

Πόσο έγκυρα είναι τα συμπεράσματα όταν η ανάλυση των πειραματικών δεδομένων περιορίζεται απλά σε έλεγχο θεωρητικών υποθέσεων με χρήση απλών στατιστικών συναρτήσεων;

Γιώργος Σ. Ιωαννίδης

Εργαστήριο Θετικών Επιστημών, Π.Τ.Δ.Ε., Πανεπιστήμιο Πατρών
gsioanni@upatras.gr - ioannidis_gs@hotmail.com

Περίληψη

Σκοπός αυτής της μελέτης είναι η διερεύνηση της εγκυρότητας των τελικών συμπερασμάτων σε διάφορες εργασίες διδακτικής θετικών επιστημών. Το κύριο πρόβλημα εντοπίζεται στην απλή χρήση στατιστικών μεθόδων στην ανάλυση των δεδομένων, με πρακτικές που είναι απλές ως προς την εφαρμογή, ενώ καταλήγουν σε συμπεράσματα συχνά απλοϊκά και μη υποστηριζόμενα από τις μετρήσεις. Ενώ η επαναλαμβανόμενη χρήση αποκλειστικά στατιστικών εργαλείων και μεθόδων είναι πραγματικά απλή αριθμητικά, παραγνωρίζει την ύπαρξη συστηματικών (μη-στατιστικών) πειραματικών σφαλμάτων μέτρησης, ο μη συνυπολογισμός των οποίων είναι αιτία ικανή να οδηγήσει στην εξαγωγή εσφαλμένων συμπερασμάτων. Ερμηνεύεται έτσι η μη επαναληψιμότητα των πειραμάτων και τα όποια αμοιβαία αντικρουόμενα συμπεράσματα, ενώ δίνονται και ορισμένες βελτιωτικές προτάσεις.

Abstract

The aim of the present study is the investigation of the validity of various results in education research, due to the way these have been derived from the data taken. The use of simple statistical tools for data analysis is seen as the main culprit, leading to results that are not supported by the quality of the data taken. The repeated use of purely statistical tools, albeit simple in execution as it really is, ignores the presence of systematic (non-statistical) measurement errors. It is precisely this failure during data analysis that very often leads to erroneous results. The non-repeatability of various experiments is thus explained, while some suggestions are offered to improve the situation.

1. Εισαγωγή

Η φύση του γενικότερου προβλήματος: η μη επαναληψιμότητα των διαφόρων πειραματικών συμπερασμάτων

Τα τελευταία έτη έχει αρχίσει να γίνεται αντιληπτό ότι ένα μεγάλο μέρος από τις ερευνητικές μετρήσεις (σε διάφορα γνωστικά πεδία), πολλές από τις οποίες δημοσιεύονται, δεν είναι δυνατόν να επιβεβαιωθούν από άλλα πειράματα που ενώ μετρούν τα ίδια πράγματα το κάνουν με διαφορετικό τρόπο. Έτσι τίθεται θέμα εγκυρότητάς τους. Συγκεκριμένα, διαδοχικά πειράματα καταλήγουν σε αντίθετα συμπεράσματα στα ίδια ερωτήματα, προκαλώντας πλήρη σύγχυση. Όπως αναφέρει και ο J.P.A. Ioannidis (2005), «τα περισσότερα ερευνητικά συμπεράσματα είναι τελείως εσφαλμένα, για τους περισσότερους ερευνητικούς σχεδιασμούς και για τα περισσότερα γνωστικά πεδία». Συμπερασματικά, «η πιθανότητα για κάποιο από τα συμπεράσματα μιας έρευνας να είναι ορθό μπορεί να εξαρτάται από το πόσο ισχυρή είναι η μελέτη

όπως και τα συστηματικά σφάλματα, τον αριθμό των άλλων μελετών που έχουν διεξαχθεί πάνω στο ίδιο θέμα, και, το σημαντικότερο, το λόγο των επαληθευμένων προς τους μη επαληθευμένους συσχετισμούς από αυτούς που διερευνήθηκαν, για κάθε επιστημονικό πεδίο». Τίθεται, λοιπόν, θέμα επαναληψιμότητας (και άρα εγκυρότητας) των διαφόρων πειραματικών συμπερασμάτων.

Σε ερευνητικά πεδία όπως η μέτρηση της εκπαιδευτικής αποτελεσματικότητας στην διδασκαλία Θετικών επιστημών, το θέμα περιπλέκεται περεταίρω. Ενώ οι μέθοδοι ανάλυσης των δεδομένων είναι παρόμοιες με αυτές της ιατρικής και φαρμακολογίας λόγω χάριν, και υπάρχει επιπλέον και η ομοιότητα του ότι έχουν να κάνουν με ανθρώπους και υφίστανται πάντα οι ατομικές διαφορές μεταξύ τους, υπάρχουν και διαφορές. Πρέπει, επίσης, να συνεκτιμηθεί το γεγονός ότι τα αποτελέσματα κάθε μορφής αλλαγής στην διδασκαλία παίρνουν πολύ περισσότερο χρόνο ώστε να πιστοποιηθούν από ότι λ.χ. ένα φάρμακο όπου οι ερευνητές μετρούν την πορεία ίασης των ασθενών (- ή αντίθετα να διαπιστώνουν άμεσα και με τραγικό τρόπο τα αποτελέσματα). Έτσι, λόγω ακριβώς της μη αμεσότητας στην επιβεβαίωση των αποτελεσμάτων μέσω πολλαπλών ελέγχων γνώσης (σε διάφορες ηλικίες των ίδιων μαθητών ή και σε άλλους μαθητές ίδιας ηλικίας), είναι ακόμα πιο σημαντικό να είμαστε σίγουροι για τα αποτελέσματα των όποιων εκπαιδευτικών παρεμβάσεων κάνουμε.

Το ότι υπάρχει, λοιπόν, πρόβλημα πρέπει να θεωρείται ως δεδομένο, τουλάχιστον δυνητικά – καθότι λίγοι ασχολούνται με την επιβεβαίωση αποτελεσμάτων άλλων, ή και προηγούμενων δικών τους. Πάντως, ανεξάρτητα με το είδος της εκπαιδευτικής παρέμβασης που μελετάται, ελάχιστες είναι οι περιπτώσεις μη θετικών αποτελεσμάτων – ακριβώς όπως και στα προαναφερθέντα ερευνητικά πεδία. Η παρούσα μελέτη ασχολείται με τις διάφορες αιτίες του γενικότερου αυτού φαινομένου, με έμφαση όμως στις εφαρμογές στην διδασκαλία Θετικών Επιστημών.

Ενώ οι αιτίες αυτής της μη επαναληψιμότητας των αποτελεσμάτων είναι σύνθετες (και κάθε φορά διαφορετικές στην λεπτομέρεια), ο πρώτος και κύριος λόγος εντοπίζεται στην μη σωστή αντιμετώπιση των εκάστοτε συστηματικών σφαλμάτων κάθε μέτρησης, είτε αυτά εμπίπτουν στην κατηγορία των bias που ίσως και να μπορούν να αντιμετωπιστούν, είτε εμπίπτουν στην κατηγορία του τελικού «παραμένοντος συστηματικού σφάλματος» οπότε απλά πρέπει να συνυπολογίζονται στην διαμόρφωση του εκάστοτε συνολικού σφάλματος, επηρεάζοντας όμως και τα συμπεράσματα, όπως είναι φυσικό. Η πρώτη, λοιπόν, αιτία έχει ως βάση την πλημμελή εφαρμογή μεθοδολογιών μέτρησης, όπως λ.χ. αυτές των Φυσικών επιστημών.

Η δεύτερη αιτία του φαινομένου της μη επαναληψιμότητας των αποτελεσμάτων μπορεί να αναζητηθεί στην απλή χρήση στατιστικών μεθόδων και συναρτήσεων στην επεξεργασία των δεδομένων, πρακτικές που βασικό πλεονέκτημα έχουν το ότι είναι συνολικά απλές ως προς την εφαρμογή τους. Σε συνδυασμό με την προαναφερθείσα αιτία (δηλαδή της μη ορθής αντιμετώπισης των εκάστοτε συστηματικών σφαλμάτων) καταλήγουν, έτσι, σε συμπεράσματα συχνά απλοϊκά, που δεν υποστηρίζονται από τις μετρήσεις. Η δεύτερη, άρα αιτία έχει ως βάση την πλημμελή χρήση μεθόδων πραγματικής στατιστικής επαγωγής, θεωρίας πιθανοτήτων και θεωρίας αποφάσεων, γενικά δηλαδή των Μαθηματικών.

2. Η έννοια του «σφάλματος» και η διαφοροποίησή του από το «λάθος»

Τα ακόλουθα αφορούν μετρήσεις κάθε μορφής και άρα επηρεάζουν έρευνες κάθε ειδικότητας, είτε αυτές αναφέρονται στις ανθρωπιστικές επιστήμες είτε στις θετικές,

είτε στις εφαρμογές τους, ισχύουν δε γενικά όποτε γίνεται κάποια μέτρηση (ή παρατήρηση) οποιουδήποτε μεγέθους με οποιονδήποτε τρόπο και αν έχει ληφθεί (λ.χ. είτε ερωτηματολόγιο είτε βολτόμετρο).

Κάθε μέτρηση ενός οποιουδήποτε μεγέθους χαρακτηρίζεται από μια αβεβαιότητα που ονομάζουμε σφάλμα. Οι έννοιες «σφάλμα» και «λάθος» είναι τελείως διαφορετικές και δεν πρέπει να συγχέονται μεταξύ τους. Τα σφάλματα μέτρησης (ή παρατήρησης) μπορούν να χαρακτηριστούν και να κατηγοριοποιηθούν. Συγκεκριμένα, μπορούν να διαχωριστούν: (α) στα συστηματικά σφάλματα και (β) στα τυχαία (ή στατιστικά) σφάλματα.

Οι δύο αυτές κατηγορίες σφαλμάτων, μπορούν να υπολογιστούν και στη συνέχεια να αθροιστούν, ώστε να προκύψει το τελικό σφάλμα μέτρησης. Όπως είναι λογικά αναμενόμενο, το άθροισμα αυτό γίνεται διανυσματικά, καθότι οι δύο κατηγορίες είναι τελείως ανεξάρτητες μεταξύ τους. Εδώ ισχύει το ότι «ανεξάρτητες ποσότητες σημαίνει κάθετα διανύσματα», ακριβώς όπως στην διανυσματική πρόσθεση ταχυτήτων σε κάθετους άξονες, ή στον υπολογισμό της σύνθετης αντίστασης RLC, στον ηλεκτρισμό. Έτσι προκύπτει το εκάστοτε τελικό σφάλμα μέτρησης, το οποίο και διαφέρει από μέτρηση σε μέτρηση (ακόμα και για την ίδια έρευνα), διότι εξαρτάται από το εκάστοτε μετρούμενο μέγεθος. Το τελικό σφάλμα που υπολογίζεται είναι διαφορετικό για κάθε σημείο διαγράμματος, και για οποιοδήποτε πειραματικό διάγραμμα ή ιστογράμμο. Αυτή η διαφοροποίηση των τελικών σφαλμάτων μεταξύ τους επηρεάζει άρα (α) κάθε σύγκριση μεταξύ διαφόρων μετρήσεων στο ίδιο διάγραμμα ή και (β) κάθε σύγκριση μεταξύ διαφορετικών διαγραμμάτων. Ερευνητικά, αυτό είναι ταυτόσημο με το να πούμε ότι η διαφοροποίηση επηρεάζει την σύγκριση μεταξύ μετρήσεων διαφόρων δειγμάτων ή και καταστάσεων. Έτσι, για να μπορούμε να εκφέρουμε γνώμη για το αν μια διδασκαλία πέτυχε ή όχι (λ.χ. πόσο καλή ήταν η πειραματική εκπαιδευτική παρέμβαση), χρειάζεται να υπολογιστεί το συνολικό σφάλμα για κάθε μια από τις μετρήσεις που έγιναν – δηλαδή για κάθε σημείο κάθε διαγράμματος. Άλλωστε, σε κανένα αποτέλεσμα δεν μπορεί να προσδοθεί οποιαδήποτε αξία, αν δεν υποστηρίζεται από μία αντικειμενική εκτίμηση του συνολικού σφάλματος μέτρησης, για κάθε τι που συγκρίνεται. Αυτό συμβαίνει διότι το αν υπάρχει ή όχι διακριτή διαφορά μεταξύ των σημείων που συγκρίνονται, θα εξαρτηθεί από το πόσο μεγάλο είναι το συνολικό πειραματικό σφάλμα για τα συγκεκριμένα σημεία που συγκρίνονται. Τέτοιες μετρήσεις αντιστοιχούν στο όποιο μετρούμενο μέγεθος όπως λ.χ. σε γνώμη-άποψη μαθητή σε σχέση με κάποιο Φυσικό φαινόμενο. Όλα τα σφάλματα εξαρτώνται από την φύση αυτού που μετριέται, αλλά διαφέρουν ακόμα και μεταξύ σημείων του ίδιου ιστογράμματος με διαφορετικό αριθμό μαθητών που επέλεξαν αυτήν την απάντηση.

3. Τα συστηματικά σφάλματα

Τα συστηματικά σφάλματα (systematic errors) είναι μη τυχαία σφάλματα, που όμως περιορίζουν την αξιοπιστία μιας μέτρησης. Θα πρέπει να τονιστεί ότι τα συστηματικά σφάλματα δεν μειώνονται καθόλου όταν αυξηθεί το πλήθος των μετρήσεων (του δείγματος). Πηγάζουν από διάφορες αιτίες, και επειδή καμία πειραματική μέθοδος μέτρησης ή παρατήρησης δεν μπορεί να αποφύγει τα συστηματικά σφάλματα, καμία έρευνα οποιασδήποτε μορφής δεν έχει μηδενικά συστηματικά. Αντιθέτως, τα συστηματικά σφάλματα είναι πανταχού παρόντα σε όλες τις μετρήσεις και (παρά την κοινή πρακτική σε ορισμένους κλάδους) δεν πρέπει να αποφεύγεται ο προσδιορισμός τους, όσο δύσκολο και αν είναι αυτό. Ο μη προσδιορισμός τους ισοδυναμεί, πρακτικά, με σιωπηλή παραδοχή ενός δήθεν μηδενισμού τους (- κάτι που είναι σίγουρα λάθος).

Τα συστηματικά, είτε είναι (α) τύπου bias, είτε (β) απλά συνεισφέρουν στο συνολικό σφάλμα ως “παραμένοντα συστηματικά σφάλματα” στα αριθμητικά αποτελέσματα, ακόμα και μετά τις όποιες τυχόν αναγωγές και αφαιρέσεις, μειώνοντας έτσι την εγκυρότητα των μετρήσεων και άρα των τελικών αποτελεσμάτων. Σημειώνεται εδώ ότι αν βρεθεί κάποιος (πειραματικά ορθός) τρόπος προσδιορισμού των τύπου bias συστηματικών σφαλμάτων αυτά μπορεί να αφαιρεθούν από την πειραματική μέτρηση – και πρέπει να γίνει. Αν δεν υπάρχει τρόπος ούτε να μετρήσουμε ούτε καν να εκτιμήσουμε το bias (λ.χ. δεν γνωρίζουμε αν είναι προς το να αυξάνει ή να ελαττώνει την μέτρησή μας) τότε, επειδή ακριβώς δεν είναι πρακτικά δυνατή η οποιαδήποτε αριθμητική παρέμβαση (με αναγωγή των μετρήσεων), θα πρέπει οπωσδήποτε να αυξηθεί η εκτίμηση του “παραμένοντος συστηματικού σφάλματος”, κατά την ανάλυση των δεδομένων.

Έχουν γίνει αρκετές προσπάθειες για την ταξινόμηση των συστηματικών σφαλμάτων σε σχέση με την προέλευσή τους (και η ταξινόμηση αυτή είναι ανεξάρτητη του αν είναι προς μία κατεύθυνση –δηλαδή bias- ή όχι). Αυτό επειδή πολλά από τα συστηματικά μπορούν και να προβλεφθούν και ίσως και να αντιμετωπιστούν (ή τελικά απλά να εκτιμηθούν ακριβέστερα), αν ταξινομηθούν και κατηγοριοποιηθούν σύμφωνα με κάποια κοινά χαρακτηριστικά, αν τους αποδοθεί ένας κατάλληλος ορισμός και αν γίνει μία περιγραφή των κατάλληλων προληπτικών μέτρων (Sackett, 1979). Η σύγχρονη προσέγγιση είναι η ταξινόμηση σε 3 κατηγορίες: (α) στα συστηματικά σφάλματα επιλογής (selection biases), (β) στα συστηματικά σφάλματα πληροφορίας (information biases) και (γ) στους συγχυτές, (confounders) – (βλέπε Grimes et al. 2002, ενώ για την μετάφραση του όρου βλέπε Mellou & Sparos 2005).

Το συστηματικά σφάλματα δεν εξορκίζονται με το να αγνοούνται στην ανάλυση δεδομένων, αντίθετα ενυπάρχουν, και συχνά ανατρέπουν τα τελικά αποτελέσματα. Έτσι, κατά τη διάρκεια κάθε πειράματος, είναι γνωστό ότι κατά πάσα πιθανότητα υποπίπτουμε σε κάποια συστηματικά σφάλματα. Ενώ αυτό είναι βέβαιο, το ζητούμενο είναι να εκτιμηθεί ποσοτικά η τάξη μεγέθους του σφάλματος που εμπεριέχεται. Όμως, αυτό καθίσταται δυνατόν μόνο αν συγκριθεί η μέθοδος που χρησιμοποιήθηκε με κάποια άλλη που θεωρείται «πρότυπο». Αυτό αποτελεί τον μόνο τρόπο ώστε να ποσοτικοποιηθούν τα συστηματικά σφάλματα.

Χαρακτηριστικό παράδειγμα διάπραξης συστηματικού σφάλματος είναι το εξής: Επιχειρείται η μέτρηση ενός μήκους (λ.χ. 8 μέτρα) με ένα χάρακα 1m. Αν ο συγκεκριμένος χάρακας που θα χρησιμοποιηθεί δεν έχει μήκος ακριβώς 1 μέτρο αλλά –λόγω κατασκευαστικών ατελειών- είναι κάπως «ελαττωματικός» με συνέπεια το πραγματικό του μήκος να μην είναι 100cm, αλλά 99,8cm για παράδειγμα, τότε το υπό μέτρηση μέγεθος (μήκος) δεν είναι $8 \times 100\text{cm} = 800\text{cm}$, αλλά $8 \times 99,8\text{cm} = 798,4\text{cm}$. Επιπλέον, μπορεί ο χάρακας να χρησιμοποιείται (ή να διαβάζεται) με ανακριβή τρόπο, κάτι που αποτελεί άλλη αιτία που προκαλεί τη γέννηση του συστηματικού σφάλματος. Άλλα παραδείγματα ακριβέστερων «πρότυπων» οργάνων (στην Φυσική) αποτελούν τα μαγνητόμετρα nuclear magnetic resonance NMR (αντί αυτών φαινομένου Hall) ή επιταχυνσιόμετρα ύψιστης ακρίβειας (αντί των απλών micro-electro-mechanical systems MEMS), τα οποία χρησιμοποιούνται για την εκ των υστέρων βαθμονόμηση των απλούστερων οργάνων. Άπαξ και εντοπιστεί οποιασδήποτε κατηγορίας και προέλευσης συστηματικό σφάλμα και υπολογιστεί αριθμητικά το μέγεθός του, μπορεί (και πρέπει) να αφαιρεθεί (με σωστό πρόσημο) από τις μετρήσεις, οι οποίες θα παρουσιαστούν απαλλαγμένες από το (όπως τότε ονομάζεται) bias. Η αναγωγή, με σκοπό την αφαίρεση του bias εκτελείται με απλή αναγωγή (αναλογικά) ή ακόμα και με χρήση εξειδικευμένων τεχνικών προσομοίωσης με H/Y - δηλαδή με τεχνικές “Monte

Carlo simulation” (για μία εισαγωγή βλέπε: https://en.wikipedia.org/wiki/Monte_Carlo_method), η ανάλυση των οποίων, όμως, υπερβαίνει την παρούσα μελέτη.

Εξειδικευμένο (αλλά κοινότατο) παράδειγμα συστηματικού σφάλματος τύπου bias είναι, όσον αφορά ιατρική και φαρμακολογία, το φαινόμενο placebo (δηλαδή του εικονικού φαρμάκου που λόγω αυθυποβολής προκαλεί κάποια ίαση). Το, κατά κάποιον τρόπο, αντίστοιχο φαινόμενο στην διδασκαλία Θετικών Επιστημών είναι η (όποια) βελτίωση μάθησης που προκαλείται λόγω της αύξησης του ενδιαφέροντος των μαθητών, λόγω του διάφορου της διδασκαλίας (π.χ. νέοι δάσκαλοι, νέα μέσα, νέοι μέθοδοι) και όπως στην ιατρική είναι πάντοτε προς θετική κατεύθυνση. Και ενώ μεν σε ενδελεχείς ερευνητικές μετρήσεις πάντα χρησιμοποιείται και δεύτερο δείγμα μαθητών (για έλεγχο) όπου η διδασκαλία γίνεται χωρίς τις νέες (πειραματικές) μεθόδους διδασκαλίας, αυτό δεν ακολουθείται πάντα. Επιπροσθέτως, πόσο όμοια είναι τα δύο δείγματα, ενώ μήπως υπάρχει σε αυξημένο βαθμό τέτοιο φαινόμενο “placebo” στην πειραματική ομάδα, σε σχέση με την μονάδα ελέγχου? Η όποια νέα πειραματική μέθοδος (συντά με H/Y) είναι, άλλωστε, αναμενόμενο να δημιουργεί πρόσθετο ενδιαφέρον στους μαθητές.

Υπάρχουν όμως και συστηματικά σφάλματα τα οποία δεν είναι δυνατόν να υπολογιστούν από μία ακριβέστερη μέτρηση, δηλαδή με κάποια ακριβέστερη διαδικασία. Λόγου χάρι, στο προηγούμενο παράδειγμα μέτρησης μήκους, για να μετρήσουμε την ακρίβεια του χάρακα, πρέπει να χρησιμοποιηθεί ένας ακριβέστερος, ενώ ταυτόχρονα η μέτρηση πρέπει να γίνει με τρόπο που να είναι ιδιαίτερα ακριβής και χρονοβόρος. Τι γίνεται όμως αν δεν υπάρχει τέτοιος ακριβέστερος χάρακας; Ακόμα και αν υπάρχει κάποιο πρότυπο, πόσο σίγουροι μπορεί να είμαστε (τελικά) για την ακρίβεια αυτού του νέου πρότυπου; Σε εκπαιδευτική έρευνα, πόσο αντιπροσωπευτικό είναι το δείγμα των παιδιών στο συγκεκριμένο σχολείο που χρησιμοποιήθηκε ερευνητικά, ή επίσης όποιοι οι δάσκαλοι των προηγούμενων τάξεων σε σχέση με τους υπολοίπους στο σχολείο; Και το ίδιο το σχολείο, πόσο αντιπροσωπευτικό είναι σε σχέση με άλλα σχολεία όλης της χώρας, και πώς μπορεί να ποσοτικοποιηθεί κάτι τέτοιο; Είναι καλύτερο από τον μέσο όρο ή ίσως χειρότερο; Το ποσοστό των παιδιών που δεν χρησιμοποιούν Ελληνικά στο σπίτι είναι και αυτό κοντά στον μέσο όρο της χώρας; Τέτοια σφάλματα που πηγάζουν από τις διαδικασίες επιλογής δείγματος είναι συστηματικά σφάλματα επιλογής. Παρ όλα ταύτα, αυτού του τύπου τα σφάλματα (λόγω δείγματος) αποτελούν απλά και μόνο μια από τις πηγές συστηματικών σφαλμάτων – υπάρχουν και πολλές άλλες. Λόγου χάρι, πόσο σίγουροι είμαστε για την ορθότητα της διαδικασίας μέτρησης που χρησιμοποιήθηκε; Πόσο ίδιοι μεταξύ τους είναι οι δάσκαλοι ερευνητές, αν το δείγμα περιλάμβανε πολλά σχολεία και πολλούς δασκάλους; Η πόσο αντιπροσωπευτικοί είναι αυτοί οι δάσκαλοι σε σχέση με τον μέσο όρο δασκάλων της χώρας; Αυτή ακριβώς η έλλειψη εμπιστοσύνης για το τελικό “μέτρο” με το οποίο συγκρίνουμε, όπως και για την διαδικασία που ακολουθήθηκε, αποτελεί το μέτρο του τελικού «παραμένουστος συστηματικού σφάλματος» για την μέτρηση αυτή, και μπορεί να προέρχεται από πολλές πηγές. Για πιο σύνθετες μετρήσεις (λ.χ. ένα οποιοδήποτε σύνθετο ερωτηματολόγιο που ελέγχει γνώσεις μαθητών) αυτό το παραμένον συστηματικό σφάλμα είναι ακόμα μεγαλύτερο, ακριβώς επειδή η μέτρηση είναι πιο σύνθετη. Λόγου χάρι, πόσο ξεκάθαρα αντιληπτή είναι κάθε ερώτηση από τους μαθητές; Είναι όλες εξίσου δύσκολα αντιληπτές (σε περίπτωση εσωτερικής σύγκρισης των μαθητικών απαντήσεων);

Συμπερασματικά, η συμβολή των συστηματικών σφαλμάτων στις μετρήσεις που γίνονται στα πλαίσια ερευνών στις ανθρωπιστικές επιστήμες είναι, στην πραγματικότητα, ιδιαίτερα υψηλές (ενίοτε ακόμα και καθοριστικές) στην διαμόρφωση

του εκάστοτε τελικού σφάλματος μέτρησης. Υπενθυμίζεται ότι αν το δείγμα είναι πραγματικά μεγάλο, ακριβώς επειδή μειώνονται τα στατιστικά, αυξάνεται (κατ' αναλογία) η συμβολή των συστηματικών στην διαμόρφωση του συνολικού σφάλματος. Ενώ, λοιπόν, ο ερευνητής νομίζει ότι μπορεί να είναι σίγουρος για α συμπεράσματά του, συμβαίνει ακριβώς το αντίθετο.

4. Τα τυχαία ή στατιστικά σφάλματα

Τα τυχαία (ή στατιστικά) σφάλματα (Random Errors or Statistical Errors), είναι σφάλματα που υπεισέρχονται σε κάθε μέτρηση και την επηρεάζουν με τυχαίο τρόπο και επομένως δεν μπορούν να αφαιρεθούν κατά την επεξεργασία τους. Εάν η μέτρηση αυτή επαναληφθεί, τότε οι τιμές που λαμβάνονται ως αποτέλεσμα της μέτρησης μπορεί να διαφέρουν κατά τι. Λόγω της τυχαίας προέλευσης που τα διακρίνει, (και σε πλήρη αντίθεση με τα συστηματικά) τα τυχαία σφάλματα μπορούν να μειωθούν με αύξηση του αριθμού των μετρήσεων – είναι δηλαδή στατιστικά αντιμετωπίσιμα, από όπου και προέρχεται το δεύτερο όνομά τους Πρόκειται για σφάλματα που σχετίζονται με την ακρίβεια μιας μέτρησης και δείχνουν τις διακυμάνσεις που έχουν οι μετρήσεις της πειραματικής διαδικασίας που γίνονται κάτω από τις ίδιες φαινομενικά συνθήκες και οδηγούν στην κατανομή των αποτελεσμάτων γύρω από μια μέση τιμή. Τα τυχαία σφάλματα συχνά (αλλά όχι πάντα) παρουσιάζουν μια κανονική κατανομή Gauss. Στις περιπτώσεις που η κατανομή είναι πραγματικά “κανονική” Gaussian, είναι εφικτό να χρησιμοποιηθούν απλές στατιστικές μέθοδοι για την μελέτη των τυχαίων σφαλμάτων. Παραδείγματα προέλευσης τέτοιων τυχαίων σφαλμάτων αποτελούν λ.χ. οι όποιες διαφορές μεταξύ των μαθητών σε μία σχολική μέτρηση (που θεωρούνται ως ενιαίο σύνολο σε μία σχολική τάξη, ενώ δεν είναι), ή η ύπαρξη κάποιας τυχαίας απόσπασης της προσοχής του μαθητή την στιγμή που απαντούσε στην συγκεκριμένη ερώτηση, ή η κούραση και φυσική κατάσταση του συγκεκριμένου μαθητή, μικροασθένεια αυτού μόνο του μαθητή, ή κάτι που είδε στην τηλεόραση τις μέρες πριν την μέτρηση. Αντίθετα, στο ίδιο παράδειγμα, αν όλοι οι μαθητές είχαν δει κάτι σχετικό στο σχολείο τους στα πλαίσια κάποιου άλλου μαθήματος, αυτό θα μετέτρεπε το σφάλμα σε (συστηματικό) bias και θα μπορούσε να επηρεάσει τις μετρήσεις είτε θετικά είτε αρνητικά σε σχέση με το όποιο μετρούμενο μέγεθος (αναλόγως του τι ειπώθηκε). Για όλους τους λόγους που προαναφέρθηκαν, τα τυχαία σφάλματα είναι, γενικά, αναπόφευκτα. Όπως άλλωστε προαναφέρθηκε, ούτε όλοι οι μαθητές είναι ίδιοι, ενώ η διδασκαλία στην τάξη είναι γενικά ενιαία, αλλά και ο πειραματικός τρόπος μέτρησης που χρησιμοποιείται αναφέρεται πάντα στο σύνολο των μαθητών.

Είναι, λοιπόν, φανερό ότι μια απλή παράθεση του μεγέθους N του δείγματος μιας έρευνας είναι και τελείως ανεπαρκής αλλά και (ακόμα χειρότερα) παραπλανητική, σε σχέση με την στάθμιση της συνολικής εγκυρότητας και συνολικής αξίας κάθε μορφής (ποσοτικής) έρευνας, ανεξαρτήτως γνωστικού πεδίου.

5. Τα εργαλεία μέτρησης της γνώσης που διαθέτουμε – ερωτηματολόγια (ανοιχτού και κλειστού τύπου) και συνεντεύξεις

Η εγκυρότητα (validity) ενός εργαλείου-ερωτηματολογίου είναι έννοια πραγματικά διαφορετική και μόνο έμμεσα σχετιζόμενη με αυτήν της πιστότητας (reliability). Ενώ η εγκυρότητα (validity) εξετάζει το κατά πόσον το εργαλείο μετράει αυτό ακριβώς που προσπαθεί να μετρήσει, η πιστότητα (reliability) μετρά το κατά πόσον το εργαλείο μετράει (κάτι οποιοδήποτε) με τρόπο συνεπή και επαναλήψιμο (Tavacol et al. 2008). Και ενώ μεν ένα εργαλείο δεν μπορεί να είναι (με πειραματικά αξιοποιήσιμο τρόπο) έγκυρο (valid) αν δεν είναι κατ' αρχή αρκετά πιστό (reliable), το αντίθετο δεν ισχύει

καθόλου, επειδή οι δύο έννοιες διαφέρουν τελείως, με το ένα (την πιστότητα) να αποτελεί (τελικά) αναγκαία αλλά όχι και ικανή συνθήκη ώστε να υπάρξει αρκετή πειραματική εγκυρότητα.

Συμπερασματικά, το αν έχει κάποιο ερωτηματολόγιο χρησιμοποιηθεί επιτυχώς στο παρελθόν δεν αποτελεί καν ένδειξη εγκυρότητάς του. Δεν υφίσταται η έννοια του “εκ των πρωτέρω” έγκυρου εργαλείου, ερωτηματολογίου, ή τύπου συνέντευξης με ερωτήσεις που έχουν προκαθοριστεί. Αντίθετα, η εγκυρότητα του κάθε ερωτηματολογίου πρέπει να αποδεικνύεται κάθε φορά εκ των υστέρων – δηλαδή μετά την εφαρμογή του με την λήψη των δεδομένων. Πρέπει άρα να αποτελεί ριζικό και αναπόσπαστο τμήμα της ανάλυσης των δεδομένων.

Γιατί; Ποια είναι η υποβόσκουσα αιτία; Η πραγματική διαφορά μεταξύ των εννοιών εγκυρότητας (validity) και πιστότητας (reliability) συνίσταται, από την σκοπιά ενός έμπειρου πειραματικού, στην ύπαρξη των συστηματικών σφαλμάτων τα οποία, όπως προαναφέρθηκε, αποτελούν ριζικά διαφορετικό τύπο σφαλμάτων από τα τυχαία (ή στατιστικά) σφάλματα και (εξ ορισμού) είναι ανεξάρτητα (και άρα δεν επηρεάζονται) από το πλήθος N του δείγματος.

Πάντως, ανεξάρτητα του σε ποια από τις δύο κατηγορίες συστηματικών ανήκουν (δηλαδή bias ή παραμένοντα συστηματικά), τα σφάλματα αυτά αποτελούν την πραγματική (και υποβόσκουσα) αιτία της ριζικής διαφοράς μεταξύ εγκυρότητας και πιστότητας. Αυτός ακριβώς είναι και ο λόγος που το αν κάποιο εργαλείο αποδείχθηκε ότι ήταν έγκυρο (valid) σε μία έρευνα (με τα όποια συστηματικά σφάλματα που αυτή περιείχε), δεν αποτελεί ούτε καν ένδειξη εγκυρότητας για μία επόμενη έρευνα, με τα δικά της συστηματικά (που γενικά είναι επόμενο να είναι πολύ διαφορετικά ως προς την φύση και το μέγεθός τους από την προηγούμενη).

6. Η φύση του προβλήματος κατά την ανάλυση των δεδομένων

Τα συστηματικά σφάλματα είναι εκ της φύσης τους πολύ δύσκολο να εκτιμηθούν. Δηλαδή να βρεθεί (α) η αιτία προέλευσής τους και (β) το μέγεθός τους. Επίσης κάθε μαθηματική πράξη που κάνουμε στα πρωταρχικά δεδομένα (raw data) έχει ως αποτέλεσμα την αλλαγή του μεγέθους των σφαλμάτων αυτών. Δηλαδή, κάθε διαίρεση πρόσθεση ή πολλαπλασιασμός «προωθεί» το σφάλμα της αρχικής μέτρησης και το διαφοροποιεί, - ποτέ όμως δεν το μειώνει στην ουσία, αντίθετα πάντα το αυξάνει. Η πραγματική αιτία της συνεχούς αύξησης του σφάλματος, πρέπει να αναζητηθεί στο δεύτερο θερμοδυναμικό αξίωμα (δηλαδή της συνεχούς αύξησης της Εντροπίας), όπως αυτό εφαρμόζεται στην Πληροφορική, κατά τον C. Shannon. Αυτή η διαφοροποίηση του σφάλματος (ώστε να φθάσουμε στο σφάλμα του τελικού πειραματικού αποτελέσματος που συνήθως παρουσιάζεται υπό μορφή κάποιου διαγράμματος) είναι δυνατόν να υπολογιστεί με την διαδικασία που ονομάζεται «διάδοση σφάλματος» ή “error propagation”. Ενώ όμως είναι δυνατόν να γίνει τέτοιος υπολογισμός, αυτό συμβαίνει πολύ σπάνια.

Διευκρινίζεται ότι ακόμα και ένας απλός υπολογισμός κάποιας συνάρτησης (από αυτές που συχνά χρησιμοποιούνται στη πειραματική ανάλυση – λ.χ. μίας στατιστικής συνάρτησης) εμπεριέχει την διάδοση σφαλμάτων που επηρεάζουν το αριθμητικό αποτέλεσμα της συνάρτησης, και ενώ αυτό θα πρέπει να συνυπολογίζεται στην εξαγωγή των επιμέρους (ή και τελικών) συμπερασμάτων, αυτό συμβαίνει και πάλι πάρα πολύ σπάνια (- ιδιαίτερα σπάνια).

Πάντως, ένας πολύ δοκιμασμένος τρόπος για να εκτιμήσει ο ερευνητής την ποιότητα των πειραματικών δεδομένων που έχει πάρει, είναι μέσω σχηματισμού διαφόρων γραφημάτων και διεξοδικής παρατήρησης τους, όπως και συσχετισμού μεταξύ των γραφημάτων αυτών.

7. Ο έλεγχος ερευνητικών υποθέσεων ως μέθοδος εξαγωγή γενικών αποτελεσμάτων

Συνεχίζεται, σε διάφορα άρθρα, η ανορθολογική χρήση διαφόρων απλών στατιστικών συναρτήσεων (μη Μπαεσιανής στατιστικής) όπως η “Kuder-Richardson Coefficient of reliability (K-R 20)”, για δυαδικά δεδομένα, ή όπως αυτή τροποποιήθηκε από τον Hoyt για εκλαϊκευθεί ως “Cronbach-alpha” για μη δυαδικά δεδομένα, δηλαδή αυτά που περιλαμβάνουν κάποια στάθμιση εκ μέρους του μαθητή (και όχι απλά ερωτήσεις σωστό/λάθος). Αυτά θεωρούνται ότι είναι κατάλληλα για τον έλεγχο ερωτηματολογίων. Τέτοιες συναρτήσεις μετράνε μόνο την “εσωτερική πιστότητα” (internal consistency) μιας συγκεκριμένης χρήσης (δοκιμής) ενός ερωτηματολογίου, σε ένα συγκεκριμένο δείγμα – μετράνε άρα μια όψη της πιστότητας (reliability) (Cortina J. 1993). Συγκεκριμένα μετριέται το πόσο πολύ οι απαντήσεις στις ερωτήσεις του ερωτηματολογίου συσχετίζονται μεταξύ τους (υπολογίζεται ανά ζεύγος αυτών) – οπότε θα μπορούσε να υποθεθεί ότι ίσως δεν είναι οι μαθητές που “παρασύρονται” αλλά ότι στην πραγματικότητα (μέχρι ένα βαθμό) οι ερωτήσεις μετρούν το ίδιο πράγμα (κάτι που μπορεί να γίνεται ακόμα και επίτηδες - ερευνητικά), ή απλά τους ερωτάται με ελαφρά διαφορετικό (και μη διακριτό από αυτούς) τρόπο, η ίδια πραγματικά ερώτηση – με στόχο τον εσωτερικό έλεγχο.

Επιπλέον η “εσωτερική πιστότητα” (internal consistency) δεν είναι καν ταυτόσημο με την πιστότητα (reliability). Οι δύο αυτές έννοιες, εκτός από την λεκτική τους διαφορά στην Αγγλική, είναι φανερό ότι διαφέρουν μεταξύ τους και από το ότι: (α) η “εσωτερική πιστότητα” ισχύει για συγκεκριμένη μέτρηση και μόνο, με συγκεκριμένο δείγμα παιδιών και συνθήκες μέτρησης, (β) ελέγχει τον αριθμητικό συσχετισμό των απαντήσεων, στο σύνολο των ερωτήσεων του ερωτηματολογίου, (γ) η τελική τιμή εξαρτάται και από τον αριθμό των ερωτήσεων όπως και το πόσο πλησίον βρίσκονται, εννοιολογικά και λεκτικά, μεταξύ τους. Γιατί, λοιπόν, να θεωρείται ότι δείχνει κατά πόσο το ερωτηματολόγιο είναι εργαλείο πιστό, δηλαδή μετράει με τρόπο συνεπή και επαναλήψιμο?

Συνήθως, ο απώτερος λόγος που υπολογίζονται τέτοιες συναρτήσεις είναι ως προετοιμασία για την ορθή εφαρμογή ενός ελέγχου κάποιας συγκεκριμένης ερευνητικής υπόθεσης (hypothesis testing). Τέτοιες διαδικασίες είναι κοινές σε κλάδους όπως κυρίως η ιατρική και ψυχολογία, διότι λ.χ. απαντούν στην απλή ερώτηση του αν μία (δοκιμαζόμενη) ουσία μπορεί να χρησιμοποιηθεί ως φάρμακο, ή (εναλλακτικά) κατά πόσο υπάρχει κάποιο κοινό υποβόσκον αίτιο που έχει ως αποτέλεσμα μερικά (ή και όλα) από όσα μετρούνται κατά την διάρκεια του πειράματος. Στην εκπαίδευση, αυτό που είναι υπό δοκιμή είναι (αντί για φάρμακο) μια νέας μορφής διδακτική παρέμβαση. Υπενθυμίζεται ότι ακόμα και στην περίπτωση της προαναφερθείσας ορθής χρήσης τους, οι συναρτήσεις πρέπει να υπολογίζονται πάντα ύστερα από κάθε εφαρμογή ερωτηματολογίου σε κάθε δείγμα (δηλαδή a posteriori), και όχι να θεωρούνται σαν (δήθεν) ορθά τα αριθμητικά αποτελέσματα που ελήφθησαν σε παλαιότερα πειράματα με χρήση του ιδίου ερωτηματολογίου (Streiner 2003). Είναι δηλαδή πληροφορίες για κάτι που συνέβη στο παρελθόν (αποκλειστικά), ενώ δεν παρέχουν τίποτα που μπορούμε να το χρησιμοποιήσουμε υποθετικά, ή ως πρόβλεψη.

Πολύ περισσότερο, αν όμως ο σκοπός της έρευνας είναι πιο σύνθετος όπως λ.χ. ερωτήματα κατανόησης ενός μεγάλου εύρους εννοιών στην διδασκαλία επιστημών, η τακτική “ελέγχου υποθέσεων” κρίνεται απλοϊκή (ως απλά δυαδική – λ.χ. αποτελεί φάρμακο ναι ή όχι), και άρα ο υπολογισμός τέτοιων συναρτήσεων είναι είτε περιττός και ακατάλληλος, είτε ακόμα και παραπλανητικός. Δεν είναι απλά ένα θέμα για το αν λ.χ. η χρήση μιας συσκευής ή ενός λογισμικού υποβοηθάει ή όχι την γνώση στους μαθητές. Αντίθετα, το πραγματικό ερώτημα είναι πόσο μεγάλη ήταν η κατανόηση σε κάθε ένα από τα θέματα της διδακτικής παρέμβασης, όπως και το πώς αυτό θα μπορούσε να βελτιστοποιηθεί στην πράξη. Η μελέτη των ιδεών των παιδιών και των (εκάστοτε) παρανοήσεων τους πριν και μετά την παρέμβαση είναι ο κύριος στόχος. Άλλωστε αυτός είναι ο γενικότερος και βαθύτερος στόχος της εκπαιδευτικής έρευνας. Επιπλέον, αυτό οδηγεί και στην ανάλυση κάθε ερώτησης (ερωτηματολογίου) ξεχωριστά, αλλά και κάθε παρανόησης ξεχωριστά από την άλλη. Η συστηματική μελέτη των απαντήσεων των παιδιών, κατάλληλα επεξεργασμένες, αποτελεί πάντα τον ερευνητικό στόχο.

Είναι λοιπόν φανερό ότι οι όποιες τιμές εξάγονται από μεθόδους σύγκρισης μεταξύ ομάδας ελέγχου και πειραματικής ομάδας του συνολικού δείγματος, ή οι όποιες τιμές παρέχονται από διάφορες στατιστικές συναρτήσεις (-υπάρχουν αρκετές), απέχουν κατά δύο φορές από το να μετρούν την εγκυρότητα. Πρώτον διότι η μετρούμενη “εσωτερική πιστότητα” (internal consistency) απέχει από την πιστότητα (reliability), και δεύτερον διότι και η reliability απέχει πολύ από την εγκυρότητα (validity). Επί παραδείγματι, όπως είναι γνωστό, η τιμή των προαναφερθέντων KR-20, και Cronbach-alpha αυξάνεται (αριθμητικά) όσο αυξάνεται ο αριθμός των ερωτήσεων του ερωτηματολογίου, ή ακόμα και όταν οι ερωτήσεις καλύπτουν ένα στενότερο φάσμα εννοιών που ερευνώνται (Schmitt N. 1996). Πώς, λοιπόν, μετά όλα αυτά, θα μπορούσαν να αποτελούν κριτήριο κάποιας υποτιθέμενης απόδειξης ορθότητας;

Άλλωστε, ο Gigerenzer (1991) αναρωτιέται “where are the error-bars in social-science journals” κάτω από την ενότητα “deficits of hypothesis testing”, τόσο στο άρθρο όσο επίσης και στο βιβλίο “The Empire of Chance” (Gigerenzer & Porter, 1990). Είναι, λοιπόν, φανερό ότι έχουν διατυπωθεί σοβαρές αμφιβολίες σε σχέση με την αλόγιστη εφαρμογή μεθόδων απλής στατιστικής επαγωγής (inference) σε κάθε μορφή ανάλυσης. Και στην Ελλάδα έχουν παρουσιαστεί οι δυσμενείς επιπτώσεις της παρουσίας πειραματικών αποτελεσμάτων χωρίς ενδείξεις για τα (συστηματικά και στατιστικά) σφάλματά τους (Ιωαννίδης, 2008). Πόσο δε μάλλον, η εξαγωγή ακόμα και συμπερασμάτων από τέτοιου τύπου μετρήσεις.

Επιπροσθέτως, για θέματα διδασκαλίας επιστημών, πρέπει να προβληματίζει οποιαδήποτε προσπάθεια γενικευμένης (και απλής) δυαδικής απάντησης του τύπου “είναι καλό ή όχι η χρήση του συγκεκριμένου εργαλείου”, αντί της πολύ πιο σύνθετης “πόσο είναι διδακτικά ωφέλιμη, και για διδασκαλία ποιων εννοιών, σε ποια ηλικία, και με ποια μέθοδο”. Δεν πρόκειται για φάρμακο.

8. Η εξαγωγή επιμέρους αποτελεσμάτων από την ανάλυση των δεδομένων

Η χρήση των διαφόρων στατιστικών συναρτήσεων ως μέθοδος ελέγχου των δεδομένων είναι, φυσικά, θεμιτή. Εδώ όμως πρέπει να τονιστεί ότι όλες αυτές οι συναρτήσεις δίνουν αριθμητικά αποτελέσματα που ισχύουν μόνο κάτω από συγκεκριμένες προϋποθέσεις, εκ των οποίων η βασικότερη είναι ότι η κατανομή σφάλματος στις μετρήσεις είναι Γκαουσιανή (Gaussian).

Ας πάρουμε ένα τελείως γενικευμένο απλό παράδειγμα, όπως την σύγκριση μεταξύ του ποσοστού “σωστών απαντήσεων” σε δύο δείγματα μαθητών Α και Β. Έστω ότι, αριθμητικά, το δείγμα Α έχει μικρότερο ποσοστό από αυτό του Β. Ο ερευνητής ρωτάει, πολύ απλά, πόσο βέβαιο είναι αυτό το αποτέλεσμα. Μπορεί δηλαδή, παρά τις μετρήσεις, να μην είναι δυνατή η διάκριση μεταξύ δειγμάτων Α και Β – ή ίσως και η πραγματικότητα να είναι αντίθετη; Για να απαντηθεί (αριθμητικά) το “αν και κατά πόσο μπορεί να γίνει διάκριση” μεταξύ Α και Β χρειάζεται να υποθέσουμε κάποια “κατανομή πιθανότητας” για τις μετρήσεις που έγιναν. Μια απλή απαρίθμηση των κοινά αναγνωρισμένων κατανομών πιθανότητας θα απαιτούσε πολλές σελίδες. Αν τώρα προστεθεί το ότι η σίγουρη ύπαρξη συστηματικών σφαλμάτων στις μετρήσεις διαφοροποιεί ακόμα και αυτές τις (μαθηματικά καθορισμένες) κατανομές πιθανοτήτων, το ερώτημα μετατίθεται στο “ποια κατανομή ισχύει” στην περίπτωση αυτή, ώστε να οδηγηθούμε σε αριθμητικό αποτέλεσμα. Επειδή η Gaussian κατανομή ισχύει για τυχαίες μεταβλητές και είναι σχετικά απλή υπολογιστικά, επιλέγεται αυτή (- ενώ προς στιγμή εξαιρούνται και τα συστηματικά σφάλματα που σίγουρα παραλλάσσουν την κατανομή και μάλιστα με τρόπο κατ’ αρχή άγνωστο). Έτσι εξάγεται ένα αριθμητικό αποτέλεσμα, το οποίο όμως δεν πρέπει να θεωρείται απόλυτο, προφανώς.

Αποτελεί αντικείμενο υποθετικής συζήτησης το αν τα συστηματικά σφάλματα πάνω στις μετρήσεις ελέγχου ποιότητας μαύρης Ιρλανδέζικης μπύρας (όπως αυτό για την πρωταρχική χρήση του “Student’s” t-test) είναι αμελητέα. Είναι πολύ πιθανό να είναι σχεδόν αμελητέα, μια και πρόκειται για γραμμή βιομηχανικής παραγωγής. Σίγουρα όμως δεν είναι αμελητέα για μετρήσεις που μελετούν την ανθρώπινη γνώση. Το ίδιο που ισχύει για απλά χ^2 -test, ισχύει και για διάφορα άλλα που χρησιμοποιούνται σε όποιες απλές αναλύσεις μετρήσεων. Η φύση και το μέγεθος των συστηματικών σφαλμάτων ξεπερνάει πάντα τέτοιες προσεγγίσεις, ακόμα και για ιδιαίτερα επιμελημένες μετρήσεις.

Ο J.P.A. Ioannidis (2005) αναφέρει ότι «πρέπει να αναγνωρίσουμε ότι ο έλεγχος στατιστικής σημαντικότητας σε μια έρευνα δίνει μια μερική εικόνα μόνο». Είναι σαφές ότι μία απλή στατιστική ανάλυση δεν είναι ικανή να δώσει απαντήσεις στο σύνθετο ερώτημα των συνολικών σφαλμάτων πειραματικής μέτρησης, που μάλιστα διαφέρουν στο μέγεθος για κάθε ένα από τα μετρούμενα σημεία. Αυτό, φυσικά, δεν σημαίνει ότι δεν έχουν νόημα οι στατιστικές αναλύσεις. Αντίθετα έχουν πολύ μεγάλο νόημα, αλλά δεν πρέπει να εξετάζονται απλοϊκά, αλλά στα πλαίσια μιας γενικότερης στατιστικής συνεπαγωγής (statistical inference) και της ευρύτερης επιστημονικής συνεπαγωγής.

9. Αποτελέσματα: Η αξιολόγηση των αποτελεσμάτων οδηγεί στην εξαγωγή ασφαλών συμπερασμάτων - ή μήπως όχι?

Τέτοιες συναρτήσεις όπως οι προαναφερόμενες είναι απλά στατιστικές στην φύση τους, ενώ οι πιο κοινές από αυτές προϋποθέτουν Gaussian κατανομές, κάτι ανέφικτο λόγω της συμφυούς ύπαρξης πολύ σημαντικών “παραμενόντων συστηματικών σφαλμάτων” (remaining systematic errors i.e. non-bias), κάτι εγγενές σε όλες τις μετρήσεις που σχετίζονται με γνώσεις – απλά και μόνο λόγω πολυπλοκότητας και ευρύτητας του ερευνώμενου αντικειμένου. Ως γνωστό, τα σφάλματα αυτά προστιθέμενα στα όποια στατιστικά τοιαύτα, καταλήγουν σε κατανομές που διαφέρουν πολύ από τις κανονικές κατανομές. Άρα, ακόμα και στις περιπτώσεις ορθής χρήσης των συναρτήσεων (δηλαδή απλά για έλεγχο υποθέσεων), η χρήση τους πρέπει να γίνεται με εξαιρετική προσοχή (Green & Thompson 2005), καθότι δεν υπάρχει τρόπος να ληφθούν υπόψη τα συστηματικά σφάλματα, εντός του όποιου υπολογισμού των

συναρτήσεων αυτών. Είναι πράγματι παράδοξο το ότι ενώ είναι πολύ σπάνιο να δημοσιεύονται έρευνες στις θετικές επιστήμες ή στην μηχανολογία χωρίς κάποια ειδική μελέτη των συστηματικών σφαλμάτων των πειραμάτων που διεξήχθησαν, στον χώρο των ανθρωπιστικών ερευνών γενικότερα, όπου τα συστηματικά σφάλματα είναι πολύ μεγαλύτερα στο μέγεθος, προτιμάται (με ελάχιστες φωτεινές εξαιρέσεις) η τακτική του να αγνοούνται επιδεικτικά. Συχνά ένας -κατά βάση απλοϊκός- στατιστικός έλεγχος σημαντικότητας χρησιμοποιείται, και ο έλεγχος υποθέσεων διεξάγεται χωρίς ορθή θεώρηση όλων των συστηματικών σφαλμάτων (διαφόρων τύπων και προέλευσης), και αυτό είναι από την φύση του βαθιά προβληματικό.

10. Συμπεράσματα - Προτάσεις

Συμπερασματικά, η διαδικασία της ανάλυσης δεδομένων για την εξαγωγή χρήσιμων πληροφοριών από πειράματα, δεν αρκεί να στηρίζεται απλά σε μια απλή στατιστική ανάλυση των δεδομένων. Μια ενδεδειγμένη ανάλυση πειραματικών δεδομένων συνήθως χρησιμοποιεί κάποιες στατιστικές μεθόδους αλλά πρέπει να επεκτείνεται ώστε να περιλάβει τα σφάλματα κάθε τύπου και κατηγορίας. Έτσι, η τελική πληροφορία που προκύπτει από την ανάλυση δεδομένων εμπεριέχει και τον βαθμό της εμπιστοσύνης την οποία πρέπει να δίνουμε στην πληροφορία. Μέσα στα όρια της εμπιστοσύνης αυτής, τα τελικά συμπεράσματα μπορούν να επιβεβαιωθούν μέσω οποιασδήποτε εναλλακτικής πειραματικής επιβεβαίωσης. Μόνο με αυτό τον τρόπο μπορούν να θεωρηθούν τα συμπεράσματα ως επιστημονικώς ορθά.

Απλή χρήση στατιστικών συναρτήσεων χρησιμοποιώντας πακέτα λογισμικού, όσο καλά και να είναι αυτά, απλά δεν αρκεί για να καταλήξουμε σε ορθά ερευνητικά συμπεράσματα. Δεν χρειάζεται να γίνει κανένα σχόλιο σε σχέση με πρακτικές απλής παράθεσης αριθμητικών αποτελεσμάτων σειράς διαφορετικών στατιστικών συναρτήσεων, για τις οποίες ούτε η μέθοδος υπολογισμού είναι γνωστή, αλλά ούτε και τηρούνται οι προϋποθέσεις κάτω από τις οποίες κάθε τέτοια συνάρτηση είναι έγκυρη και χρήσιμη.

Δεν υπάρχει, τελικά, καμία στατιστική ή άλλου τύπου αριθμητική συνάρτηση, και καμία αυτοματοποιημένη στατιστική ή άλλου τύπου διαδικασία ελέγχου οποιασδήποτε θεωρητικής υπόθεσης, που να μπορεί να λειτουργεί ως (απλό μεθοδολογικό) υποκατάστατο της επιστημονικής πειραματικής συνεπαγωγής (experimental inference). Σύμφωνα με αυτή γίνεται κάθε δυνατή διαδικασία ελαχιστοποίησης των συστηματικών σφαλμάτων, εκτιμώνται τα παραμένοντα συστηματικά, υπολογίζονται τα (τελείως ανεξάρτητα προς αυτά) στατιστικά σφάλματα, και από αυτά τα δύο υπολογίζονται τα συνολικά σφάλματα για κάθε μέτρηση, για να ακολουθήσει μετά ενδεδειγμένη επιστημονική σκέψη, η συνεπαγωγή πραγματικά έγκυρων επιστημονικών συμπερασμάτων.

11. Βιβλιογραφία

1. Ιωαννίδης Γ. Σ., «Ραβδογράμματα, Ψέματα, και ατυχείς Μεταφράσεις» ανακοίνωση ύστερα από πρόσκληση, στα πρακτικά του 12^{ου} Πανελληνίου Συνεδρίου Ένωσης Ελλήνων Φυσικών, Καβάλα 20-23 Μαρτίου 2008, ISBN 978-960-89657-3-7, σελ. 14-15
2. Cortina J. (1993). What is coefficient alpha: an examination of theory and applications. *Journal of applied psychology*, 78, 98-104.

3. Gigerenzer, G., & Porter, T. (1990). *The empire of chance: How probability changed science and everyday life* (Vol. 12). Cambridge University Press.
4. Gigerenzer, G., et al. (1991). On the tyranny of hypothesis testing in the social sciences. *Contemporary psychology*, 36(2), 102-105.
5. Green S. & Thompson M. (2005). Structural equation modelling in clinical psychology research In: Roberts M, Ilardi S, [Eds.] *Handbook of research in clinical psychology*, Oxford: Wiley-Blackwell, 138 - 175
6. Grimes, D. A., & Schulz, K. F. (2002). Bias and causal associations in observational research. *The Lancet*, 359(9302), 248-252.
7. Ioannidis, J.P.A. (2005). Why most published research findings are false. *PLoS Med*, 2(8), e124, 696-701.-(Retrieved from: <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124>)
8. Mellou, K., & Sparos, L. (2005). Systematic errors in etiological epidemiological studies, *Arch Hellen Med*, 22(2), 199-207.
9. Sackett, D.L. (1979). Bias in analytic research, *Journal of Chronic Diseases*, 32, 51 - 63
10. Schmitt N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8, 4, 350-353.
11. Streiner D.L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency, *Journal of personality assessment*, 80 (1), 99-103.
12. Tavakol M., Mohagheghi M.A., Dennick R. (2008). Assessing the skills of surgical residents using simulation, *Journal Surgical Education*, 65 (2), 77-83.