module runs in Excel 2007, as well as Excel 2003. The modules

Done

University of New South Wales - School of Psychology - PSY - A program for contrast analysis - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.psy.unsw.edu.au/research/resources/psyprogram.html

Most Visited Getting Started Latest Headlines

Principles and Practice of Clinical Resea... | Confidence Intervals | ESCI - Exploratory Software for Confid... | University of New South Wales - ...

THE UNIVERSITY OF NEW SOUTH WALES • SYDNEY • AUSTRALIA

UNSW psychology faculty of science

Search

Contacts | Sitemap

Future Students | Current Students | People & Contacts | Research | Alumni & Supporters | News & Events | Careers | About the School

PSY Statistical Program

PSY: A program for contrast analysis - Kevin Bird, Dusan Hadzi-Pavlovic, and Andrew Isaac

Some background - Why PSY was developed:

The APA Task Force on Statistical Inference has recommended that interval estimates of effect sizes should always be presented for primary outcomes (Wilkinson and the Task Force on Statistical Inference, 1999). When fixed-effects analysis of variance is used to analyse data, effect sizes are usually expressed as raw or standardized values of contrasts on means. It is often difficult or impossible to obtain appropriate confidence intervals on contrast values from statistical packages, particularly the simultaneous confidence intervals required for unrestricted or post hoc analyses. The PSY program provides confidence intervals on contrasts for a number of designs and analysis strategies.

PSY is very easy to use for planned or post hoc analyses of single-factor designs (with or without repeated measures) and for two-factor designs with one between-subjects and one within-subjects factor. It can be used for the analysis of more complex factorial designs, but it is a little less user-friendly when simultaneous confidence intervals are required from complex designs. PSY can accept user-supplied critical constants for the purpose of confidence interval construction, so advanced users can control familywise error rates for non-standard analyses. PSY can also supply standard or Bonferroni-adjusted critical values from t, F, GCR (greatest characteristic root) and SMR (studentized maximum root) distributions. Examples of PSY analyses of data from various designs may be found in Bird (2004).

Those who wish to use PSY to carry out tests or construct confidence intervals based on the SMR approach to the analysis of product interaction contrasts (Boik, 1993) may be interested in downloading an SPSS macro that calculates SMR statistics in analyses including more general product contrasts. You can download the macro as a **text file** or an SPSS **syntax file**.

To run PSY, you need Windows 95 (or later). Click **here** to download PSY

Before running PSY, consult the ReadMe document. Please send feedback to Kevin Bird (K.Bird@unsw.edu.au).

References

- Bird, K.D. (2004). *Analysis of variance via confidence intervals*. London: Sage Publications.
- Boik, R.J. (1993). *The analysis of two-factor interactions in fixed effects linear models*. Journal of Educational Statistics, 18, 1- 40.
- Wilkinson, L., & the Task Force on Statistical Inference, APA Board of Scientific Affairs (1999). *Statistical methods in psychology journals: guidelines and explanations*. American Psychologist, 54, 594-604.

Printer friendly page

Done