



ΕΛΛΗΝΙΚΟ ΣΤΑΤΙΣΤΙΚΟ ΙΝΣΤΙΤΟΥΤΟ (Ε.Σ.Ι.)

Π Ρ Α Κ Τ Ι Κ Α
15^{ΟΥ} ΠΑΝΕΛΛΗΝΙΟΥ ΣΥΝΕΔΡΙΟΥ
ΣΤΑΤΙΣΤΙΚΗΣ

Β' Τόμος



ΙΩΑΝΝΙΝΑ, 8-11 Μαΐου 2002

ΑΝΑΛΥΣΗ ΙΣΧΥΟΣ ΤΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΕΛΕΓΧΩΝ: ΜΙΑ ΠΡΩΤΗ ΠΡΟΣΕΓΓΙΣΗ

Γεώργιος Μενεξές

Τμήμα Εφαρμοσμένης Πληροφορικής, Πανεπιστήμιο Μακεδονίας

ΠΕΡΙΛΗΨΗ

Ο έλεγχος υποθέσεων στην επιστημονική έρευνα, έχει να επιδείξει μια σαφή προτίμηση στη χρησιμοποίηση της στατιστικής σημαντικότητας ως κριτηρίου ή προσέγγισης σχετικά με την απόρριψη ή όχι της μηδενικής υπόθεσης H_0 , με αποτέλεσμα, να δοθεί μεγαλύτερη έμφαση στον έλεγχο και στη μελέτη του Σφάλματος Τύπου I. Σχετικά με τη διαδικασία του ελέγχου της σημαντικότητας της H_0 έχει ασκηθεί κριτική ήδη από τη δεκαετία του 60. Τα τελευταία χρόνια και ιδιαίτερα μετά τις εργασίες του Jacob Cohen, σχετικά με την ανάλυση ισχύος των στατιστικών ελέγχων στις Επιστήμες της Συμπεριφοράς, η προσοχή των ερευνητών αρχίζει να στρέφεται κυρίως στον έλεγχο του Σφάλματος Τύπου II και, επομένως, στην αναγκαιότητα ανάλυσης της ισχύος (γ) των στατιστικών ελέγχων. Στην παρούσα εργασία περιγράφεται συνοπτικά ο ρόλος του «μεγέθους του αποτελέσματος» (*Effect Size-ES*) το οποίο μαζί με το επίπεδο σημαντικότητας α , την ισχύ γ και το μέγεθος του δείγματος (n), αποτελούν ένα κλειστό σύστημα ($\alpha\gamma nES$) όπου η γνώση ή ο προκαθορισμός τριών από τα στοιχεία του καθορίζει το τέταρτο. Το γεγονός αυτό θέτει τις βάσεις για μια περαιτέρω έρευνα στο πρόβλημα των στατιστικών ελέγχων.

ΕΙΣΑΓΩΓΗ

Κατά παράδοση, ο έλεγχος υποθέσεων στην επιστημονική έρευνα έχει να επιδείξει μια σαφή προτίμηση στη χρησιμοποίηση της στατιστικής σημαντικότητας ως κριτηρίου απόρριψης ή όχι της μηδενικής υπόθεσης H_0 (Τσάντας κ.ά., 1999), με αποτέλεσμα να δοθεί μεγαλύτερη έμφαση στον έλεγχο και στη μελέτη του Σφάλματος Τύπου I. Όμως, τα τελευταία χρόνια και ιδιαίτερα μετά τις εργασίες του Cohen (1962, 1965), σχετικά με την ανάλυση ισχύος των στατιστικών ελέγχων στις Επιστήμες της Συμπεριφοράς, η προσοχή των ερευνητών αρχίζει να στρέφεται και

στον έλεγχο του Σφάλματος Τύπου II καθώς και στην αναγκαιότητα Ανάλυσης της Ισχύος των στατιστικών ελέγχων (Cohen 1988, Murphy και Myers 1998).

ΜΕΘΟΔΟΛΟΓΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ ΣΤΟΥΣ ΕΛΕΓΧΟΥΣ ΥΠΟΘΕΣΕΩΝ

Υπάρχουν διάφορες μεθοδολογικές προσεγγίσεις σχετικά με τον έλεγχο σημαντικότητας της H_0 (βλέπε Huck, 2000). Αναφέρουμε συνοπτικά μερικές:

A) Η κλασική μέθοδος των έξι βημάτων. Σύμφωνα με την κλασική μέθοδο σε ένα στατιστικό έλεγχο θα πρέπει να ακολουθούνται αυστηρά τουλάχιστον τα παρακάτω ιεραρχικά ταξινομημένα έξι βήματα:

1. Σαφής καθορισμός και διατύπωση της μηδενικής υπόθεσης H_0 .
2. Σαφής καθορισμός και διατύπωση της εναλλακτικής υπόθεσης H_1 .
3. Προκαθορισμός του επιπέδου σημαντικότητας α .
4. Εφαρμογή της κατάλληλης για την περίπτωση μεθόδου συλλογής και σύνοψης των δειγματικών ή πειραματικών δεδομένων. Ο όρος σύνοψη αναφέρεται στον υπολογισμό μέσω όρων, ποσοστών, διασπορών, συντελεστών συσχέτισης, κ.ά.
5. Καθορισμός ενός κριτηρίου για την αξιολόγηση της δειγματικής μαρτυρίας (καθορισμός της απορριπτικής περιοχής του ελέγχου).
6. Απόφαση για την απόρριψη ή όχι της μηδενικής υπόθεσης.

Είναι σημαντικό να τονιστεί ότι η H_0 και η H_a θα πρέπει να καθορίζονται εξ αρχής και πριν από τη συλλογή και επεξεργασία των δεδομένων (Huck 2000, Pagano και Gauvreau 2000, Zar 1996). Σε αντίθετη περίπτωση ο έλεγχος δεν έχει καμία λογική βάση.

B) Η μέθοδος των επτά βημάτων. Η διαδικασία περιλαμβάνει:

Τα βήματα 1 έως 6 είναι ίδια με αυτά της κλασικής μεθόδου.

7^a) Υπολογισμός ενός δείκτη που να εκφράζει την ένταση ή το βαθμό της σχέσης μεταξύ των μεταβλητών (π.χ. r , r^2 , η^2 , ω^2 , ϕ , κ.ά.).

7^b) Εκτίμηση του μεγέθους του αποτελέσματος (*effect size-ES*).

7^γ) Υπολογισμό της ισχύος γ του στατιστικού ελέγχου (*Post Hoc Power Analysis*).

Γ) Η μέθοδος των εννέα βημάτων. Η διαδικασία περιλαμβάνει:

Τα βήματα 1 έως 3 είναι ίδια με αυτά της κλασικής μεθόδου και επιπλέον:

- 4) Καθορισμό του μεγέθους του αποτελέσματος *ES*.
- 5) Καθορισμό του επίπεδου της ισχύος γ του στατιστικού ελέγχου (*A priori Power Analysis*).
- 6) Καθορισμό του απαιτούμενου μεγέθους δείγματος ή δειγμάτων.

Ακολουθούν τα βήματα 7, 8 και 9 τα οποία είναι ίδια με τα βήματα 4, 5 και 6 της κλασικής μεθόδου αντίστοιχα

Δ) Η προσέγγιση του ελέγχου της στατιστικής σημαντικότητας. Η διαδικασία περιλαμβάνει:

- 1) Σαφή καθορισμό και διατύπωση της μηδενικής υπόθεσης H_0 .
- 2) Απόφαση σχετικά με το εάν ο έλεγχος θα είναι μονόπλευρος ή δίπλευρος.
- 3) Εφαρμογή της κατάλληλης για την περίπτωση μεθόδου συλλογής και σύνοψης των δεδομένων. Υπολογισμός του στατιστικού του ελέγχου.
- 4) Καθορισμό της πιθανότητας που σχετίζεται με τη δειγματική μαρτυρία κάτω από την ισχύ της μηδενικής υπόθεσης (*p-value*).
- 5) Απόφαση σχετικά με το εάν η τιμή p της παρατηρούμενης στάθμης σημαντικότητας αποτελεί αρκετή ένδειξη για την απόφαση να απορριφθεί η μηδενική υπόθεση.

Στην προσέγγιση αυτή η στάθμη σημαντικότητας δεν προκαθορίζεται από τον ερευνητή. Τα ίδια τα δεδομένα μαρτυρούν το κατά πόσο είναι συμβατά με τη μηδενική υπόθεση με δεδομένη την ισχύ (ορθότητα) της μηδενικής υπόθεσης.

Ε) Η υβριδική προσέγγιση στον έλεγχο της μηδενικής υπόθεσης

Η υβριδική προσέγγιση αποτελεί συνδυασμό της κλασικής μεθόδου και του ελέγχου της στατιστικής σημαντικότητας. Πιο συγκεκριμένα, εφαρμόζεται η διαδικασία των 6, 7, ή 9 βημάτων και στο τέλος ο ερευνητής παρουσιάζει την τιμή p με σκοπό να δηλώσει τη μαρτυρία για το κατά πόσο τα δεδομένα υποστηρίζουν ή όχι την H_0 . Αποτελεί την πιο συχνά χρησιμοποιούμενη προσέγγιση. Ιδιαίτερα δημοφιλής είναι η υβριδική η οποία προκύπτει από το συνδυασμό της κλασικής μεθόδου των έξι βημάτων και του ελέγχου της στατιστικής σημαντικότητας.

ΚΡΙΤΙΚΗ ΣΤΟΥΣ ΕΛΕΓΧΟΥΣ ΣΗΜΑΝΤΙΚΟΤΗΤΑΣ ΤΗΣ H_0

Σχετικά με τη διαδικασία ελέγχου της σημαντικότητας της H_0 (Null-Hypothesis Significance-Test Procedure-NHSTP) έχει ασκηθεί κριτική ήδη από τη δεκαετία του 60 η οποία και επαναλαμβάνεται περιοδικά από διάφορους επικριτές (Yates 1951, Kish 1959, Rozeboom 1960, Bakan 1966, Morrison και Henkel 1970, Parkhurst 1985, Pratt 1976, Cox 1977, Chatfield 1991, Loftus 1991, Carver 1978, Oakes 1986, Yocczuz 1991, Schmidt 1996, Guttman 1985). Μπορούμε να συνοψίσουμε τα βασικά επιχειρήματα κατά της NHSTP ως εξής:

- Η στατιστική σημαντικότητα ενός αποτελέσματος μπορεί να είναι αποτέλεσμα της κατάλληλης επιλογής του μεγέθους του δείγματος και του επιπέδου σημαντικότητας α .
- Η H_0 δε μπορεί ποτέ να είναι αληθινή.
- Από τη στατιστική σημαντικότητα δε μπορούμε να εξάγουμε συμπεράσματα για την αντίστροφη πιθανότητα της υπόθεσης, δηλαδή την πιθανότητα η H_0 να είναι αληθής με βάση τα δεδομένα.
- Η στατιστική σημαντικότητα δε δίνει πληροφορίες σχετικά με τις τιμές των παραμέτρων του ή των πληθυσμών.
- Ο έλεγχος του σφάλματος Τύπου II παραμελείται αδικαιολόγητα.
- Από τη στατιστική σημαντικότητα δε μπορεί να εξαχθεί κάποιο συμπέρασμα σχετικά με την πρακτική ή κλινική σημαντικότητα ενός αποτελέσματος.
- Η δυαδική λογική της NHSTP (η H_0 απορρίπτεται ή όχι) δε συμβαδίζει με το γεγονός ότι η γνώση αποχτιέται βήμα προς βήμα.
- Η διαδικασία εγκυμονεί κινδύνους για στοχαστικά και λογικά σφάλματα καθώς και για παρανοήσεις (Μενεξές και Οικονόμου, 2001).

Όμως, στις περισσότερες περιπτώσεις, η κριτική που έχει ασκηθεί στηρίζεται σε λόγους που δεν αφορούν στην ίδια τη στατιστική διαδικασία αλλά κυρίως στο γεγονός ότι οι λανθασμένες αντιλήψεις των ερευνητών και ο στοχαστικός αναλφαβητισμός είναι οι παράγοντες που οδηγούν σε εσφαλμένη χρήση και ερμηνεία των αποτελεσμάτων των στατιστικών ελέγχων σημαντικότητας της H_0 .

ΑΝΑΛΥΣΗ ΙΣΧΥΟΣ (Α.Ι.)

Ισχύς γ ενός στατιστικού ελέγχου είναι η πιθανότητα να απορρίψουμε μία όντως λανθασμένη H_0 και καθορίζεται από το β (πιθανότητα να διαπράξουμε Σφάλμα Τύπου II): $\gamma = 1 - \beta = P(\text{απόρριψης της } H_0 / H_0 \text{ λανθασμένη})$. Η Α.Ι. πραγματοποιείται, συνήθως, κατά τη διάρκεια σχεδιασμού της έρευνας ή του πειράματος, δηλαδή πριν ακόμα από τη συλλογή των δεδομένων (a priori), και χρησιμοποιείται για την εκτίμηση της πιθανότητας να απορρίψουμε μια όντως λανθασμένη H_0 . Με άλλα λόγια, με την Α.Ι. προσπαθούμε να προσεγγίσουμε το βαθμό εμπιστοσύνης που θα αποδώσουμε στην «ικανότητα» του ελέγχου να αναδείξει όντως στατιστικά σημαντικά αποτελέσματα. Η ισχύς ενός στατιστικού ελέγχου γ εξαρτάται κυρίως από τρεις παράγοντες (βλέπε Cohen 1988, Murphy και Myers 1998):

- α) Το επίπεδο σημαντικότητας α (πιθανότητα να διαπράξουμε Σφάλμα Τύπου I).
- β) Το μέγεθος του δείγματος n .
- γ) Το μέγεθος ή την 'ένταση' του αποτελέσματος (*effect size-ES*). Το μέγεθος του αποτελέσματος μπορεί να οριστεί γενικά ως η έκταση ή η ένταση του υπό εξέταση φαινομένου (Cohen και Cohen, 1983). Αποτελεί ένα μέτρο του βαθμού στον οποίο ένα φαινόμενο πραγματοποιείται (Cohen, 1965). Από μια άλλη σκοπιά, το *ES* μπορεί να θεωρηθεί ως ο βαθμός απόκλισης του παρατηρούμενου αποτελέσματος από την H_0 (Kramer and Rosental, 1999). Σε κάθε στατιστικό έλεγχο αντιστοιχεί και διαφορετικό *ES* και μπορεί να μετρηθεί με δύο τρόπους (Cohen 1988, Kramer and Rosental 1999, Murphy και Myers 1998): α) Ως Διαφορά, τυποποιημένη ή μη (π.χ. Cohen's d , Hedges's g , Glass's δ), ή β) Ως Συσχέτιση ή Συνάφεια (π.χ. r , r^2 , η^2 , ω^2 , ϕ).

Οι τρεις παραπάνω παράγοντες μαζί με την ισχύ αποτελούν ένα κλειστό σύστημα (το οποίο προτείνουμε να ονομαστεί «σύστημα α - γ - n -*ES*»), με την έννοια ότι εάν τρία από τα στοιχεία του συστήματος είναι γνωστά και σταθερά τότε το τέταρτο μπορεί να καθοριστεί πλήρως (Cohen και Cohen, 1983). Σκοπός της Α.Ι. είναι να εξισορροπήσει κατάλληλα τους 4 παράγοντες του συστήματος λαμβάνοντας υπόψη τόσο τους θεωρητικούς όσο και τους πρακτικούς στόχους της έρευνας σε συνδυασμό

με τους πόρους (οικονομικούς, χρονικούς και τεχνολογικούς) που έχει στη διάθεσή του ο ερευνητής. Η εξισορρόπηση αυτή δε θα πρέπει να αντιτίθεται στους ηθικούς-δεοντολογικούς περιορισμούς της έρευνας. Ο Πίνακας 1 (βλέπε τέλος) παρουσιάζει παραδείγματα απλών στατιστικών ελέγχων, τα αντίστοιχα Τυποποιημένα ES και τις αντίστοιχες τιμές με βάση τις οποίες το ES μπορεί να χαρακτηριστεί, κατά σύμβαση, ως «μικρό», «μέτριο» και «μεγάλο» για την έρευνα στις Επιστήμες της Συμπεριφοράς (βλέπε Cohen, 1988).

ΣΥΜΠΕΡΑΣΜΑΤΑ-ΠΡΟΤΑΣΕΙΣ

Σε πρακτικό επίπεδο η Α.Ι., μεταξύ άλλων, μπορεί να απαντήσει στα δύο παρακάτω βασικά ερωτήματα: α) Ποιο είναι το ελάχιστο μέγεθος δείγματος n ώστε σε επίπεδο σημαντικότητας α και για επίπεδο ισχύος γ ο στατιστικός έλεγχος που θα εφαρμοστεί να διαγνώσει ως στατιστικά σημαντικό ένα ES d ; Στην περίπτωση αυτή το d , π.χ. 0,20, αποτελεί μια εκτίμηση του μικρότερου ES που έχει πρακτική ή κλινική σημαντικότητα για τον ερευνητή και έχει αξία να ανιχνευθεί ως στατιστικά σημαντικό. β) Δοθέντος του μεγέθους δείγματος n , του επιπέδου σημαντικότητας α και του παρατηρούμενου ES , ποια είναι ισχύς γ του στατιστικού ελέγχου; Όμως, ένας έμπειρος στατιστικός αναλυτής, ο οποίος κάνει κακή χρήση της Στατιστικής (χωρίς α - γ - n -ές προθέσεις), μπορεί να σχεδιάσει μια έρευνα με τρόπο ώστε να εξισορροπήσει κατάλληλα τα Σφάλματα Τύπου I και Τύπου II, με αποτέλεσμα, μετά από κατάλληλη επιλογή των α , β και του μεγέθους δείγματος, να προσανατολίσει τα συμπεράσματα προς ορισμένες «επιθυμητές» κατευθύνσεις.

Οι παραπάνω διαπιστώσεις σε συνδυασμό με τα επιχειρήματα αυτών που ασκούν κριτική στη διαδικασία της NHSTP θέτουν τις βάσεις για μια περαιτέρω έρευνα στο πρόβλημα των υποθέσεων των στατιστικών ελέγχων. Για παράδειγμα: Ένα αποτέλεσμα μη στατιστικά σημαντικό είναι πιο ενδιαφέρον από ένα αντίστοιχο στατιστικά σημαντικό σε περίπτωση που το μέγεθος του δείγματος και το παρατηρούμενο ES είναι αρκετά μεγάλα (άλλωστε τα μη στατιστικά σημαντικά αποτελέσματα σπάνια δημοσιεύονται). Στο πλαίσιο αυτό, προτείνουμε: α) A priori ή/και Post Hoc Ανάλυση Ισχύος, β) Ενεργή συμμετοχή των ερευνητών για τον

καθορισμό του ES και γ) Αποφυγή του στατιστικού υπερκαταναλωτισμού (βλέπε p -value) και τετριμμένων συμπερασμάτων του τύπου "*Η Γη γυρίζει, ($p < 0,05$)*".

ABSTRACT

Traditionally, data collected in a research study is submitted to a significance test to assess the viability of the null hypothesis. A power analysis, usually executed when the study is being planned, is used to anticipate the likelihood that the study will yield a significant effect. In this paper we briefly demonstrate the role of *Effect Size-ES* in power analysis. Effect Size together with alpha (the probability of committing Error Type I), sample size and power, form a closed system - once any three are established, the fourth is completely determined. The goal of a power analysis is to find an appropriate balance among these factors by taking into account the substantive goals of the study, and the resources available to the researcher. But the usage of the criterion of statistical significance for the rejection or not of the null-hypothesis has been the target of criticism since the early sixties and is still the target of periodical criticisms. These and some other issues lay the foundation for an epistemological revision of information, construed as knowledge, which results from the statistical hypotheses testing.

ΑΝΑΦΟΡΕΣ

- Bakan, D., (1966): The test of significance in psychological research. *Psychological Bulletin*, 66, 423-437.
- Carver, P., (1978): The case against statistical testing. *Harvard Educational Review*, 48, 378-399.
- Chatfield, C., (1991): Avoiding statistical pitfalls. *Statistical Science*, 6, 240-268.
- Cohen, J., (1962): The Statistical Power of Abnormal-Social Psychological Research: A Review, *Journal of Abnormal and Social Psychology*, 65, 145-153.
- Cohen, J., (1965): Some Statistical Issues in Psychological Research, in B. B. Wolman (Ed.), *Handbook of Clinical Psychology*. McGraw-Hill, New York, p.p. 95-121.
- Cohen, J., (1988): *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates Inc., New Jersey.
- Cohen, J., and Cohen, P., (1983): *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates Inc., New Jersey.

- Cox, R., (1977): The role of significance tests. *Scandinavian Journal of Statistics*, 4, 49-70.
- Guttman, L., (1985): The illogic of statistical inference for cumulative science. *Applied Stochastic Models and Data Analysis*, 1, 3-10.
- Huck, S., (2000): *Reading Statistics and Research*. Addison Wesley Longman, Inc.
- Kish, L., (1959): Some statistical problems in research design. *American Sociological Review*, 24, 328-338.
- Kramer, S., and Rosental, R., (1999): Effect Sizes and Significance Levels in Small-Sample Research, in R. Hoyle (Ed.), *Statistical Strategies for Small Sample Research*. Sage Publications Inc., Thousand Oakes.
- Loftus, R., (1991): On the tyranny of hypothesis testing in the social sciences. *Contemporary Psychology*, 36, 102-105.
- Μενεξές, Γ., και Οικονόμου, Α., (2001): Σφάλματα και Παρανοήσεις στους Στατιστικούς Ελέγχους Υποθέσεων: Υπέρβαση μέσω της Ανάλυσης Δεδομένων. *Τετράδια Ανάλυσης Δεδομένων* (έγινε δεκτή προς δημοσίευση).
- Morrison, E., and Henkel, E., (1970): Significance tests in behavioral research: Skeptical conclusions and beyond, in D. E. Morrison and R. E. Henkel (Ed.), *The Significance Test Controversy---A Reader*. Aldine, Chicago.
- Murphy, K., and Myors, B., (1998): *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*. Lawrence Erlbaum Associates Inc., New Jersey.
- Oakes, M., (1986): *Statistical Inference: A Commentary for the Social and Behavioral Sciences*. John Wiley & Sons Inc., Chichester.
- Pagano, M. and Gauvreau, K., (2000): *Αρχές Βιοστατιστικής*. Εκδόσεις ΕΛΛΗΝ, Περιστέρι-Αθήνα.
- Parkhurst, F., (1985): Interpreting failure to reject a null hypothesis. *Bulletin of the Ecological Society of America*, 66, 301-302.
- Pratt, W., (1976): A discussion of the question: for what use are tests of hypotheses and tests of significance. *Communications in Statistics, Series A* 5, 779-787.

- Rozeboom, W., (1960): The fallacy of the null-hypothesis significance test. *Psychological Bulletin*, 57, 416-428.
- Schmidt, L., (1996): Statistical significance testing and cumulative knowledge in psychology: implications for training of researchers. *Psychological Methods*, 1(2), 115-129.
- Τσάντας, Ν., Μωϋσιάδης, Χ., Μπαγιάτης, Ν., και Χατζηπαντελής, Θ., (1999): *Ανάλυση Δεδομένων με τη βοήθεια Στατιστικών Πακέτων*. Εκδόσεις ΖΗΤΗ, Θεσσαλονίκη.
- Yates, F., (1951): The influence of Statistical Methods for Research Workers on the development of the science of statistics. *Journal of the American Statistical Association*, 46, 19-34.
- Yoccz, G., (1991): Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bulletin of the Ecological Society of America*, 72, 106-111.
- Zar, J., (1996): *Biostatistical Analysis*. Prentice-Hall International, Inc.

ΠΑΡΑΡΤΗΜΑ

Ειδικό Λογισμικό για Ανάλυση Ισχύος

Στις παρακάτω διευθύνσεις μπορεί κανείς να βρει ειδικό λογισμικό για ανάλυση ισχύος των στατιστικών ελέγχων. Το πρόγραμμα Gpower είναι freeware.

- Gpower <http://www.psych.uni-duesseldorf.de/aap/projects/gpower/>
- Power and Precision (from Dataxiom, Inc., <http://www.dataxiom.com/>)
- Sample Power (from SPSS Inc., <http://www.spss.com/>)
- PASS (NCSS Statistical Software, <http://www.ncss.com/>)
- Statistica (from StatSoft Inc., <http://www.statsoft.com>)
- On line: <http://calculators.stat.ucla.edu/powercalc/>
<http://www.math.yorku.ca/SCS/Demos/power/>,
<http://home.clara.net/sisa/sampshlp.htm>, <http://www.mp1-pwrc.usgs.gov/powcase/powlinks.html>.

Πίνακας 1 Αντιστοιχία Στατιστικών Ελέγχων και Τυποποιημένων ES. Συμβάσεις κατά Cohen

Στατιστικός έλεγχος	Effect Size-ES	Συμβάσεις κατά Cohen
Τεστ για τη μέση τιμή μ ενός πληθυσμού Έλεγχος μονόπλευρος ή δίπλευρος	$d = M_0 - M_1 / \text{Std. D.}$ όπου: M_0 : η μέση τιμή του πληθυσμού M_1 : η μέση τιμή του δείγματος Std. D.: η τυπική απόκλιση του δείγματος	Θεωρητικά: $d \in [0, +\infty)$ Στην πράξη: $d = 0,20$ (μικρό) $d = 0,50$ (μέτριο) $d = 0,80$ (μεγάλο)
Τεστ για τη διαφορά δύο μέσων τιμών Δείγματα ανεξάρτητα Έλεγχος μονόπλευρος ή δίπλευρος	$d = M_1 - M_2 / \text{Std. D.}$ όπου: M_1 : η μέση τιμή του 1ου δείγματος (ομάδας) M_2 : η μέση τιμή του 2ου δείγματος (ομάδας) Std. D.: η κοινή ανάμεσα στα δείγματα (ομάδες) τυπική απόκλιση	Θεωρητικά: $d \in [0, +\infty)$ Στην πράξη: $d = 0,20$ (μικρό) $d = 0,50$ (μέτριο) $d = 0,80$ (μεγάλο)
Τεστ για τη διαφορά δύο μέσων τιμών Δείγματα εξαρτημένα Έλεγχος μονόπλευρος ή δίπλευρος	$d = M_1 - M_2 / \text{Std. D.}$ όπου: M_1 : η μέση τιμή του 1ου δείγματος (ομάδας) M_2 : η μέση τιμή του 2ου δείγματος (ομάδας) Std. D.: η τυπική απόκλιση των διαφορών	Θεωρητικά: $d \in [0, +\infty)$ Στην πράξη: $d = 0,20$ (μικρό) $d = 0,50$ (μέτριο) $d = 0,80$ (μεγάλο)
Τεστ για την αναλογία p ενός πληθυσμού Έλεγχος μονόπλευρος ή δίπλευρος	$d = p_0 - p_1 $ όπου: p_0 : αναλογία (ή %) στον πληθυσμό p_1 : αναλογία στο δείγμα	Θεωρητικά: $d \in [0, 1)$ Στην πράξη: $d : (0,60 \text{ vs } 0,50)$ (μικρό) $d : (0,74 \text{ vs } 0,50)$ (μέτριο) $d : (0,86 \text{ vs } 0,50)$ (μεγάλο)
Τεστ για τη σύγκριση δύο αναλογιών p_1 και p_2 Δείγματα ανεξάρτητα Έλεγχος μονόπλευρος ή δίπλευρος	$d = p_1 - p_2 $ όπου: p_1 : αναλογία στο 1ο δείγμα p_2 : αναλογία στο 2ο δείγμα	Θεωρητικά: $d \in [0, 1)$ Στην πράξη: $d : (0,40 \text{ vs } 0,50)$ (μικρό) $d : (0,40 \text{ vs } 0,65)$ (μέτριο) $d : (0,40 \text{ vs } 0,78)$ (μεγάλο)