Γραπτή εξέταση των μαθητών και επιστημονική έρευνα

Κώστας Παγώνης, *Παιδαγωγικό Ινστιτούτο*

Περίληψη

Η παρούσα μελέτη, υποστηρίζοντας ότι η γραπτή εξέταση των μαθητών έχει ομοιότητες με την επιστημονική έρευνα, στηρίζεται στη διαπίστωση ότι τόσο η μία όσο και η άλλη παρουσιάζουν την ίδια σειρά από απαραίτητες φάσεις. Η τεκμηρίωση αυτού του ισχυρισμού γίνεται με αναφορά σε στατιστικές μεθόδους.

Εισαγωγή

Ο ισχυρισμός ότι η εξέταση των μαθητών έχει ομοιότητες με την επιστημονική έρευνα στηρίζεται στη διαπίστωση ότι τόσο η μία όσο και η άλλη παρουσιάζουν την ίδια σειρά από απαραίτητες φάσεις, οι οποίες απεικονίζονται στον παρακάτω πίνακα. Οι πρώτες τρεις από τις φάσεις λέγονται *επιτελικές*, ενώ οι τελευταίες τρεις *εκτελεστικές*. Ως προς την εξέταση των μαθητών, η υποδιαίρεση στις συγκεκριμένες παρακάτω φάσεις ταιριάζει κυρίως για γραπτή εξέταση, ενώ, ως προς την επιστημονική έρευνα, ταιριάζει κυρίως για ορισμένα είδη της, όπως είναι η έρευνα πεδίου, η μελέτη περίπτωσης (case study), κ.τ.λ. Οι προφορικές εξετάσεις διαφοροποιούνται από τις γραπτές (και είναι συνήθως δυσκολότερες για τον εκπαιδευτικό), επειδή σε αυτές συμπίπτουν χρονικά οι δύο πρώτες εκτελεστικές φάσεις, δηλαδή η τέταρτη και η πέμπτη του πίνακα I.

Ο κ. Κώστας Παγώνης είναι Εκπαιδευτικός Δ.Ε., Δρ Φυσικός, κάτοχος ενός Μεταπτυχιακού και δύο Μεταδιδακτορικών (P.D.R.), με οργανική θέση στο 6ο ΕΠΑΛ Αθήνας.

Πίνακας Ι
Οι φάσεις που διαιρούνται η εξέταση των μαθητών και η επιστημονική έρευνα

Φάσεις	Εξέταση μαθητών	Επιστημονική έρευνα
1η	Επιλογή της λειτουργίας της αξιολόγησης	Εντοπισμός του στόχου της έρευνας και διατύπωση των ερευνητικών υποθέσεων
2η	Κατάστρωση του προγράμματος της εξέτασης	Κατάστρωση του σχεδίου της έρευνας
3η	Επινόηση ή υιοθέτηση συγκεκριμένων “θεμάτων” εξέτασης δηλαδή ενός “paper”	Κατασκευή ή υιοθέτηση ενός οργάνου μέτρησης (π.χ. ερωτηματολόγιο, έντυπο παρατήρησης τάξης, κ.τ.λ.).
4η	Δόσιμο της εξέτασης	Συλλογή των δεδομένων
5η	Διόρθωση των γραπτών	Επεξεργασία των δεδομένων
6η	Αξιολόγηση της επίδοσης του κάθε μαθητή και εξαγωγή συμπερασμάτων αναφορικά με την τάξη και την υλοποίηση του προγράμματος	Ερμηνεία των δεδομένων και εξαγωγή συμπερασμάτων αναφορικά με τις ερευνητικές υποθέσεις

Παράλληλα με την ύπαρξη πολλών ομοιοτήτων, η εξέταση των μαθητών και η επιστημονική έρευνα παρουσιάζουν και ορισμένες διαφορές. Μία από αυτές είναι ότι, ενώ η επεξεργασία των δεδομένων μιας έρευνας γίνεται σχεδόν πάντα με εφαρμογή στατιστικών μεθόδων, κατά την εξέταση των μαθητών αυτό συμβαίνει σπανιότερα και χρονικά αργότερα, δηλαδή όταν έχει ολοκληρωθεί η διόρθωση και βαθμολογία των γραπτών.

Στατιστική επεξεργασία των αποτελεσμάτων της εξέτασης

Συνήθως ο καθηγητής (ή ένα ίδρυμα ή οργανισμός) θεωρεί ότι ολοκλήρωσε την εξέταση των μαθητών, όταν καταλήξει σε έναν κατάλογο με τα ονόματα και τους βαθμούς που πήραν οι μαθητές. Τόσο η συστηματική ανάλυση των λαθών όσο και η στατιστική επεξεργασία των αποτελεσμάτων αγνοούνται τελείως. Πρέπει, όμως, να θεωρούνται αναπόσπαστο τμήμα της εξέτασης και βρίσκονται στην έκτη και τελευταία φάση του πίνακα I, αφού επιτρέπουν την εξαγωγή συμπερασμάτων. Η στατιστική επεξεργασία των αποτελεσμάτων, ειδικότερα, εξυπηρετεί διάφορους παιδαγωγικούς σκοπούς και μπορεί να γίνει για να:

1) *Κατανοήσουμε* την αξία του βαθμού που εξασφάλισε ένας μαθητής:
   α) συγκριτικά με τη μέχρι εκείνη τη στιγμή επίδοσή του,
   β) συγκριτικά με την επίδοση όλων των συμμαθητών του κατά την ίδια εξέταση,
   γ) συγκριτικά με την επίδοση άλλων μαθητών, άλλης τάξης, που υποβλήθηκαν στην ίδια εξέταση,
   δ) συγκριτικά με τη μέση επίδοσή του στα άλλα μαθήματα, του προγράμματος σπουδών του κ.τ.λ.

2) *Εξακριβώσουμε*:
   α) το βαθμό επίτευξης των στόχων του προγράμματος από την τάξη,
   β) την αποδοτικότητα της μεθόδου διδασκαλίας,
   γ) την αποδοτικότητα του διδακτικού υλικού που χρησιμοποιήθηκε,
   δ) την καταλληλότητα του γνωστικού περιεχομένου της διδασκαλίας για τους συγκεκριμένους μαθητές.

3) *Ελέγξουμε*:
   α) την αξιοπιστία του τεστ,
   β) την εγκυρότητα της εξέτασης, κ.τ.λ.

4) *Ανιχνεύσουμε* διάφορα χαρακτηριστικά του δασκάλου:
   α) συγκρίνοντας, για παράδειγμα, τη μέση επίδοση των μαθητών του με αυτή των μαθητών άλλων τάξεων που είχαν άλλο δάσκαλο και που υποβλήθηκαν στο ίδιο τεστ,
   β) συγκρίνοντας τη διασπορά της βαθμολογίας (που αποκαλύπτει το βαθμό ανομοιογένειας ή ομοιογένειας μιας τάξης) των μαθητών του με εκείνη των μαθητών των τάξεων άλλων δασκάλων κ.τ.λ.

Για να πετύχουμε όλα τα παραπάνω δεν αρκεί πάντα η εκ των υστέρων στατιστική επεξεργασία των αποτελεσμάτων. Στις περιπτώσεις, για παράδειγμα, 2β και 2γ χρειάζεται σχεδιασμός και εκτέλεση πειραματικής έρευνας. Πρέπει, επομένως, ήδη από την πρώτη και δεύτερη φάση της εξέτασης (βλ. πίνακα Ι) να έχει προβλεφθεί τι είδους στατιστική θα γίνει και να έχει γίνει ο κατάλληλος σχεδιασμός και των άλλων εκτελεστικών φάσεων. Ένα πρώτο βήμα προς τη στατιστική επεξεργασία των αποτελεσμάτων (που είναι απαραίτητο, ιδιαίτερα όταν υπάρχει ένας σημαντικός αριθμός μαθητών που εξετάστηκε) γίνεται όταν, από την κατάσταση βαθμολογίας στην οποία καταλήγει σχεδόν πάντα ο εκπαιδευτικός, κατασκευαστεί ένας πίνακας “κατανομής συχνοτήτων” (frequency distribution). Η κατάσταση βαθμολογίας περιλαμβάνει συνήθως τα ονόματα των μαθητών τοποθετημένα σε αλφαβητική σειρά και το βαθμό που πήρε ο καθένας από αυτούς. Έχει επομένως την παρακάτω μορφή: Η παραπάνω κατάσταση, όταν γίνεται πίνακας κατανομής συχνοτήτων, παίρνει την παρακάτω μορφή του Πίνακα ΙΙΙ. Τα δύο κύρια πλεονεκτήματα του πίνακα κατανομής συχνοτήτων σε σύγκριση με την κατάσταση βαθμολογίας είναι: α) κάνει απρόσωπη τη βαθμολογία και μετατοπίζει το ενδιαφέρον μας από την επίδοση των μεμονωμένων μαθητών στην επίδοση της τάξης του μαθητικού κοινού, β) παρουσιάζει με τον οικονομικότερο τρόπο το αποτέλεσμα, αφού, στην προκειμένη περίπτωση, θα είχε την ίδια έκταση, ακόμη κι αν οι μαθητές δεν ήταν σαράντα οκτώ (48), αλλά εκατοντάδες ή χιλιάδες.

Πίνακας ΙΙ
Κατάσταση βαθμολογίας

Ονοματεπώνυμο	Βαθμός
Βαρέλας Αριστοτέλης	7
Βερυκούκης Γιάννης	7

Ονοματεπώνυμο	Βαθμός
Δημητριάδη Κατερίνα	6
κ.ο.κ.	...

Πίνακας ΙΙΙ
Πίνακας κατανομής συχνοτήτων

Βαθμός (χ)	Συχνότητα (f)	Αθροιστική συχνότητα (cf)	Αναλογία (%)
2	3	3	6,25
3	6	9	12,50
Σύνολα:	N=9		100,00

Περιγραφικοί στατιστικοί δείκτες

Ο συντομότερος και ακριβέστερος τρόπος υπολογισμού και κοινοποίησης των αποτελεσμάτων της στατιστικής ανάλυσης μιας βαθμολογίας είναι η χρησιμοποίηση «δεικτών». Ένας δείκτης είναι απλά ένα νούμερο που, για κάποιον που έχει γνώσεις στατιστικής, αποκαλύπτει σε ποιο σημείο της βαθμολογικής κλίμακας συγκεντρώνονται οι περισσότεροι βαθμοί, ή πόσο σκόρπιοι είναι ή τι είδους και πόση σχέση υπάρχει ανάμεσα σε μια βαθμολογία και μιαν άλλη, κ.τ.λ. Εδώ δε θα ασχοληθούμε με το "κ.τ.λ.", δηλαδή δε θα εξετάσουμε άλλου είδους δείκτες, εκτός από αυτούς που δείχνουν πού συγκεντρώνονται οι περισσότεροι βαθμοί και ονομάζονται δείκτες "κεντρικής τάσης", αυτούς που δείχνουν πόσο σκόρπιοι είναι οι βαθμοί και που ονομάζονται δείκτες "διασποράς" και, τέλος, στο επόμενο κεφάλαιο, αυτούς που δείχνουν την ποιότητα και ποσότητα μιας σχέσης και που ονομάζονται δείκτες "συνάφειας". Οι δείκτες κεντρικής τάσης (measures of central tendency), καθώς φανερώνουν όπως ειπώθηκε το σημείο της βαθμολογικής κλίμακας προς το οποίο τείνουν, συγκεντρώνονται, οι περισσότεροι βαθμοί, είναι χρήσιμοι για πολλούς λόγους και γι' αυτό χρησιμοποιούνται ευρέως. Είναι, επίσης, αυτοί που ευκολότερα υπολογίζονται, ακόμη και από όποιον δεν έχει γνώσεις στατιστικής. Οι δείκτες αυτής της κατηγορίας είναι τρεις και θα τους εξετάσουμε από τον λιγότερο "καλό" προς τον πιο "καλό". Το πόσο καλός είναι ένα δείκτης προσδιορίζεται από το πόσο εύκολα αλλάζει, όταν αλλάξουν μερικά δεδομένα, δηλαδή μερικοί βαθμοί. Όταν ένας δείκτης αλλάζει εύκολα, μας ξεγελάει, δε μας δίνει ικανοποιητική εικόνα της κεντρικής τάσης μιας κατανομής. Ο πρώτος δείκτης, που δε χρειάζεται καμία αριθμητική πράξη για να τον βρούμε, είναι η "συχνότερη" ή "δεσπόζουσα" ή "επικρατούσα" τιμή (mode). Αυτή συμβολίζεται διεθνώς με το "Mo" και είναι απλά ο βαθμός που συγκέντρωσε, στον πίνακα κατανομής συχνοτήτων τους περισσότερους μαθητές. Δηλαδή η βαθμίδα της κλίμακας στην οποία εμφανίστηκε μεγαλύτερη συχνότητα. Στο παράδειγμα βαθμολογίας που δόθηκε στο προηγούμενο κεφάλαιο το Mo = 6. Αν οι μαθητές που πήραν 7 ήταν δεκατρείς κι αυτοί που πήραν 6 ήταν δώδεκα, τότε θα είχαμε Mo=7. Επειδή ακριβώς η μικρή μετατόπιση ενός μόνο βαθμού αλλάζει κατά μία μονάδα την επικρατούσα τιμή, αυτή δε θεωρείται καλός δείκτης. Αυτό όμως δε σημαίνει ότι είναι και άχρηστος δείκτης. Υπάρχουν περιπτώσεις που μας είναι χρήσιμος, δηλαδή καταλληλότερος από αυτούς που θα εξετάσουμε παρακάτω. Όταν, τέλος, σε μία κατανομή τη μεγαλύτερη συχνότητα την έχουν, σε ίση ποσότητα, δύο διαφορετικοί βαθμοί, τότε η κατανομή ονομάζεται "διμοδική" (bimodal). Λίγο καλύτερος δείκτης είναι η “διάμεσος” ή διχοτόμος ή μέση τιμή (median). Αυτός ο δείκτης συμβολίζεται με το “Μ” και δεν είναι τίποτε άλλο από το βαθμό αριστερά, από τον οποίο τοποθετούνται οι μισοί βαθμοί της κατανομής και δεξιά οι άλλοι μισοί. Στο παράδειγμά μας το Μ=6. Ο χρησιμότερος και συνηθέστερος δείκτης κεντρικής τάσης είναι ο “αριθμητός μέσος” (arithmetic mean), δηλαδή ο γνωστός σε όλους “μέσος όρος”. Αυτός συμβολίζεται με το “$\bar{X}\bar{X}$” και υπολογίζεται με τον τύπο (1).

Το άθροισμα ΣX_i είναι το άθροισμα των βαθμών, και Ν, όπως ήδη φάνηκε στον πίνακα (ΙΙ) κατανομής συχνοτήτων του προηγούμενου κεφαλαίου, σημαίνει αριθμός βαθμών, δηλαδή μαθητών που εξετάστηκαν. Με βάση τα στοιχεία του παραδείγματος, το $\bar{x}$ = 5,375. Οι δείκτες διασποράς (measures of variability) που περισσότερο χρησιμοποιούνται στην εκπαίδευση είναι μόνο δύο, αλλά θα εξετάσουμε και τρεις άλλους, γιατί χρησιμεύουν για να γίνει κατανοητό τι είναι “τυπική απόκλιση”, που είναι ο “καλύτερος” από όλους τους δείκτες διασποράς. Ο πιο εύκολος να υπολογιστεί αλλά λιγότερο χρήσιμος δείκτης είναι το “εύρος” (range). Το εύρος βγαίνει αν αφαιρέσουμε το μικρότερο βαθμό από το μεγαλύτερο (αδιαφορώντας για τη συχνότητα που έχουν) και προσθέσουμε μια μονάδα. Στο παράδειγμά μας το εύρος της κατανομής είναι 8-2+1 = 7. Αν ένας από τους μαθητές που πήραν 8 είχε πάρει 10, τότε το εύρος θα ήταν 9. Η ευκολία με την οποία χαλάει το εύρος, κάνει αυτόν το δείκτη κατάλληλο μόνο στις περιπτώσεις που η κατανομή δεν έχει ακραίους βαθμούς “χασματικά” απομακρυσμένους από το σημείο της κεντρικής τάσης

$$\bar{X} = \frac{\sum_{i=1}^{i=N} X_i}{N} \qquad (1)$$

Για να εξουδετερώσουν τα μειονεκτήματα του εύρους, οι στατιστικολόγοι δημιούργησαν το δείκτη “τεταρτημοριακή διασπορά” (quartile deviation), που συμβολίζεται με το “Q” ή το “QD” και υπολογίζεται με τον τύπο:

$$Q = \frac{QP_3 - QP_1}{2} \qquad (2)$$

Τα σύμβολα QP_1 και QP_3 σημαίνουν πρώτο και τρίτο “τεταρτημοριακό σημείο”. Τεταρτημοριακό σημείο λέγεται το σημείο στο οποίο κόβεται μια κατανομή, αν θέλουμε να τη χωρίσουμε σε τέσσερα ίσα κομμάτια. Για παράδειγμα, έστω ότι έχουμε 12 μαθητές που εξετάστηκαν, και που πήραν τους παρακάτω βαθμούς: 1, 2, 3, 4, 5, 5, 5, 5, 6, 7, 8 και 9. Αν διαιρέσουμε αυτούς τους μαθητές με το τέσσερα, αφού σε τέσσερα κομμάτια θέλουμε να χωρίσουμε την κατανομή, βρίσκουμε ότι σε κάθε κομμάτι θα μπουν 3 μαθητές. Επομένως το πρώτο θα περιλαμβάνει αυτούς που πήραν 1 , 2 και 3, άρα το πρώτο τεταρτημοριακό σημείο βρίσκεται ανάμεσα στο 3 και το 4, δηλαδή είναι 3,5. Το δεύτερο κομμάτι περιλαμβάνει αυτούς που πήραν 4, 5 και 5, άρα το σημείο τομής, δηλαδή το δεύτερο τεταρτημοριακό σημείο είναι ανάμεσα στο 5 και το 5, δηλαδή 5,0. Το τρίτο κομμάτι περιλαμβάνει αυτούς που πήραν 5, 5 και 6, άρα το τρίτο τεταρτημοριακό σημείο βρίσκεται ανάμεσα στο 6 και το 7, δηλαδή είναι 6,5. Βάζοντας αυτά τα νούμερα στη θέση των σύμβολων του τύπου βρίσκουμε ότι γι' αυτή την κατανομή το Q=1,5.

Τεταρτημοριακά σημεία

Η τεταρτημοριακή διασπορά είναι πιο αξιόπιστος δείκτης από το εύρος, παρουσιάζει όμως το μειονέκτημα ότι δεν απεικονίζει τα ενδεχόμενα κενά σε μια κατανομή. Αν, για παράδειγμα, οι δώδεκα μαθητές αντί για την παραπάνω βαθμολογία είχαν την παρακάτω, αυτό δε θα φαινόταν στην τεταρτημοριακή διασπορά, η οποία θα έπαιρνε και πάλι τον τιμή 1,5: 0, 1, 3, 4, 5, 5, 5, 5, 6, 7, 9 και 10. Αυτό οδήγησε σε έναν άλλο δείκτη, που λέγεται "μέση απόκλιση" (mean deviation) και συμβολίζεται με το "MD". Σε αυτόν το δείκτη λαμβάνεται υπόψη ο αριθμητικός μέσος και ο χρησιμοποιούμενος τύπος είναι:

$$MD = \frac{\sum_{i=1}^{i=N} \left(X_i - \bar{X} \right)}{N} \qquad (3)$$

Αφού γνωρίζουμε ήδη όλα τα σύμβολα του τύπου, καταλαβαίνουμε ότι η μέση απόκλιση προκύπτει αν: α) βρούμε τον αριθμητικό μέσο μιας κατανομής β) αφαιρέσουμε κάθε ένα βαθμό μαθητή από το μέσο, γ) μετά προσθέσουμε το αποτέλεσμα όλων των αφαιρέσεων και τέλος δ) διαιρέσουμε αυτό που θα βρούμε με τον αριθμό των μαθητών. Αν κάνουμε αυτή τη δουλειά με τα δεδομένα της κατανομής που δόθηκε σαν παράδειγμα, έχουμε: MD=2. Οι ίδιοι υπολογισμοί βασισμένοι στα δεδομένα της προηγούμενης παραγράφου θα έδιναν διαφορετική τιμή στη μέση απόκλιση, πράγμα που σημαίνει ότι αυτός είναι καλύτερος δείκτης από τον προηγούμενο. Και αυτός, όμως έχει ένα μειονέκτημα: μας υποχρεώνει να αγνοήσουμε αν η διαφορά ανάμεσα σε κάθε βαθμό και στον αριθμητικό μέσο είναι θετική ή αρνητική. Δηλαδή το αποτέλεσμα της κάθε αφαίρεσης θεωρείται απόλυτος αριθμός. Αφού αυτό συμβαίνει, είναι σκόπιμο να υψώσουμε στο τετράγωνο κάθε αποτέλεσμα αφαίρεσης, πράγμα που κάνει την εξαφάνιση του θετικού ή αρνητικού σημείου θεμιτή. Ο νέος δείκτης που προκύπτει έτσι λέγεται "διακύμανση" (variance), συμβολίζεται με το "S" και τον βρίσκουμε με τον τύπο:

$$S^2 = \sqrt{\frac{\sum_{i=1}^{i=N}\left(X_i - \bar{X}\right)^2}{N}} \qquad (4)$$

Σαν αντιστάθμισμα, όμως, στην ύψωση στο τετράγωνο έπρεπε να βγει η τετραγωνική ρίζα του συνόλου των διαφορών. ΄Ετσι προέκυψε ο δείκτης που μας ενδιαφέρει και που λέγεται "τυπική απόκλιση" (standard deviation). Η τυπική απόκλιση συμβολίζεται με το "S" ή το "SD" και υπολογίζεται με τον τύπο:

$$SD = \sqrt{\frac{\sum_{i=1}^{i=N}\left(X_i - \bar{X}\right)^2}{N}} \qquad (5)$$

Πρέπει να γίνει εδώ σαφές ότι ο υπολογισμούς της τυπικής απόκλισης, που είναι βέβαια κουραστικός και χωρίς ενδιαφέρον, δεν αναμένεται να γίνεται δίχως "scientific calculator".

Αυτό που θεωρείται απαραίτητο είναι να γίνει αντιληπτό τι ακριβώς είναι ο κάθε δείκτης, για να μπορεί να χρησιμοποιείται κατά την ερμηνεία των αποτελεσμάτων της εξέτασης των μαθητών: ενδιαφέρει η "έννοια" τυπική απόκλιση, όχι ο ίδιος ο τύπος και οι πράξεις που απαιτεί.

΄Ελεγχος δυσκολίας και διακριτικότητας των ‘items’ ενός test

Πίνακας IV
“Items Analysis”

		Ερώτηση $1^{η}$	Ερώτηση $2^{η}$	...	Ερώτηση $15^{η}$ βαθμός	Τελικός
Οι 3		1	1		1	15
καλύτεροι	1	0	1		1	14
μαθητές	2	1	1		1	14
(25%)	3	1	0		1	12
	6	0	1		0	–
	7	0	0		1	–
	8	1	1	...	0	–
	9	0	0		0	–
Οι 3	0	0	1		0	5
χειρότεροι	1	1	0		0	4
μαθητές	2	1	0		0	4
(25%)		0	1		0	3
Α: Σωστές απαντήσεις		6	7	.	5	
Β: Σύνολο απαντήσεων		12	12		12	
Γ: Σωστές των καλύτερων		2	3	.	3	
Δ: Σωστές των χειρότερων		2	1	.	0	

Σε περίπτωση αξιολόγησης των μαθητών σε ευρεία κλίμακα (π.χ. κατά την πιστοποίηση γλωσσομάθειας), είναι απαραίτητο να μην περιορίζεται κανείς σε υπολογισμό της κεντρικής τάσης και της διασποράς των βαθμών, αλλά να ελέγχει και το ίδιο το τεστ που χρησιμοποίησε. Υπάρχει ένας ολόκληρος κλάδος που είναι γνωστός με τον αγγλικό ορό “item analysis”, στο πλαίσιο του οποίου αναπτύχθηκαν τεχνικές ελέγχου των τεστ και των μεμονωμένων ερωτήσεων ή μερών μιας δραστηριότητας ή άσκησης. Οι δύο πιο γνωστοί χρήσιμοι και απλοί στον υπολογισμό τους δείκτες ελέγχου των items είναι ο *δείκτης δυσκολίας* (difficulty index) και ο *δείκτης διάκρισης* (discrimination index). Το παρακάτω παράδειγμα θα κάνει φανερό και πώς υπολογίζονται και πώς πρέπει να ερμηνεύονται. ΄Εστω ότι σε ένα τεστ για έλεγχο κατανόησης προφορικού λόγου είχαμε περιλάβει 15 ερωτήσεις πολλαπλής επιλογής. Στο τεστ πήραν μέρος 12 μαθητές. Στον παρακάτω πίνακα IV σημειώνεται 0, όταν ο μαθητής έκανε λάθος σε μία ερώτηση και 1, όταν την απάντησε σωστά. Ο δείκτης δυσκολίας βρίσκεται διαιρώντας το Α με το Β, δηλαδή το σύνολο των σωστών απαντήσεων με το σύνολο των απαντήσεων. Στην πρώτη επομένως ερώτηση έχουμε 6:12=0,500. Στη δεύτερη ερώτηση ο δείκτης δυσκολίας είναι 7:12=0,583. Στη δέκατη πέμπτη ερώτηση έχουμε 5:12=0,417. ΄Οπως είναι φανερό, αυτός ο δείκτης μπορεί να έχει τιμή από μηδέν έως ένα και όσο μεγαλύτερος είναι, όσο πλησιάζει στη μονάδα, τόσο ευκολότερη ήταν ερώτηση. Πράγμα που δε δικαιολογεί το όνομά του: θα ήταν σωστότερο να τον ονομάσει κανείς “δείκτη ευκολίας. ή δείκτης διάκρισης, ή καλύτερα “διακριτικής ικανότητας”, υπολογίζεται διαιρώντας το Γ με το Γ + Δ. Στην πρώτη ερώτηση επομένως, έχουμε 2:2+2=0,500. Στη δεύτερη ερώτηση έχουμε 3:3+1=0,750 και στη δέκατη πέμπτη ερώτηση 3:3+0=1,000. Θεωρητικά και αυτός ο δείκτης μπορεί να πάρει τιμή από μηδέν έως ένα, αλλά αυτό συμβαίνει σπάνια. Συνήθως κυμαίνεται από 0,500 έως 1,000. Στις περιπτώσεις που ο δείκτης είναι μικρότερος από 0,700, η ερώτηση δεν ήταν καλή, αφού δεν επέτρεψε να διακρίνουμε αυτούς που μπορούν να κατανοήσουν ένα προφορικό κείμενο.

Συσχετίσεις βαθμολογίας και άλλων δεδομένων

Μια άλλη κατηγορία στατιστικών δεικτών είναι αυτοί που δείχνουν αν υπάρχει σχέση μεταξύ δύο ομάδων δεδομένων, αν η σχέση είναι αρνητική ή θετική και πόση ακριβώς είναι. ΄Οταν λέμε “αρνητική” ή “θετική”, δεν εννοούμε αν αυξάνοντας η κάθε τιμή σε μια από τις δύο ομάδες δεδομένων υπάρχει τάση μείωσης ή αύξησης αντίστοιχα της άλλης ομάδας δεδομένων (αρνητική συνάφεια) ή, αντίθετα, αν όσο μεγαλύτερος είναι ο κάθε βαθμός στη μια ομάδα τόσο μεγαλύτερος είναι και ο αντίστοιχος του στην άλλη (θετική συνάφεια). 0 μικρότερος δείκτης συνάφειας που μπορεί να εξαχθεί με την εφαρμογή ενός τύπου είναι το 0. Μηδέν συνάφεια σημαίνει ότι δεν υπάρχει καμιά σχέση μεταξύ των δεδομένων: οι βαθμοί αλλάζουν προς τα πάνω ή προς τα κάτω μέσα στην κάθε ομάδα δεδομένων, άσχετα με το αν αλλάζουν και εκείνοι της άλλης ομάδας. Ο μεγαλύτερος δείκτης συνάφειας που μπορούμε να βρούμε είναι το 1. Η μονάδα δείχνει μέγιστη σχέση, εξάρτηση, μεταξύ των δεδομένων, αδιάφορα αν είναι αρνητική ή θετική (±1). Υπάρχουν πολλοί δείκτες συνάφειας. ΄Ενας πολύ σημαντικός για την περίπτωση έρευνας στην εκπαίδευση ή για ερμηνεία των αποτελεσμάτων εξέτασης είναι ο “συντελεστής συνάφειας Spearman” (Spearman coefficient of correlation), που συμβολίζεται με το “r_s” και υπολογίζεται με τον τύπο:

$$r_s = 1 - \frac{6\sum D^2}{N(N^2-1)} \quad (6)$$

Το μόνο νέο σύμβολο σε αυτόν τον τύπο είναι το D, που συμβολίζει τη διαφορά μεταξύ δύο βαθμών που πήρε ένας μαθητής σε διαφορετικά μαθήματα ή τη διαφορά μεταξύ δύο αριθμητικών μέσων βαθμολογίας μιας τάξης κ.τ.λ. Προτού υπολογίσει κάποιος όμως τη διαφορά μεταξύ δύο δεδομένων, πρέπει να τα μετατρέψει (αν δεν είναι ήδη τέτοια) σε δεδομένα εκφρασμένα με βάση τακτική κλίμακα. Ας θυμηθούμε ότι μια βαθμολογία είναι ήδη το αποτέλεσμα εφαρμογής ή χρησιμοποίησης τακτικής κλίμακας. Η βαθμολογία όμως παρουσιάζεται με το “ένδυμα” αναλογικής κλίμακας. Έτσι, πρέπει να μπούμε στον κόπο να την μετατρέψουμε και “επίσημα” σε τακτική, δηλαδή να κάνουμε τα απόλυτα αριθμητικά (οκτώ, έξι, δέκα, κ.τ.λ.) να γίνουν τακτικά (όγδοος, έκτος , δέκατος, κ.τ.λ.), με τον απλό τρόπο που παρουσιάζεται παρακάτω στο παράδειγμα. Έστω ότι έχουμε μια τάξη με 5 μέλη τα οποία υποβλήθηκαν σε δύο εξετάσεις σε δύο διαφορετικά μαθήματα: για παράδειγμα στη Φυσική και στη γλώσσα. Θέλουμε να δούμε αν οι μαθητές που είναι καλοί σε ένα από αυτά είναι συνήθως καλοί και στο άλλο, ή αν αντίθετα αυτοί που είναι καλοί στα “θεωρητικά” μαθήματα έχουν πρόβλημα με τα “θετικά” και αντίστροφα. Για να υπολογίσουμε το δείκτη συνάφειας, πρέπει πρώτα να βρούμε το ΣD^2 μετατρέποντας την κλίμακα σε τακτική, αφαιρώντας το βαθμό που πήρε κάθε μαθητής στο ένα μάθημα από το βαθμό που πήρε στο άλλο υψώνοντας τη διαφορά στο τετράγωνο και προσθέτοντας το αποτέλεσμα, όπως στον παρακάτω πίνακα:

Πίνακας V
Υπολογισμός του ΣD^2

α/α	Ονοματεπώνυμο	Βαθμός Φυσικής	Βαθμός Γλώσσας	Τάξη Φυσικής	Τάξη Γλώσσας	Διαφορά	D^2
1	Πέτρος Βλάχος	17	16	2(-ος)	1(-ος)	1	1
2	Μαρία Στίννη	11	13	5(-η)	3(-η)	2	4
3	κ.ο.κ.	...	...	...	...	...	...

Από τη στιγμή που η παραπάνω εργασία μας επέτρεψε να βρούμε ότι το ΣD^2 είναι 11, αντικαθιστώντας στον τύπο βρίσκουμε $r_s = +0{,}450$. Το νούμερο αυτό δείχνει ότι η σχέση είναι θετική και έχει κάποια τιμή: 0,450. Για να αποφασίσουμε αν η σχέση είναι στατιστικά σημαντική, δηλαδή δεν οφείλεται στην τύχη, πρέπει να συγκρίνουμε το μέγεθος της συνάφειας που βρήκαμε με αυτό που δίνεται σε πίνακες εγχειριδίων στατιστικής. Στους πίνακες αυτούς δίνεται η "κριτική τιμή" (critical value) του δείκτη ανάλογα με τον αριθμό περιπτώσεων, δηλαδή, στην περίπτωσή μας, ανάλογα με τον αριθμό μαθητών (N). Αν ο δείκτης συνάφειας του Spearman που βρήκαμε είναι μεγαλύτερος από αυτόν που μας δίνει ο πίνακας (αδιάφορα αν είναι θετικός ή αρνητικός), τότε λέμε ότι υπάρχει στατιστικά σημαντική σχέση. Αντιγράφω παρακάτω έναν τέτοιο πίνακα από το σύγγραμμα του Wright, υπό μορφή παραδείγματος.

Πίνακας VI
Κριτικές τιμές του r_S

N	Επίπεδο σημαντικότητας 5%
5	1,000
6	0,886
7	0,786
8	0,738
9	0,683
10	0,648

Στον πίνακα εμφανίζεται ο όρος "επίπεδο σημαντικότητας" (level of significance), τον οποίο δεν χρειάζεται εδώ να εξετάσουμε προσεκτικά: αρκεί να γίνει αντιληπτό ότι αυτός δηλώνει πόσο ανεκτικοί είμαστε με τον ορισμό των κριτικών τιμών. Η παρουσίαση ενός δείκτη συνάφειας έγινε, κυρίως, για να συλλάβει ο αναγνώστης την έννοια "δείκτης συνάφειας" και την ενδεχόμενη χρησιμότητα αυτών των δεικτών. Εδώ, στο παραπάνω παράδειγμα, ο δείκτης του Spearman μας επιτρέπει να ελέγξουμε την υπόθεση της ύπαρξης σχέσης ανάμεσα στην επίδοση των μαθητών σε θεωρητικά και θετικά μαθήματα: η σχέση δεν είναι σημαντική, αφού με μόνο 5 μαθητές, για να είναι σημαντική η σχέση, έπρεπε να βρούμε δείκτη με τιμή 1,000 και όχι μόνο 0,450.

Παραγοντική ανάλυση (factor analysis)

Αν και η μαθηματική θεμελίωση της μεθόδου έγινε στις αρχές του αιώνα μας (Pearson K., 1901), μόλις τις τελευταίες δύο περίπου δεκαετίες χρησιμοποιείται ευρέως από τους ερευνητές, κυρίως εξαιτίας της χρήσης ηλεκτρονικών υπολογιστών. Η μέθοδος αυτή είναι ένα σύνολο τεχνικών σχεδιασμένων έτσι ώστε από τον υπολογισμό των συσχετίσεων στο υπό εξέταση σύνολο μεταβλητών (συνήθως ισχυρά συσχετισμένων μεταξύ τους), να εξάγει ένα νέο σύνολο «λανθανόντων» ή κρυφών μεταβλητών (factors), ασυσχέτιστων μεταξύ τους και πολύ μικρότερων σε αριθμό από τις αρχικές μεταβλητές. Το αρχικό στάδιο της παραγοντικής ανάλυσης είναι η δημιουργία του πίνακα των συσχετίσεων (R-matrix), οι όροι του οποίου είναι οι συντελεστές συσχέτισης του εξεταζόμενου συνόλου των μεταβλητών. Ακολούθως ελέγχεται η συμβατότητα των στοιχείων με την εφαρμογή της μεθόδου. Οι στατιστικές δοκιμασίες που χρησιμοποιούνται (Spss-user's guide, USA,1994) είναι ο έλεγχος σφαιρικότητας (Bartlett's test of sphericity), κατά τον οποίο ελέγχεται η υπόθεση (ηull hypothesis) περί της μοναδιαίας φύσης του πίνακα συσχετίσεων ($R=\delta_{ij}$) και η μέτρηση της «δειγματοληπτικής επάρκειας» (KMO), σύμφωνα με την οποία συγκρίνονται οι μερικοί με τους απλούς συντελεστές συσχέτισης των μεταβλητών. Στόχος των δοκιμασιών αυτών είναι να εξετασθεί η ισχυρή ή όχι συσχέτιση των μεταβλητών, ώστε να τεκμηριωθεί η υπόθεση για την ύπαρξη κοινών «κρυφών» παραγόντων μεταξύ των μεταβλητών. Ακολουθεί ο καθορισμός των παραγόντων (factor extraction), όπου η πλέον κατάλληλη μέθοδος για εκπαιδευτικές μελέτες είναι η «ανάλυση σε κύριες συνιστώσες» (Principal Component Analysis, PCA). Με τη βοήθεια της Γραμμικής Άλγεβρας υπολογίζονται οι ιδιοτιμές (eigenvalues) και τα ιδιοανύσματα (eigenvectors) του πίνακα συσχετίσεων. Από τις ιδιοτιμές, οι οποίες εκφράζουν το ποσοστά της συνολικής μεταβλητότητας που ερμηνεύει κάθε παράγοντας, είναι δυνατόν με απλά κριτήρια (Jolliffe, 1986) να εντοπιστούν οι στατιστικά σημαντικοί παράγοντες. Επίσης, προσδιορίζονται οι συντελεστές πολλαπλής συσχέτισης των παραγόντων με τις αρχικές μεταβλητές (factor loadings), οι οποίοι ακολούθως χρησιμοποιούνται για πολλαπλές αναλύσεις (όπως για την εύρεση τάσεων, περιοδικοτήτων κ.λπ.). Τέλος, προκειμένου να μεγιστοποιηθεί η εξάρτηση μεταξύ των αρχικών μεταβλητών και κάποιων εκ των παραγόντων, χρησιμοποιείται η διαδικασία «περιστροφής των αξόνων», με πιο διαδεδομένη μέθοδο την Varimax (Manly, 1986). Η παραγοντική μέθοδος πέραν του μαθηματικού της στόχου, δηλαδή από ένα αρχικά σύνολο εξαρτημένων μεταβλητών να δημιουργήσει ένα νέο μικρότερου αριθμού ανεξαρτήτων μεταβλητών, που να ερμηνεύει όσο το δυνατό μεγαλύτερο ποσοστά της μεταβλητότητας του αρχικού, αποτελεί ένα ερευνητικό εργαλείο, με την έννοια ότι οι προσπάθειες των ερευνητών εστιάζουν στην ανεύρεση της φυσικής ερμηνείας των παραγόντων. Για παράδειγμα μπορούμε, εφαρμόζοντας τη μέθοδο με κάποιο στατιστικό πακέτο, να αποφανθούμε σε ποιο υποσύνολο μαθημάτων έχει ή δεν έχει απόδοση κάποιος μαθητής.

Μέθοδος ανάλυσης σε ομάδες (Cluster Analysis)

Σε όλους τους τομείς της Επιστήμης είναι υπαρκτή η ανάγκη της ταξινόμησης και ομαδοποίησης των παραμέτρων των γνωστικών τους αντικειμένων. Στις περιπτώσεις που είναι γνωστοί οι κανόνες ταξινόμησης ή οι ιδιότητες των μελών των ομάδων χρησιμοποιείται η στατιστική μέθοδος της «διακριτής ανάλυσης» (discriminant analysis), ενώ όταν τα παραπάνω δεν είναι γνωστά χρησιμοποιείται η «ανάλυση σε ομάδες» (cluster Analysis). Σκοπός της cluster ανάλυσης είναι η διάσπαση ενός αρχικού συνόλου δεδομένων σε γνήσια υποσύνολα αυτού (ομάδες), τα οποία να εμφανίζουν μεταξύ τους υψηλή ετερογένεια, ενώ τα μέλη των ομάδων να είναι κατά το δυνατόν ομογενή. Υπάρχουν διάφοροι τρόποι και κριτήρια ομαδοποίησης, τα οποία διαφέρουν μεταξύ τους στον τρόπο υπολογισμού της απόστασης δύο σημείων και στον τρόπο κατά τον οποίο δύο υποομάδες συνιστούν μία ομάδα. Η πλέον συνήθης τεχνική ομαδοποίησης είναι η «Ιεραρχική», (hierarchical cluster) η οποία χρησιμοποιεί την Ευκλείδεια απόσταση στις «τυποποιημένες» τιμές (Z-values) των στοιχείων. Με βάση το κριτήριο ομαδοποίησης υπάρχουν διάφορες μορφές Ιεραρχικής Ανάλυσης (Romesburg, 1984), όπως η «single linkage», η «average linkage», η «cendroide linkage». Τέλος ο αριθμός των ομάδων ελέγχεται με διάφορα κριτήρια, συνηθέστερο των οποίων είναι η δοκιμασία με την F-κατανομή.

Φασματική ανάλυση (Spectrum Analysis)

Η φασματική ανάλυση (Blackman-Tukey, 1958) είναι μια μέθοδος διερεύνησης των περιοδικών μεταβολών που ενδεχομένως εμφανίζουν οι υπό εξέταση χρονοσειρές. Σκοπός της μεθόδου είναι να αναλυθεί η μεταβλητότητα της χρονοσειράς σε περιοδικές αρμονικές συνιστώσες και να προσδιορισθεί η φασματική ισχύς που αποτελεί το μέτρο κατανομής της μεταβλητότητας στις διάφορες φασματικές περιοχές. Επίσης, η χρονοσειρά δε θεωρείται απαραίτητα ότι αποτελείται από ένα πεπερασμένο πλήθος αρμονικών με διακριτά μήκη κύματος, αλλά μάλλον ότι συνίσταται από μικρές κυμάνσεις με συνεχή κατανομή μηκών κύματος και με τιμές από το διάστημα [2d, + ∞,] άπου d είναι το χρονικά βήμα της σειράς, ενώ σε άπειρο μήκος κύματος αντιστοιχεί η γραμμική της τάση. Ο αλγόριθμος που χρησιμοποιείται για την εύρεση της φασματικής ισχύος είναι ο εξής: Σε μια χρονοσειρά X_i, i = 1,2,...N, υπολογίζονται αρχικά όλες οι συνδιακυμάνσεις C_r (serial covariances) για τις χρονικές υστερήσεις (lags) από r=1 και r = m < N, σύμφωνα με την εξίσωση

$$C_r = \frac{1}{N-r}\sum_{i=1}^{i=N-r}\left(X_i - \bar{X}\right)\left(X_{i+r} - \bar{X}\right) \qquad (7)$$

Ακολούθως ο μετασχηματισμός Fourier των m + 1 το πλήθος C_r που θα παράγει τις ακατέργαστες φασματικές εκτιμήτριες του φάσματος ισχύος, από τις οποίες θα προκύψει το περιοδάγραμμα. Οι ακατέργαστες φασματικές εκτιμήτριες προκύπτουν από τις ακόλουθες εξισώσεις:

$$So = \frac{1}{2m}\left(C_m + C_0\right) + \frac{1}{m}\sum_{r=1}^{r=m-1} C_r$$

$$Sm = \frac{1}{2m}\left[(-1)^m\left(C_m + C_0\right)\right] + \sum_{r=1}^{r=m-1} C_r \qquad (8)$$

$$S\varkappa = \frac{C_0}{2m} + \frac{1}{m}\left[(-1)^m C_m\right] + \frac{2}{m}\sum_{r=1}^{r=m-1} C_r \cos\frac{\pi\varkappa\,\mathrm{r}}{\mathrm{m}}$$

S_o είναι η μηδενικής τάξης εκτιμήτρια αντιστοιχούσα σε άπειρο μήκος κύματος, το οποίο υποδηλώνει την ενδεχόμενη τάση (trend) της χρονοσειράς S_m είναι η φασματική εκτιμήτρια που αντιστοιχεί στο μικρότερο μήκος κύματος (διπλάσιο της απόστασης μεταξύ δύο διαδοχικών παρατηρήσεων). Τέλος η τρίτη σχέση δίνει τις κ-τάξης εκτιμήτριες $S_\varkappa$. Προκειμένου να αυξηθεί η διακριτική ικανότητα των ακατέργαστων φασματικών εκτιμητριών, έτσι ώστε να αναδειχθούν οι συχνότητες με την μεγαλύτερη φασματική πυκνότητα, υποβάλλονται σε εξομαλύνσεις με αλγορίθμους «κινητών μέσων». Οι τελικές φασματικές εκτιμήτριες μετά την εξομάλυνση με την μέθοδο Hamming (WMO, 1966) υπολογίζονται από τις ακόλουθες σχέσεις:

$$So = \frac{1}{2}\left(So + S_1\right),$$

$$Sm = \frac{1}{2}\left(S_{m-1} + S_m\right) \qquad (9)$$

$$S\varkappa = \frac{1}{4}\left(S_{\varkappa-1} + 2S_{\varkappa} + S_{\varkappa+1}\right)$$

Η μορφή του φάσματος αποκαλύπτει τα διάφορα είδη διαταραχών που εμπεριέχει η χρονοσειρά, π.χ. μια περιοδική μη αρμονική συνιστώσα θα δώσει ένα οξύ μέγιστο στην βασική αρμονική και δευτερεύουσες κορυφές στις άλλες αρμονικές. Επίσης αν η χρονοσειρά συνίσταται από καθαρά τυχαίες μεταβολές τότε το φάσμα τείνει σε μια ορθογώνια μορφή, με την έννοια ότι όλες οι φασματικές εκτιμήτριες εμφανίζουν το ίδιο πλάτος. Η μορφή αυτή είναι γνωστή με τον όρο «λευκός θόρυβος», σε αντιστοιχία με το λευκό φως, όπου η ισχύς ισοκατανέμεται σε όλες τις συχνότητες. Στην περίπτωση κατά την οποία η χρονοσειρά παρουσιάζει «εμμονή», εμφανίζεται μια ελάττωση της φασματικής ισχύος από τα μεγάλα μήκη κύματος στα μικρότερα και το φαινόμενο ονομάζεται «ερυθρός θόρυβος», σε αντιστοιχία με την κατανομή της φασματικής ισχύος της ερυθράς ακτινοβολίας.

Ο έλεγχος σημαντικότητας του φάσματος γίνεται με την εφαρμογή της υπόθεσης περί του «μηδενικού συνεχούς» (null-hypothesis continuum), δηλαδή ότι η χρονοσειρά δεν αποτελείται από στατιστικά σημαντικές περιοδικότητες και ότι το φάσμα είναι «θόρυβος». Συγκεκριμένα αν ο συντελεστής αυτοσυσχέτισης με υστέρηση ένα (r_1), δε διαφέρει του μηδενός, τότε το μηδενικό συνεχές είναι του λευκού θορύβου, και είναι μία οριζόντια γραμμή της οποίας οι τιμές είναι ίσες με τη μέση τιμή όλων των φασματικών εκτιμητριών. Αν $r_1 = 0$ και ταυτόχρονα ισχύουν κατά προσέγγιση οι σχέσεις $r_2 = r_1^2$, $r_3 = r_1^3$ ..., τότε το μηδενικό συνεχές είναι ερυθράς θόρυβος. Στην περίπτωση του ερυθρού θορύβου το μηδενικά συνεχές δίνεται από την εξής εξίσωση:

$$S\varkappa = \frac{\overline{S}\left(1 - r_1^2\right)}{1 - r_1^2 - 2r_1 \cos\frac{\pi\varkappa}{m}} \qquad (10)$$

όπου $\overline{S}$ είναι η μέση τιμή των ακατέργαστων φασματικών εκτιμητριών του φάσματος. Επειδή ο λόγος της φασματικής εκτιμήτριας προς το αντίστοιχο συνεχές ακολουθεί την X^2 κατανομή (Tukey, 1950), είναι δυνατή η αξιολόγηση, συνήθως σε επίπεδο 95%, της στατιστικής σημαντικότητας των περιοδικοτήτων που ενδεχομένως εμφανίζει το φάσμα. Τέλος, πρέπει να τονιστεί το γεγονός ότι η χρήση καταλλήλων αλγορίθμων, όπως ο Fast Fourier Transform (FFT) απλοποιεί σημαντικά το πρόβλημα της φασματικής ανάλυσης, εφόσον η μετάβαση από την χρονική στην φασματική περιγραφή της χρονοσειράς δεν απαιτεί τον υπολογισμό των συναρτήσεων αυτοσυσχέτισης. Πρέπει, επομένως, να διαθέτουμε μια χρονοσειρά των βαθμών του μαθητή προκειμένου να βρούμε περιοδικότητες και φασματική ισχύ.

Στατιστικές δοκιμασίες

X^2-test

Χρησιμοποιείται στις περιπτώσεις που οι παρατηρήσεις είναι ποιοτικής φύσης και βασίζεται στη σύγκριση των συχνοτήτων μιας κατανομής που αναμένονται θεωρητικά (Ε), με τις συχνότητες που παρατηρούνται (Ο).

Οι βαθμοί ελευθερίας ν της μεθόδου ορίζονται ως διαφορά μεταξύ του αριθμού των δεδομένων και του πλήθους των σταθερών που χαρακτηρίζουν την υπό έλεγχο κατανομή. Ορίζοντας

$$X^2 = \sum_{i=1}^{i=N} \frac{\left(O-E\right)_i^2}{E_i} \qquad (11)$$

και ν = Ν - κ, από τους αντίστοιχους στατιστικούς πίνακες της X^2-κατανομής συγκρίνουμε την τιμή του πίνακα με την ευρεθείσα, προκειμένου να δεχτούμε ή όχι τη «μηδενική» υπόθεση.

Παρατήρηση: Όταν οι βαθμοί ελευθερίας βρεθούν ν=1 τότε εφαρμόζεται η διόρθωση Yate's σύμφωνα με την οποία ελαττώνεται η διαφορά |Ο - Ε| κατά 0.5.

t-test

Η μέση τιμή ενός δείγματος μετρήσεων αποτελεί την κατά προσέγγιση εκτίμηση της μέσης τιμής του γενικότερου συνόλου (πληθυσμός), από το οποίο προήλθε το δείγμα. Μέτρο της ενδεχόμενης απόστασης μεταξύ της πραγματικής και της υπολογισθείσης μέσης τιμής αποτελεί το «τυπικό σφάλμα, (standard error = SE) της μέσης τιμής, το οποίο ορίζεται από τη σχέση:

$$SE_{\bar{x}} = \frac{S}{\sqrt{n}} \qquad (12)$$

(S η τυπική απόκλιση και η το μέγεθος του δείγματος). Αν ο πληθυσμός είναι κανονικά κατανεμημένος, τότε στο διάστημα «μέση τιμή ± 1.96 S», περιλαμβάνεται το 95% των παρατηρήσεων. Κατ' αναλογία αποδεικνύεται ότι επί μεγάλου αριθμού παρατηρήσεων (η > 500) η περιοχή που ορίζεται με κέντρο την υπολογισθείσα μέση τιμή και ακτίνα 1,96 SE, δηλαδή $\overline{X} \pm 1{,}96\ SE_{\overline{x}}$ περιέχει την πραγματική μέση τιμή με πιθανότητα 95%. Αν το δείγμα είναι μικρό (μέχρι πλήθος 30-40 παρατηρήσεις), τότε χρησιμοποιείται η t-κατανομή με η-1 βαθμούς ελευθερίας. Τέλος, θα πρέπει με κατάλληλα διαγράμματα που μπορεί να κατασκευάσει κάποιος, να παρουσιάσει με εποπτικό τρόπο το αποτέλεσμα μιας εξέτασης. Κυρίως χρησιμοποιείται το "ιστόγραμμα συχνοτήτων" και το "πολύγωνο συχνοτήτων".

Βιβλιογραφία

Ασημακόπουλος, Δ. (1981): «Σημειώσεις Εφαρμοσμένης Στατιστικής», Μεταπτυχιακό Φυσικού τμήματος, Αθήνα.

Βαλσαμάκη-Ράλλη, Η. (1979): «Εξέταση και βαθμολογία του μαθητή», Αθήνα: Αρουανός.

Brooks, C-Carruthers.(1953):«Handbook of Statistical methods in Meteorology», UK: Pitman.

Conrad, V-Pollak, L. (1950): «Methods in Climatology», U.S.A: Holden Day.

Essenwanger, O. (1976): «Applied Statistics in Atmospheric Science», U.S.A.: The Netherlands.

Jolliffe, I. (1986): «Principal Component Analysis», New York: Wiley.

Μπόρα-Σέντα, Ε., Μωϋσιάδης, Χ. (1992): «Εφαρμοσμένη Στατιστική», Θεσσαλονίκη: Ζήτη.

Παπακωνσταντίνου, Ε., Κάϊτσας, Γ. (1995): «Στατιστική», Αθήνα: ΄Ιων.

Τριχόπουλος, Δ. (1975): «Ιατρική Στατιστική», Αθήνα: Γ. Κ. Παρισιάνος.

Ritchman, M. (1986): Rotation of PCA, J. of Climatology., 6, 293-335.

Spiegel, M.R. (1975): «Πιθανότητες και Στατιστική» (Μτφ. Σ. Περσίδη). Αθήνα: ΕΣΠΙ.

Statistical Package for Social Sciences (1994): «User – Guide», U.S.A.: SPSS

W.M.O. (1966): Climatic change, No 195, T.P. 100, Switzerland: World Meteorological Organization, Department of Climatic change.

Φίλιας, Β. (1977): «Εισαγωγή στη μεθοδολογία και τις τεχνικές των κοινωνικών ερευνών», Αθήνα: Τεχνολογία-Επιστήμη.

Wright, D. (1976): «Understanding Statistics», New York: McGraw-Hill.