

Α. ΕΜΒΑΛΩΤΗΣ Α. ΚΑΤΣΗΣ Γ. ΣΙΔΕΡΙΔΗΣ

**ΣΤΑΤΙΣΤΙΚΗ ΜΕΘΟΔΟΛΟΓΙΑ
ΕΚΠΑΙΔΕΥΤΙΚΗΣ ΕΡΕΥΝΑΣ**

Α΄ ΕΚΔΟΣΗ

ΙΩΑΝΝΙΝΑ 2006

- Α. Εμβαλωτής:** Επίκουρος Καθηγητής Πανεπιστημίου Ιωαννίνων
Α. Κατσής: Επίκουρος Καθηγητής Πανεπιστημίου Πελοποννήσου
Γ. Σιδερίδης Αναπληρωτής Καθηγητής Πανεπιστημίου Κρήτης

Τα DataSets είναι διαθέσιμα στη θέση:
<http://research.edu.uoi.gr/aemvalot>



Όπου στο βιβλίο αναφέρεται ο όρος SPSS αφορά στο εγκεκριμένο λογισμικό SPSS Inc© [www.spss.com, www.spss.gr]

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

	ΠΡΟΛΟΓΟΣ	
1.	ΕΙΣΑΓΩΓΗ	7
1.1	Η έννοια της μεταβλητής	8
1.2	Κλίμακες Μέτρησης	11
1.3	Γραφική παρουσίαση των δεδομένων- ευρημάτων	13
1.4	Παρατηρήσεις αναφορικά με τις γραφικές παραστάσεις των δεδομένων (διαγράμματα)	17
2.	ΑΝΑΛΥΣΗ ΣΥΣΧΕΤΙΣΗΣ	19
2.1	Συντελεστής συσχέτισης Pearson r	21
2.2	Η προϋπόθεση της γραμμικότητας	26
2.3	Η προϋπόθεση της κανονικής κατανομής	28
2.4	Υπολογισμός και ερμηνεία του συντελεστή συσχέτισης Pearson r	30
2.5	Μερικές παρατηρήσεις για τον συντελεστή συσχέτισης	34
2.6	Μη-παραμετρικοί συντελεστές συσχέτισης	35
3.	ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ	37
3.1	Υπολογισμός της γραμμής παλινδρόμησης με τη χρήση του SPSS	42
3.2	Εφαρμογή της προβλεπτικής εξίσωσης	47
4.	ΕΛΕΓΧΟΣ ΜΕΣΟΥ ΟΡΟΥ	49
4.1	Εισαγωγή	50
4.2	Στατιστική Μεθοδολογία	51
4.2.1	Έλεγχος για το μέσο όρο ενός πληθυσμού	52
4.2.1.1	Το σκεπτικό του ελέγχου	52
4.2.1.2	Το κριτήριο ελέγχου	52
4.2.1.3	Εξαγωγή συμπερασμάτων	54
4.3	Σχόλια για τον έλεγχο του μέσου όρου ενός πληθυσμού	56
4.4	Εφαρμογές του ελέγχου για το μέσο όρο ενός πληθυσμού	56
4.5	Έλεγχος για τη σύγκριση μέσων όρων δύο πληθυσμών	61
4.5.1	Ανεξάρτητα δείγματα	61
4.5.1.1	Εφαρμογές για ανεξάρτητα δείγματα	62

4.5.2	Εξαρτημένα δείγματα: Κριτήριο ελέγχου και εφαρμογή	67
4.6	Σχόλια για τον έλεγχο μεταξύ των μέσων όρων δύο πληθυσμών	69
5.	ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ	71
5.1	Εισαγωγή	72
5.2	Μεθοδολογία ANOVA	73
5.2.1	Το σκεπτικό για τη διαδικασία ANOVA	73
5.2.2	Post hoc κριτήρια	74
5.3	Σχόλια για την ANOVA	75
5.4	Εφαρμογή της ANOVA	75
6.	ΤΟ ΚΡΙΤΗΡΙΟ χ^2	79
6.1	Εφαρμογή του χ^2	80
6.2	Τυπολόγιο ανάγνωσης ενδείξεων για τους πίνακες του SPSS	87
7.	ΠΑΡΑΡΤΗΜΑΤΑ	89
7.1	Έλεγχος κανονικότητας	90
7.2	Επιλογή κριτηρίου ελέγχου	92
7.2.1	Σύγκριση της ίδιας μεταβλητής σε δύο δείγματα	92
7.2.2	Σύγκριση της ίδιας μεταβλητής σε περισσότερα από δύο δείγματα.	93
7.2.3	Σύγκριση δύο μεταβλητών στο ίδιο δείγμα	94
8	ΒΙΒΛΙΟΓΡΑΦΙΑ	97

ΠΡΟΛΟΓΟΣ

Στη σύγχρονη κοινωνική & εκπαιδευτική έρευνα η Στατιστική είναι ένα σημαντικότατο εργαλείο για κάθε ερευνητή που ασχολείται με την (ποσοτική κυρίως) ανάλυση δεδομένων. Το κείμενο αποτελεί μέρος μιας ευρύτερης και φιλόδοξης προσπάθειας των συγγραφέων η οποία πρόκειται να ολοκληρωθεί μετά την πιλοτική εφαρμογή του συγκριμένου κειμένου. Το βιβλίο απευθύνεται σε φοιτητές, αλλά και ερευνητές που κάνουν χρήση των εφαρμογών της Στατιστικής στο χώρο κυρίως των Κοινωνικών Επιστημών & των Επιστημών της Αγωγής. Στόχος είναι η παρουσίαση και ανάλυση των κυριότερων εφαρμογών της Στατιστικής Μεθοδολογίας με έμφαση στα παρακάτω σημεία:

- (α) επεξήγηση της θεωρίας με παραδείγματα από το χώρο των Κοινωνικών Επιστημών και των Επιστημών της Αγωγής,
- (β) επιλογή και παρουσίαση του κατάλληλου κάθε φορά στατιστικού κριτηρίου και
- (γ) ερμηνεία των αποτελεσμάτων.

Ένας τομέας στον οποίο δόθηκε ιδιαίτερη προσοχή ήταν η (περιορισμένη) χρήση μαθηματικών εννοιών και όρων. Εκτιμήθηκε ότι για τη μελέτη και αξιοποίηση του σχετικού υλικού δεν θα πρέπει να θεωρείται ως προϋπόθεση η σε βάθος προσέγγιση μαθηματικών εννοιών. Για το σκοπό αυτό καταβλήθηκε προσπάθεια να περιοριστούν στο ελάχιστο οι αναφορές στη μαθηματική ορολογία και τυπολογία, ενώ παρουσιάζονται τα σημαντικότερα και απολύτως απαραίτητα μαθηματικά εργαλεία με τρόπο όσο γίνεται πιο κατανοητό. Ο αναγνώστης δεν χρειάζεται να επιμείνει στην (πλήρη) κατανόηση ενός μαθηματικού τύπου, αν αυτό του δημιουργεί πρόβλημα, αλλά στην ποιοτική διάσταση της μεθόδου και των αποτελεσμάτων της. Από την άλλη πλευρά όμως, αυτή η προσπάθεια "απο-μαθηματικοποίησης", δε σημαίνει ότι το κείμενο δεν είναι έγκυρο και δεν απαιτείται σοβαρή και συστηματική προσπάθεια για την κατανόησή του.

Ας σημειωθεί ότι η σειρά με την οποία εμφανίζονται οι συγγραφείς είναι αλφαβητική.

1

ΕΙΣΑΓΩΓΗ

1. ΕΙΣΑΓΩΓΗ

1.1. Η έννοια της μεταβλητής

Η έρευνα στις Κοινωνικές Επιστήμες και ειδικότερα στις Επιστήμες της Αγωγής, βασίζεται στη μελέτη παραγόντων-χαρακτηριστικών που συμμετέχουν στα υπό διερεύνηση φαινόμενα. Ονομάζουμε αυτούς τους παράγοντες "**μεταβλητές**" (variables). Γενικώς μεταβλητή είναι κάθε τι το οποίο επιδέχεται περισσότερες από μία τιμές. Η σχολική επίδοση είναι μια κλασική μεταβλητή που χρησιμοποιείται στην εκπαιδευτική έρευνα. Δέχεται περισσότερες από μία τιμές, οι οποίες μάλιστα δεν χρειάζεται να είναι κατ' ανάγκη αριθμητικές. Στην περίπτωση που οι τιμές μιας μεταβλητής είναι αριθμοί (όπως στην περίπτωση του βάρους ή της ηλικίας επί παραδείγματι) η μεταβλητή ονομάζεται αριθμητική. Οι τιμές της μεταβλητής μπορεί επίσης να αναφέρονται σε κάποια ιδιότητα, όπως στην περίπτωση της μεταβλητής του φύλου (αγόρι, κορίτσι).

Στην κοινωνική έρευνα έχουν προταθεί και καθιερωθεί μια σειρά ταξινομικών συστημάτων, τα οποία διακρίνουν τύπους μεταβλητών. Η συνηθέστερη ταξινόμηση ορίζει "ποιοτικές" (qualitative) και "ποσοτικές" (quantitative) μεταβλητές.

Ποιοτικές μεταβλητές (qualitative variables) ονομάζονται όσες μεταβάλλονται σε είδος, όπως το φύλο, το θρήσκευμα, η εθνικότητα, η οικογενειακή κατάσταση, κ.τ.ό. Σημειώνεται ότι οι τιμές των ποιοτικών μεταβλητών θα πρέπει να συγκροτούν ευδιάκριτες και μη επικαλυπτόμενες κατηγορίες. Έτσι, στη μεταβλητή με την ενδεικτική περιγραφή "πολιτική τοποθέτηση" (θα πρέπει να) ανήκουν σαφώς προσδιορισμένες τιμές. Οι γενικές διατυπώσεις "προοδευτικός", "κεντρώος" και "συντηρητικός" ως τιμές στην παραπάνω μεταβλητή -εάν δεν περιγραφούν/οριστούν με ακρίβεια- δημιουργούν συνθήκες ανασφάλειας ως προς την εγκυρότητα των σχετικών μετρήσεων.

Ποσοτικές μεταβλητές (quantitative variables) ονομάζονται όσες μεταβάλλονται ποσοτικά, όπως το βάρος, η ηλικία, το εισόδημα (όταν αποτυπώνεται αριθμητικά), κ.τ.ό. Ιδιαίτερη σημασία έχει η επιλογή της καταλληλότερης μονάδας μέτρησης, προκειμένου να είναι συγκρίσιμες οι επιμέρους τιμές. Έτσι λοιπόν η θερμοκρασία σε μια μέτρηση θα πρέπει να αποτυπώνεται για το σύνολο των μετρήσεων σε μια και μόνο κλίμακα (κλίμακα Celsius, κλίμακα Kelvin ή κλίμακα Fahrenheit), η οποία καλό θα ήταν μάλιστα να επιλεγεί πριν την έναρξη των μετρήσεων.

Οι ποσοτικές μεταβλητές διακρίνονται σε **συνεχείς** (continuous) και **διακριτές** (discrete).

Η συνεχής (ποσοτική) μεταβλητή επιδέχεται όλες τις τιμές ενός εύρους τιμών, με ελάχιστο κάτω όριο την ελάχιστη τιμή και ανώτατο πάνω όριο τη μέγιστη τιμή ενός συνόλου πραγματικών αριθμών (real numbers). Παράδειγμα συνεχούς ποσοτικής μεταβλητής είναι το βάρος ενός υποκειμένου, το οποίο δέχεται

τιμές από το εύρος τιμών μέτρησης βάρους. Έτσι οι τιμές μπορεί να είναι 67.2 (Kgr), 67.3 (Kgr), 67.4 (Kgr), κ.τ.ό.

Οι διακριτές (ποσοτικές) μεταβλητές επιδέχονται επιλεγμένες τιμές, οι οποίες πρέπει να είναι ακέραιοι αριθμοί. Παράδειγμα διακριτής μεταβλητής είναι η σύνθεση της οικογένειας, όταν η αποτύπωση αναφέρεται στον αριθμό των προσώπων που την αποτελούν. Έτσι, τιμές της παραπάνω μεταβλητής μπορεί να είναι 3 άτομα, 4 άτομα, 5 άτομα, κ.τ.ό, αλλά ποτέ 2.5 άτομα.

Ένας άλλος τύπος διάκρισης και ταξινόμησης των μεταβλητών αναφέρεται σε **ανεξάρτητες** (independent) και **εξαρτημένες** (dependent). Η συγκεκριμένη τυπολογία συναντάται συχνά στην εκπόνηση ερευνητικών σχεδίων και ιδιαίτερα στην ενότητα της διατύπωσης ερευνητικών υποθέσεων.

Ανεξάρτητη μεταβλητή ονομάζεται η μεταβλητή η οποία επηρεάζει αιτιωδώς μια ή περισσότερες εξαρτημένες μεταβλητές. Σε ορισμένες (στατιστικές) αναλύσεις η ανεξάρτητη μεταβλητή ορίζεται και ως προβλεπτικός παράγοντας (predictor variable), με την έννοια ότι προβλέπει την τιμή της (εξαρτημένης) μεταβλητής που επηρεάζει.

Εξαρτημένη μεταβλητή ονομάζεται η μεταβλητή της οποίας οι τιμές εξαρτώνται από τις μεταβολές της ανεξάρτητης μεταβλητής. Σε ορισμένες (στατιστικές) αναλύσεις η εξαρτημένη μεταβλητή ορίζεται και ως αποτέλεσμα (outcome/response variable).

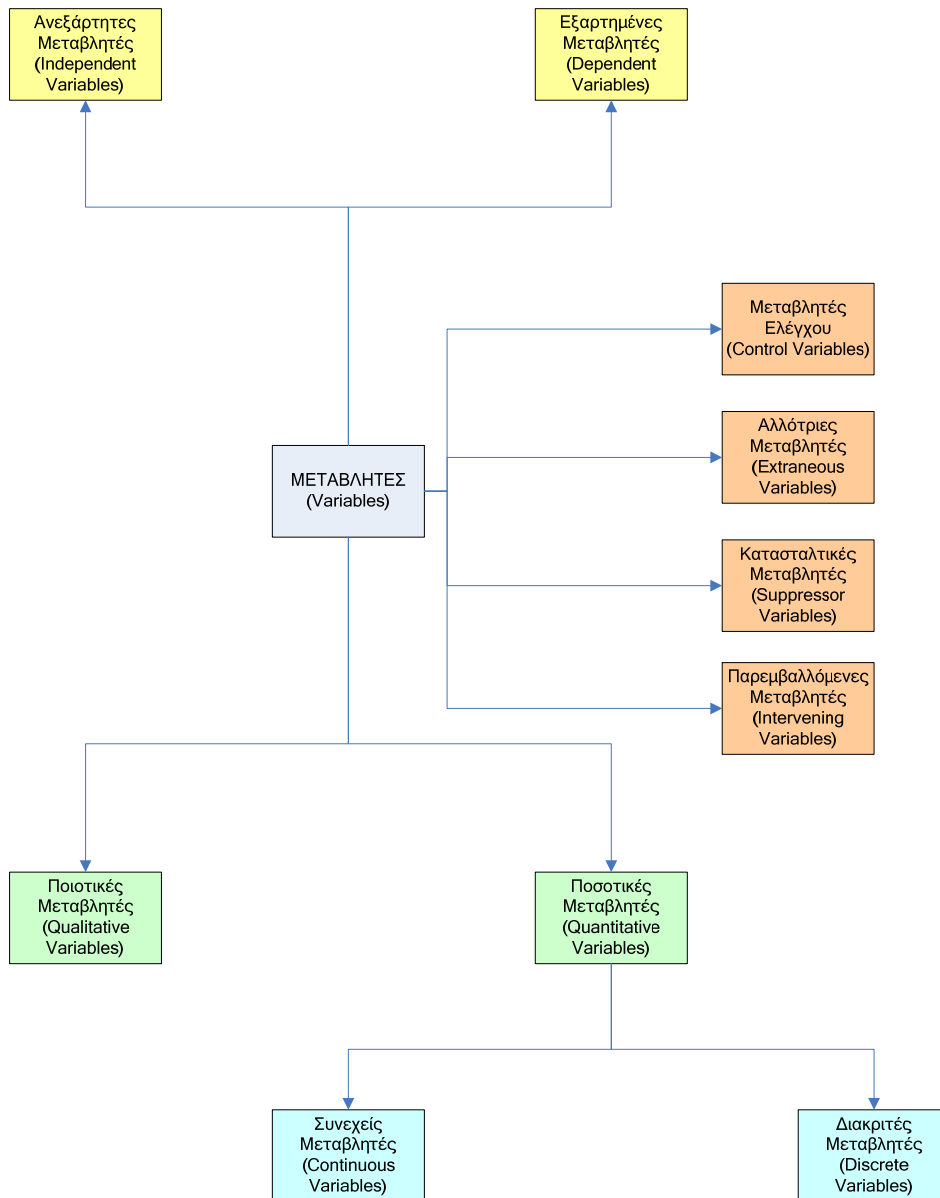
Συμπληρωματικά της παραπάνω ταξινόμησης στις αναλύσεις εντοπίζουμε μεταβλητές **ελέγχου** (control), **αλλότριες** (extraneous) μεταβλητές, **κατασταλτικές** (suppressor) μεταβλητές και **παρεμβαλλόμενες** (intervening) μεταβλητές.

Μεταβλητές ελέγχου ονομάζονται όσες μεταβλητές ουδετεροποιούνται από τον ερευνητή, ή των οποίων η δραστική επιρροή περιορίζεται από τον ερευνητή, προκειμένου να μετρηθεί η επίδραση της ανεξάρτητης στην εξαρτημένη μεταβλητή.

Αλλότριες μεταβλητές ονομάζονται όσες (ανεξάρτητες) μεταβλητές δεν απασχολούν τον ερευνητή, παρότι ενδέχεται να επηρεάζουν την εξαρτημένη μεταβλητή.

Κατασταλτικές μεταβλητές ονομάζονται όσες (ανεξάρτητες) μεταβλητές παρεμβαίνουν τόσο δραστικά στη σχέση ανεξάρτητης – εξαρτημένης μεταβλητής ώστε να περιορίζεται η εμβέλεια της ανεξάρτητης μεταβλητής.

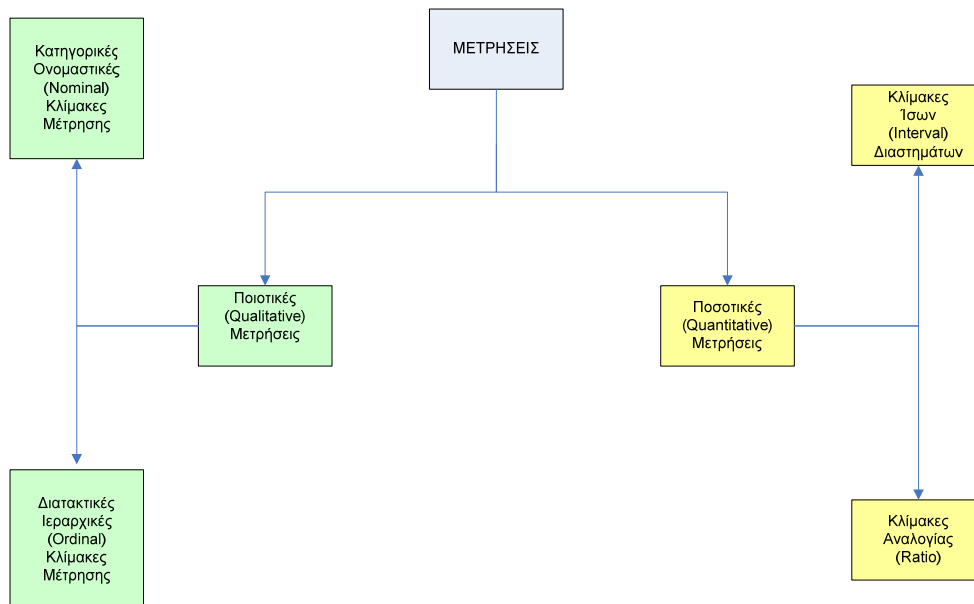
Παρεμβαλλόμενες μεταβλητές ονομάζονται όσες μεταβλητές παρεμβάλλονται κατά τη διάρκεια της έρευνας και ενδεχομένως η επίδρασή τους δεν είχε προβλεφθεί ή δεν είναι δυνατό να περιοριστεί ή να ακυρωθεί.



Διάγραμμα 1

1.2. Κλίμακες μέτρησης

Η έναρξη της συστηματικής συζήτησης για τις κλίμακες μέτρησης ορίζεται με αναφορά στις εργασίες του Stevens (1946). Οι ταξινομήσεις αναφορικά με τις μετρήσεις που προκύπτουν στο ερευνητικό πεδίο οργανώνονται σε δύο τύπους, ανάλογα με τον τύπο της μεταβλητής που διερευνάται. Η βασική διάκριση γίνεται ανάμεσα σε **ποσοτικές & ποιοτικές** μετρήσεις. Οι ποσοτικές μετρήσεις προκύπτουν από την επεξεργασία των ποσοτικών (quantitative) μεταβλητών, ενώ οι ποιοτικές μετρήσεις προκύπτουν από την επεξεργασία των ποιοτικών (qualitative) μεταβλητών. Στη μέτρηση ποιοτικών μεταβλητών εντοπίζουμε **κατηγορικές** ή **ονομαστικές** (nominal) και **διατακτικές** ή **ιεραρχικές** (ordinal) **κλίμακες** μέτρησης, ενώ στις ποσοτικές μεταβλητές εντοπίζουμε **κλίμακες ίσων διαστημάτων** (interval) και **κλίμακες αναλογίας** (ratio).



Διάγραμμα 2

Η ονομαστική κλίμακα εφαρμόζεται όταν επιχειρείται απλή ταξινόμηση χαρακτηριστικών. Το σύνολο τιμών δεν έχει κάποια ιδιότητα και η σημασία του αναγνωρίζεται μόνο στο πλήθος των (διακριτών) τιμών. Παραδείγματα ονομαστικής κλίμακας αναγνωρίζονται στη μεταβλητή του φύλου ή της οικογενειακής κατάστασης ενός υποκειμένου. Οφείλει να επισημάνει κανείς ότι στις μετρήσεις που προκύπτουν με χρήση ονομαστικής κλίμακας απαιτείται ιδιαίτερη προσοχή στην εκτίμηση των μέτρων θέσεως και διασποράς¹.

¹ Βλ. σχετική ενότητα

Η διατακτική κλίμακα εφαρμόζεται όταν οι τιμές της μεταβλητής αναπτύσσονται κατά τέτοιο τρόπο ώστε να είναι δυνατή η ανίχνευση κάποιας σχέσης διάταξης. Για παράδειγμα, η σειρά τερματισμού σε κάποιο αγώνισμα ή η αξιολόγηση των κινηματογραφικών ταινιών της εβδομάδας είναι περιπτώσεις εφαρμογής διατακτικής κλίμακας. Στον έλεγχο διατακτικής κλίμακας ίσες διαφορές μεταξύ των τιμών της μεταβλητής δεν συνεπάγονται και ίσες διαφορές για το χαρακτηριστικό που συνοδεύει τη μεταβλητή. Για παράδειγμα, η διαφορά στο χρόνο μεταξύ του πρώτου και του δεύτερου αθλητή σε έναν (μαραθώνιο) αγώνα δρόμου μπορεί να είναι 2 δευτερόλεπτα, ενώ η διαφορά του δεύτερου από τον τρίτο 30 λεπτά(!) Με βάση αυτή την κλιμάκωση το μέγεθος της διαφοράς δεν έχει σημασία, ενώ εξετάζεται μόνο η σειρά (πρώτος, δεύτερος, τρίτος, κλπ.). Στις μετρήσεις που προκύπτουν με χρήση διατακτικής κλίμακας, ορισμένα μέτρα θέσης μπορούν να χρησιμοποιηθούν με προσοχή ως δείκτες τάσης.

Στην κλίμακα ίσων διαστημάτων οι τιμές της μεταβλητής που επιδέχονται όχι μόνο σχέση διάταξης, αλλά και ίσες διαφορές μεταξύ των τιμών, συνεπάγονται και ίσες διαφορές του χαρακτηριστικού που μετράει η μεταβλητή. Παράδειγμα εφαρμογής κλίμακας διαστήματος έχουμε στην περίπτωση ελέγχου θερμοκρασιακών μετρήσεων. Η διαφορά ανάμεσα στους 10°C και 20°C (10°C) στη στήλη του θερμομέτρου, είναι ίση με τη διαφορά ανάμεσα στους 80°C και 90°C (10°C). Ιδιαίτερο γνώρισμα των μεταβλητών διαστήματος είναι ότι η αρχική τιμή ορίζεται αυθαίρετα. Αυτό επιτρέπει σε κάθε γραμμικό μετασχηματισμό της μορφής $y=a+bx$ (όπου x η αρχική τιμή, y η μετασχηματισμένη b μια θετική και a κάποια τυχαία σταθερά) να διατηρεί τις ιδιότητες της κλίμακας (Τσάντας, κ.ά. 1999:1-28)².

Η κλίμακα αναλογίας εφαρμόζεται όταν οι τιμές της μεταβλητής αντιστοιχούν αναλογικά στην ποσότητα του χαρακτηριστικού της μεταβλητής που μετρούν. Στην περίπτωση αυτή η αρχική τιμή (το μηδέν επί παραδείγματι), ορίζεται με σαφήνεια και καθιστά την αφετηρία βαθμονόμησης (και βαθμολόγησης) των μεταβολών στις τιμές. Κλίμακες αναλογίας επιτρέπουν μετασχηματισμούς της μορφής $y=bx$ (Τσάντας, κ.ά. 1999: 1-28). Παραδείγματα τέτοιων μετασχηματισμών είναι οι μετατροπές δραχμών σε € (1 € = 340,750 δραχμές)

² Μελετήστε το κεφάλαιο της ανάλυσης παλινδρόμησης σε αυτές τις σημειώσεις για την επεξήγηση αυτής της συνάρτησης.

1.3. Γραφική παρουσίαση των δεδομένων – ευρημάτων

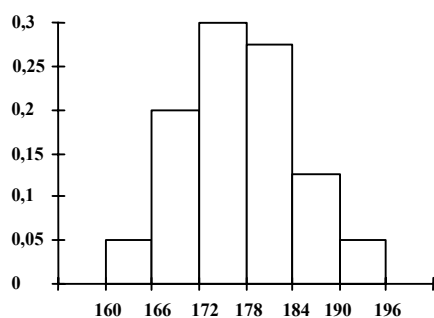
Από τις κρισιμότερες ενέργειες-αποφάσεις του ερευνητή που ενδιαφέρεται να παρουσιάσει ευρήματα κατανομών (και όχι μόνο), είναι η περιεκτική και αντιπροσωπευτική (γραφιστική) απόδοσή τους με τη βοήθεια κατάλληλων (γραφικών) παραστάσεων. Μια γραφική παράσταση οφείλει να είναι:

- (α) παραστατική, διευκολύνοντας την αναγνώριση και κατανόηση των βασικών χαρακτηριστικών της εξεταζόμενης μεταβλητής,
- (β) σαφής, αποφεύγοντας τις συγχύσεις ανάμεσα στις τιμές των μεταβλητών και
- (γ) ακριβής, ώστε να αποφεύγονται οι (οπτικές συνήθως) πλάνες.

Οι συνηθέστερες γραφικές παραστάσεις είναι το ιστόγραμμα συχνοτήτων (histogram), το πολύγωνο συχνοτήτων (frequency polygon), η καμπύλη συχνοτήτων (frequency curve), το ραβδόγραμμα (bar chart), το κυκλικό διάγραμμα (pie chart) και το θηκόγραμμα (boxplot), ενώ σπανιότερες το διάγραμμα μίσχου – φύλλων (stem and leaf diagram) και το σημειόγραμμα (dot diagram).

- Το ιστόγραμμα (histogram)

Πρόκειται για τη συνηθέστερη επιλογή γραφικής παράστασης ποσοτικών μεταβλητών. Οι τιμές της μεταβλητής διατάσσονται (συνήθως) στον οριζόντιο άξονα κατ' αύξουσα σειρά με αφετηρία την αρχή των αξόνων. Το ύψος κάθε στήλης αντιστοιχεί στη συχνότητα κάθε τιμής³ και δίνεται πάντοτε σημασία στην καταλληλότητα της κλίμακας μέτρησης στους άξονες αναφοράς⁴ (βλ. ενδεικτικά και <http://www.shodor.org/interactivate/activities/histogram/>).



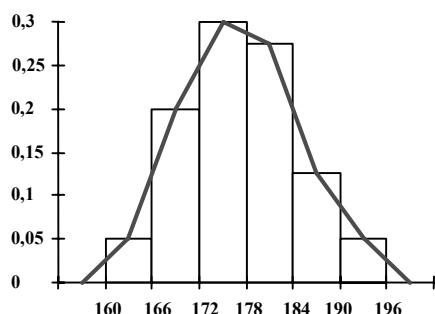
Εικόνα 1: Ιστόγραμμα

³ Εάν πρόκειται για ομαδοποιημένη κατανομή αναφέρεται το εύρος τιμών.

⁴ Σε πολλές περιπτώσεις το βήμα (scale) επιλέγεται από τον ερευνητή μετά την επεξεργασία των δεδομένων και την παραγωγή του σχετικού πίνακα συχνοτήτων.

- Το πολύγωνο συχνοτήτων (frequency polygon)

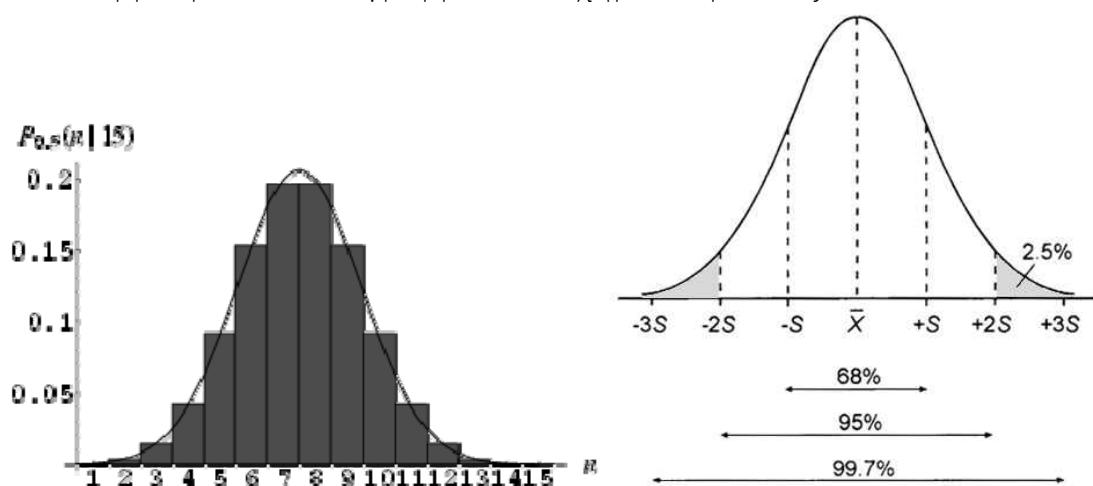
Το πολύγωνο συχνοτήτων σχηματίζεται συνδέοντας την κεντρική τιμή κάθε στήλης με ευθύγραμμα τμήματα. Το βασικό χαρακτηριστικό του είναι ότι το εμβαδόν της περιοχής που περικλείεται από την πολυγωνική γραμμή και τον οριζόντιο άξονα x , ισούται με το σύνολο των συχνοτήτων.



Εικόνα 2: Πολύγωνο (σχετικών) συχνοτήτων

- Η καμπύλη συχνοτήτων (frequency curve)

Η καμπύλη συχνοτήτων παράγεται ως γραφική παράσταση της θεωρητικής κατανομής των συχνοτήτων⁵ και μπορεί να εξαχθεί από το πολύγωνο συχνοτήτων. Οι συνηθέστερες (και γνωστότερες) κατανομές είναι η ομοιόμορφη κατανομή (rectangular distribution), η κανονική κατανομή (normal distribution), η κατανομή student και η F κατανομή. Ιδιαίτερη αναφορά χρειάζεται να γίνει στην κανονική κατανομή (ή κατανομή Gauss) η οποία θεωρείται από τις σημαντικότερες στην ανάλυση ερευνητικών δεδομένων λόγω της συμμετρίας και της ευρείας χρήσης της. Η κανονική κατανομή αξιοποιείται κυρίως προκειμένου να αναλυθούν μετρήσεις, οι οποίες αναπτύσσονται σε ένα συμμετρικό ιστόγραμμα σε σχήμα καμπάνας.



Εικόνες 3 & 4: Ιστόγραμμα και καμπύλη κανονικής κατανομής

⁵ Οι θεωρητικές κατανομές συχνοτήτων διακρίνονται σε συνεχείς (κανονική, εκθετική, student, X^2 , κ.λπ.) και ασυνεχείς (διωνυμική, γεωμετρική, υπεργεωμετρική, ομοιόμορφη, κ.λπ.).

- Το ραβδόγραμμα (bar chart)

Το ραβδόγραμμα επιτρέπει τη γραφική απεικόνιση ποιοτικών μεταβλητών. Η συνηθέστερη πρακτική είναι η τοποθέτηση στον οριζόντιο άξονα (x) των κατηγοριών της μεταβλητής και η ανάπτυξη ράβδων με ύψος ανάλογο της συχνότητας για κάθε κατηγορία. Συχνά το ραβδόγραμμα αξιοποιείται από τους ερευνητές προκειμένου να συγκρίνουν ομάδες περιπτώσεων (Norusis 2002:570-573). Μια διασκευή του ραβδογράμματος είναι το διάγραμμα Pareto (Pareto's chart), το οποίο εμφανίζει τις κατηγορίες της ποιοτικής μεταβλητής σε φθίνουσα σειρά (βλ. ενδεικτικά και http://nces.ed.gov/nceskids/Graphing/bar_pie_data.asp?ChartType=bar).

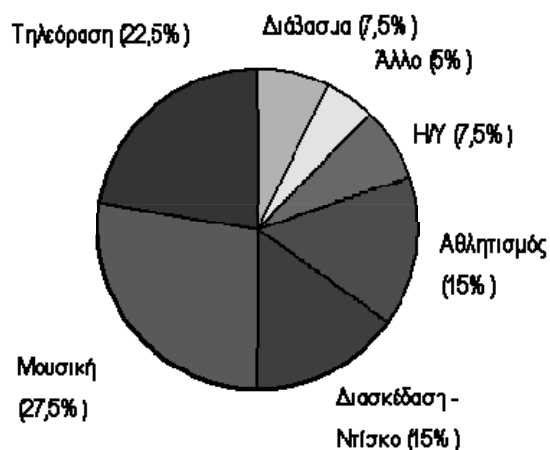
- Το κυκλικό διάγραμμα (pie chart)

Πρόκειται για μια εναλλακτική (του ραβδογράμματος) γραφική απεικόνιση κατηγορικών δεδομένων και βασίζεται στην απλή παραδοχή ότι κάθε κύκλος μπορεί να διαιρεθεί σε 360 τμήματα (360 μοίρες). Για τη δημιουργία του, η συχνότητα κάθε τιμής πολλαπλασιάζεται επί 3.6 (όση και η γωνία που αντιστοιχεί στο 1% του κυκλικού δίσκου), με αποτέλεσμα το γινόμενο να αντιστοιχεί στη γωνία του αντίστοιχου τμήματος στο κυκλικό διάγραμμα. Κατ' αυτό τον τρόπο η απεικόνιση της (σχετικής) συχνότητας 34 είναι η γωνία με άνοιγμα $34 \times 3.6 = 122.4$ μοίρες. Η χρήση των κυκλικών διαγραμμάτων προτείνεται όταν (α) οι τιμές της μεταβλητής αθροίζουν σε σταθερό άθροισμα (το 100% των σχετικών συχνοτήτων), (β) όταν οι κατηγορίες αναπτύσσονται κατά τρόπο αισθητά διαφορετικό⁶ και (γ) όταν ο αριθμός των επιμέρους κατηγοριών δεν είναι ιδιαίτερα μεγάλος⁷.

(βλ. ενδεικτικά http://nces.ed.gov/nceskids/Graphing/bar_pie_data.asp?ChartType=pie, <http://www.shodor.org/interactivate/activities/piechart/index.html>).

⁶ Ένα κυκλικό διάγραμμα το οποίο εμφανίζει οκτώ τομείς (κατηγορίες) με την ίδια συχνότητα δεν είναι παραστατικό.

⁷ Ένα κυκλικό διάγραμμα με πολλούς τομείς δεν είναι (επίσης) παραστατικό.



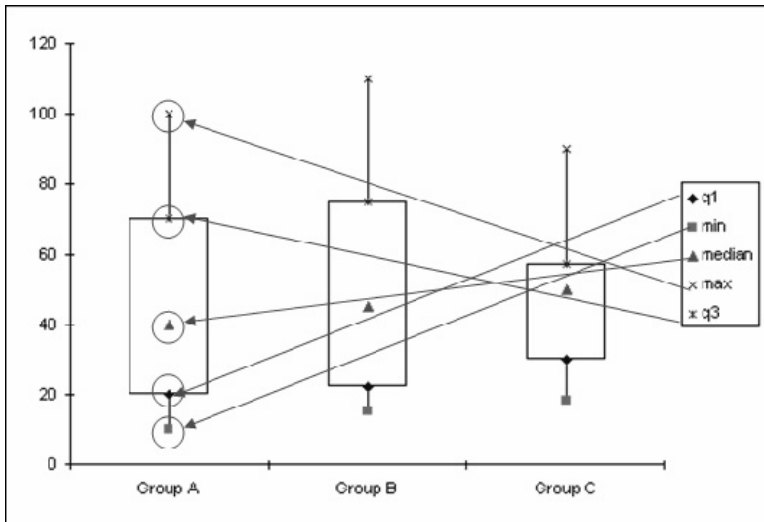
Εικόνα 5: Κυκλικό διάγραμμα

- Το θηκόγραμμα (boxplot)

Η συνήθης εφαρμογή του θηκογράμματος είναι ο σχεδιασμός του προκειμένου να συγκριθούν κατανομές. Για κάθε κατανομή δημιουργείται με τη βοήθεια πέντε στατιστικών στοιχείων θέσεως (ελάχιστο, μέγιστο, 1^ο-2^ο & 3^ο τεταρτημόριο) ένα ορθογώνιο πλαίσιο, το μήκος του οποίου αντιστοιχεί στο ενδοτεταρτημοριακό εύρος της μεταβλητής. Στα άκρα κάθε (ορθογώνιου) πλαισίου (βλ. Εικόνα 6) αναπτύσσονται γραμμές (απολήξεις⁸), οι οποίες συνδέουν τις μέγιστες και ελάχιστες τιμές της μεταβλητής⁹. Με τη βοήθεια του θηκογράμματος ο ερευνητής έχει άμεσα τις πρώτες ενδείξεις αναφορικά με τη συμμετρία της κατανομής.

⁸ Στη σχετική βιβλιογραφία το boxplot αναφέρεται και ως διάγραμμα πλαισίου - απολήξεων.

⁹ Εκτός από τις έκκεντρες (outlier) ή τις ακραίες (extreme) τιμές



Box-plot

Εικόνα 6: Θηκόγραμμα

1.4 Παρατηρήσεις αναφορικά με τις γραφικές παρουσιάσεις των δεδομένων (διαγράμματα)¹⁰

1. Τα διαγράμματα παρέχουν την οπτικοποίηση¹¹ των δεδομένων ενός πίνακα. Πίνακας και διάγραμμα θα πρέπει να συμφωνούν (ως προς τις τιμές) απόλυτα.
2. Κάθε διάγραμμα θα πρέπει να είναι πλήρες. Θα πρέπει δηλαδή ο αναγνώστης μελετώντας το διάγραμμα να ενημερώνεται πλήρως για το περιεχόμενό του γραφήματος, χωρίς να χρειάζεται να δοθούν σχετικές διευκρινίσεις.
3. Κάθε διάγραμμα θα πρέπει να συνοδεύεται από:
 - a. αρίθμηση
 - b. τίτλο (κατανοητό και πλήρη)
 - c. πηγή (όταν τα δεδομένα δεν είναι εμπειρικά).

¹⁰ Για μια ενδιαφέρουσα και αναλυτική σχετική συζήτηση βλ. Κατσιλλής, Ι. (1997). *Περιγραφική Στατιστική: Εφαρμοσμένη στις κοινωνικές επιστήμες και την εκπαίδευση, με έμφαση στην ανάλυση με υπολογιστές*. Αθήνα: Gutenberg. Και Κατσιλλής, Ι. (1998). *Οι μικροϋπολογιστές στις κοινωνικές επιστήμες: επιστημονική εμπειρική έρευνα και στατιστικές αναλύσεις*. Αθήνα: Gutenberg.

¹¹ Η οπτικοποίηση των δεδομένων έχει απασχολήσει τους ερευνητές ορισμένοι εκ των οποίων έχουν επισημάνει την ιδιαίτερη προσοχή που οφείλει να έχει ο ερευνητής προκειμένου να αποφευχθούν οπτικές πλάνες

2

ΑΝΑΛΥΣΗ ΣΥΣΧΕΤΙΣΗΣ

2. ΑΝΑΛΥΣΗ ΣΥΣΧΕΤΙΣΗΣ

Υπάρχει σχέση¹² (και αν ναι, ποια) μεταξύ διαβάσματος και επίδοσης, ευφυΐας και επαγγελματικής αποκατάστασης; Σε σχετικές ερευνητικές ερωτήσεις καλείται να απαντήσει συχνά η έρευνα στις Κοινωνικές Επιστήμες και τις Επιστήμες της Αγωγής. Προκειμένου να απαντηθούν τα ερωτήματα χρειαζόμαστε αξιόπιστα στατιστικά εργαλεία, με τα οποία θα αξιολογηθούν τα ερωτήματα και θα επιχειρηθεί να δοθούν πειστικές απαντήσεις.

Ας υποθέσουμε ότι η ερευνητική μας υπόθεση αφορά στις διαφορές μεταξύ αγοριών και κοριτσιών αναφορικά με την επίδοση τους στην ανάγνωση. Σε σχετική μέτρηση διαπιστώθηκε ότι τα αγόρια έχουν μέσο όρο 5 ενώ τα κορίτσια 7. Είμαστε βέβαιοι ότι τα κορίτσια έχουν καλύτερη επίδοση από τα αγόρια; Ενδεχομένως οι περισσότεροι να ισχυριστούν "ναι", άλλοι όμως μπορεί να είναι περισσότερο επιφυλακτικοί. Αν το άριστα δεν είναι το 10 αλλά το 20; Θα υπάρχουν περισσότεροι που θα απαντήσουν "όχι"; Αν το άριστα είναι το 50; Προκειμένου να δοθεί απάντηση στα παραπάνω ερωτήματα η στατιστική παρέχει στους ερευνητές "αντικειμενικά κριτήρια", επιτρέποντας να απαντήσουν με τρόπο κατηγορηματικό στο αν οι διαφορές που παρατηρήθηκαν δεν είναι "τυχαίες" (random) -δηλαδή δεν οφείλονται σε τυχαίους παράγοντες αλλά σε συστηματικούς που αξίζει να μελετηθούν- ή αντίθετα οφείλονται σε τυχαίους παράγοντες και δεν επιδέχονται περαιτέρω ανάλυση. Για τον καθορισμό αυτής της τυχειότητας απαιτείται η χρήση πιθανοτήτων. Χωρίς να μπούμε σε μαθηματικές αναλύσεις, αναφέρουμε ότι στη στατιστική κάτι θεωρείται τυχαίο όταν είναι αρκετά συνηθισμένο, συμβαίνει δηλαδή αρκετές φορές, τουλάχιστον στο 5% των περιπτώσεων. Αντιθέτως, όταν συμβαίνει λιγότερες από 5 στις 100 φορές, τότε θεωρούμε ότι το συγκεκριμένο αποτέλεσμα δε συμβαίνει κατά τύχη, είναι "σπάνιο" και αξίζει να συζητηθεί. Είναι φανερό ότι επιλογές-αποφάσεις του τύπου "μη τυχαίο - τυχαίο" (ή αντιστοίχως "σπάνιο - συνηθισμένο"), δημιουργούν άλλα προβλήματα. Για παράδειγμα κάτι που συμβαίνει 4 φορές στις 100 είναι μη τυχαίο (σπάνιο), ενώ κάτι που συμβαίνει 5 φορές στις 100 δεν είναι; Ουσιαστικά κάθε ερευνητής αποφασίζει για το επίπεδο τυχειότητας που ταιριάζει στο "πρόβλημα" που μελετά. Τα τυπικά επίπεδα τυχειότητας που χρησιμοποιούνται (συνήθως) είναι το 10%, το 5%, το 1% και το 1%.

Επομένως, στο παραπάνω παράδειγμα που εξετάζουμε αν υπάρχουν ουσιαστικές διαφορές επίδοσης μεταξύ αγοριών και κοριτσιών, ο στατιστικός δείκτης που θα χρησιμοποιηθεί πρέπει να σχετίζεται με μια πιθανότητα εμφάνισης του συγκεκριμένου αποτελέσματος, η οποία να συγκρίνεται με το 5% (εκτός και αν οι ερευνητές έχουν αποφασίσει να ορίσουν πιο "αυστηρό"

¹² Σε αυτό το κεφάλαιο χρησιμοποιούμε τους όρους «σχέση» και «συνάφεια» σαν να είναι ταυτόσημοι.

κριτήριο). Μόνο αν η πιθανότητα είναι μικρότερη του 5% έχουμε στατιστικά σημαντικό εύρημα, το οποίο (υπο)δηλώνει μη-τυχαίες διαφορές και άρα διαπιστώνεται σχέση μεταξύ φύλου και επίδοσης. Σε αντίθετη περίπτωση θεωρούμε ότι δεν υπάρχει στατιστικά σημαντική σχέση. Η υπόθεση της μη στατιστικά σημαντικής σχέσης ονομάζεται και "αρχική" ή "μηδενική" υπόθεση (H_0), ενώ η υπόθεση της στατιστικά σημαντικής σχέσης ονομάζεται και "εναλλακτική" ή "ερευνητική" υπόθεση ή και "υπόθεση εργασίας" (H_1 , ή H_A , ή H_e). Η διατύπωση των υποθέσεων είναι απαραίτητη για την απεικόνιση των μεταβλητών και την επιλογή των κριτηρίων στατιστικού ελέγχου.

2.1. Συντελεστής συσχέτισης Pearson r

Ο συντελεστής συσχέτισης Pearson r είναι ο κατάλληλος (παραμετρικός) στατιστικός δείκτης προκειμένου να αξιολογηθεί αν υπάρχει συνάφεια μεταξύ δύο μεταβλητών (π.χ., ωρών διαβάσματος και επίδοσης). Με τον όρο παραμετρικός αναφερόμαστε σε μια σειρά στατιστικών δεικτών που ικανοποιούν συγκεκριμένες προϋποθέσεις. Αν αυτές οι προϋποθέσεις "απουσιάζουν", τότε γίνεται χρήση μη-παραμετρικών στατιστικών δεικτών, οι οποίοι δεν επηρεάζονται από την τήρηση ή "καταπάτηση" των προϋποθέσεων. Για τον συντελεστή συσχέτισης Pearson r οι προϋποθέσεις αυτές είναι:

- (α) οι μεταβλητές πρέπει να είναι συνεχείς σε κλίμακα ίσων διαστημάτων και να κατανομούνται κανονικά &
- (β) να υπάρχει γραμμική σχέση μεταξύ των δύο μεταβλητών.

Αυτό που απαιτείται για την ανίχνευση των σχέσεων, είναι να υπάρχει αναλογική σχέση μεταξύ των δύο μεταβλητών για όλα τα επίπεδά τους. Δηλαδή, όσο μεγαλώνει η μία να μεγαλώνει και η άλλη και το αντίθετο. Οι προϋποθέσεις αυτές γίνονται κατανοητές στη συνέχεια με τη χρήση του διαγράμματος σκεδασμού και με σχετικές αναλύσεις κανονικότητας.

Το διάγραμμα σκεδασμού είναι ένα πάρα πολύ χρήσιμο διάγραμμα, το οποίο απεικονίζει το είδος και (κατά προσέγγιση), το μέγεθος της σχέσης των δύο μεταβλητών που μελετά ο ερευνητής. Μπορεί να χαρακτηριστεί και ως μια "φωτογραφία" της υπό μελέτη σχέσης. Στο διάγραμμα αυτό, κάθε μια από τις δύο μεταβλητές απεικονίζεται σε έναν από τους δύο άξονες (σε ποιον άξονα δεν έχει σημασία για την ανάλυση της συσχέτισης) και μέσα στο σχήμα κάθε "κουκίδα" αποτελεί μια παρατήρηση (ένα άτομο), η οποία είναι συνάρτηση των τιμών που έχει το άτομο για (στις) δύο μεταβλητές (πρόκειται για το σημείο τομής αν χαράξουμε παράλληλες γραμμές από την μια τιμή προς τον ένα άξονα και από την άλλη τιμή προς τον άλλο άξονα).

Ας δούμε ένα παράδειγμα:

Υποθέσεις (μηδενική και εναλλακτική):

H_0 : ΔΕΝ υπάρχει σχέση μεταξύ ωρών διαβάσματος και επίδοσης

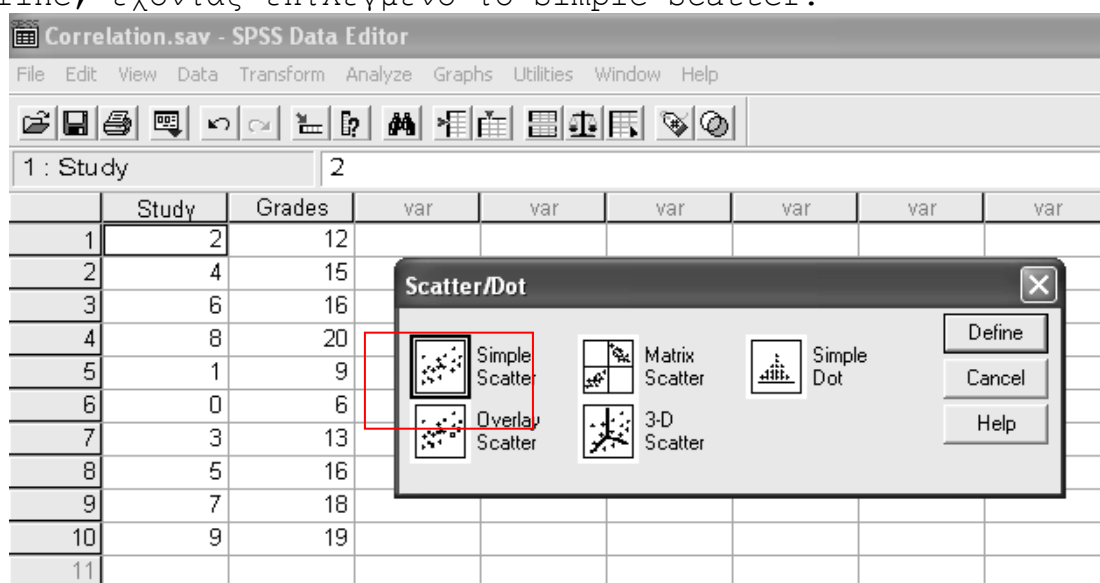
$H_ε$: Υπάρχει σχέση μεταξύ ωρών διαβάσματος και επίδοσης

Αν υποθέσουμε ότι έχουμε 10 μαθητές, για τους οποίους έχουμε τα εξής δεδομένα:

Πίνακας 1: Δεδομένα από υποθετική μελέτη για τη σχέση ωρών διαβάσματος και επίδοσης

Συμμετέχοντες	1	2	3	4	5	6	7	8	9	10
Ώρες Διαβάσματος	2	4	6	8	1	0	3	5	7	9
Επίδοση	12	15	16	20	9	6	13	16	18	19

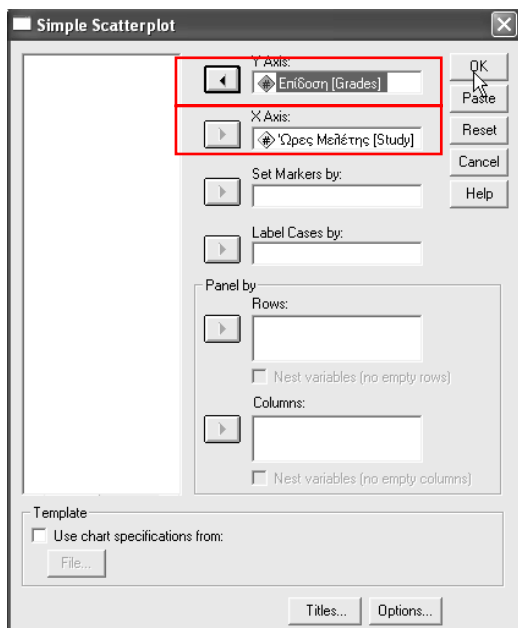
Η δημιουργία του διαγράμματος σκεδασμού (διασποράς) (scatterplot) για την οπτική απεικόνιση του είδους και του μεγέθους της σχέσης στο SPSS έχει ως εξής: Από το μενού Graphs του SPSS επιλέξτε Scatter/Dot. Στη συνέχεια κάντε κλικ στο Define, έχοντας επιλεγμένο το Simple Scatter.



Εικόνα 7: Επιλογές για τη δημιουργία απλού διαγράμματος σκεδασμού.

Στη συνέχεια επιλέξτε τις μεταβλητές που θα τοποθετηθούν στον οριζόντιο και τον κάθετο άξονα. Τοποθετείστε τις Ώρες Μελέτης

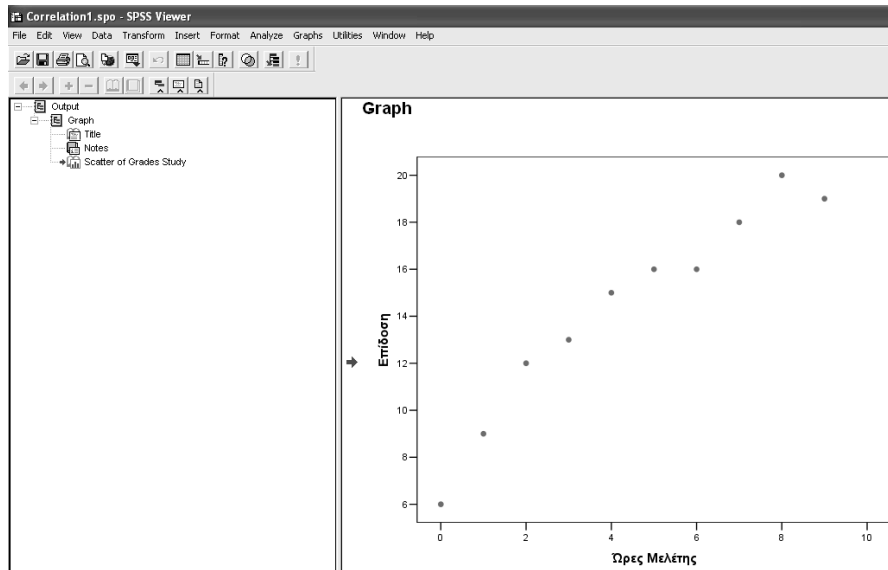
(Study) στον οριζόντιο άξονα και την Επίδοση (Grades) στον κάθετο άξονα¹³. Στη συνέχεια επιλέξτε OK.



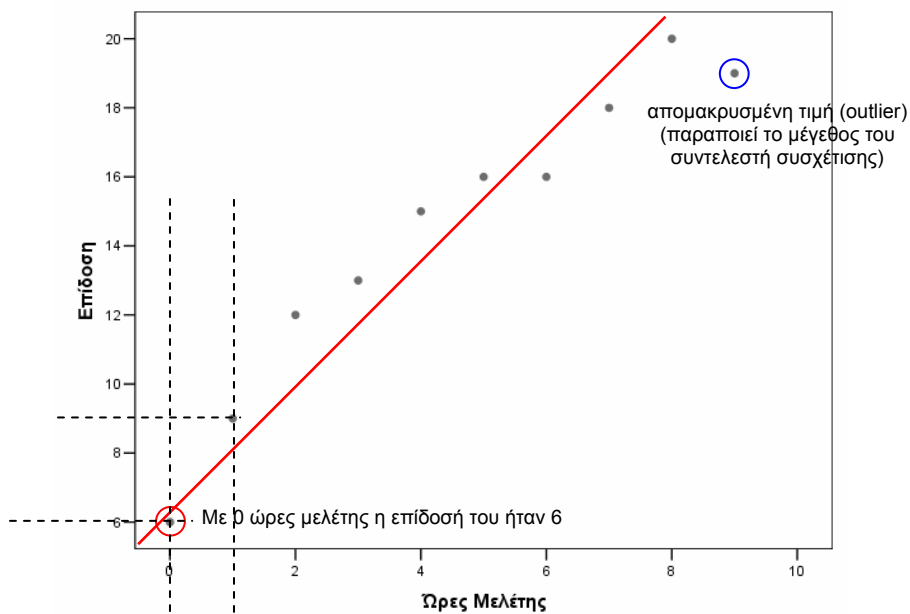
Εικόνα 8: Επιλογή των μεταβλητών σε άξονες για τη δημιουργία απλού διαγράμματος σκεδασμού.

Οι Εικόνες 9α & 9β δείχνουν το διάγραμμα σκεδασμού όπως προκύπτει από το SPSS για τη σχέση μεταξύ ωρών μελέτης και επίδοσης. Όπως φαίνεται στις εικόνες 9α & 9β, όσο περισσότερες ώρες διάβασε κάποιος/α, τόσο μεγαλύτερο βαθμό πήρε (με κάποιες διακυμάνσεις). Για την ακρίβεια, όποτε το σχήμα με τις παρατηρήσεις στο διάγραμμα σκεδασμού έχει την κατεύθυνση από αριστερά και κάτω προς τα δεξιά και πάνω, τότε υπάρχει θετική σχέση μεταξύ των δύο μεταβλητών (Εικόνα 10). Όταν το σχήμα έχει κατεύθυνση από αριστερά και πάνω προς τα δεξιά και κάτω, τότε έχουμε αρνητική σχέση (βλέπε Εικόνα 11), ενώ όποτε μας είναι δύσκολο να εντοπίσουμε οπτικά κάποια κατεύθυνση (όπως η περίπτωση να είναι οι παρατηρήσεις διάσπαρτες στο χώρο) τότε έχουμε μηδενική σχέση (Εικόνα 12).

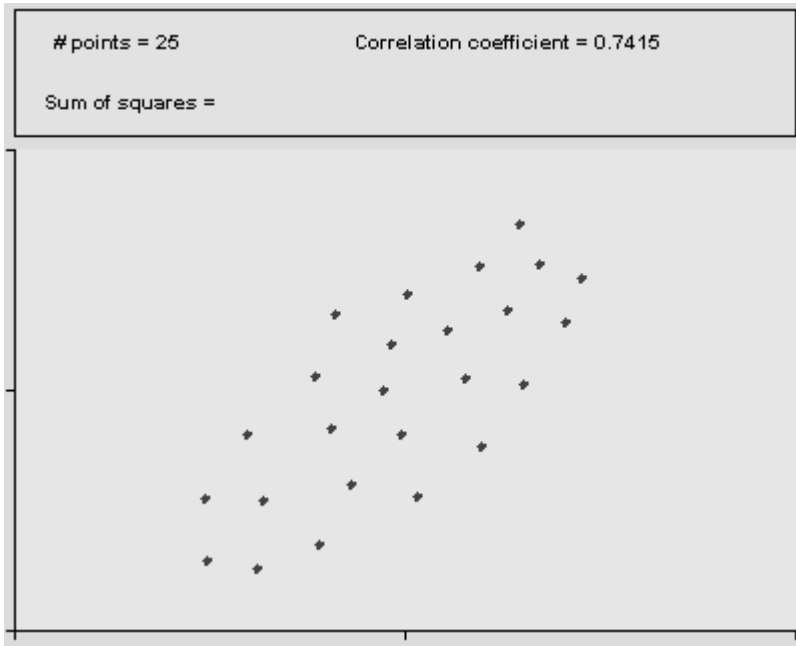
¹³ Στην περίπτωση της απλής συσχέτισης δεν έχει ιδιαίτερη σημασία το ποια μεταβλητή θα τοποθετηθεί σε ποιον άξονα αφού η σχέση $A - B$ είναι ίδια με την $B - A$.



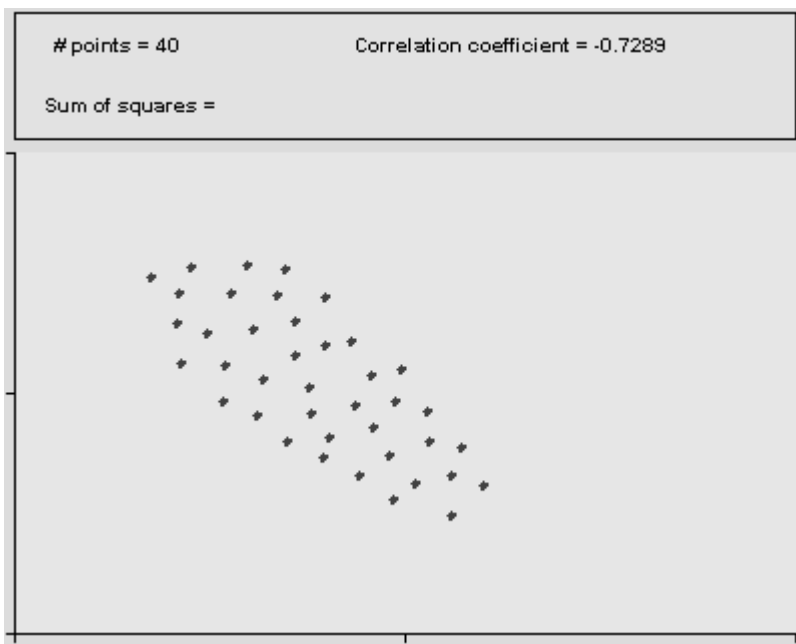
Εικόνα 9α: Διάγραμμα σκεδασμού για τη σχέση ωρών διαβάσματος και επίδοσης.



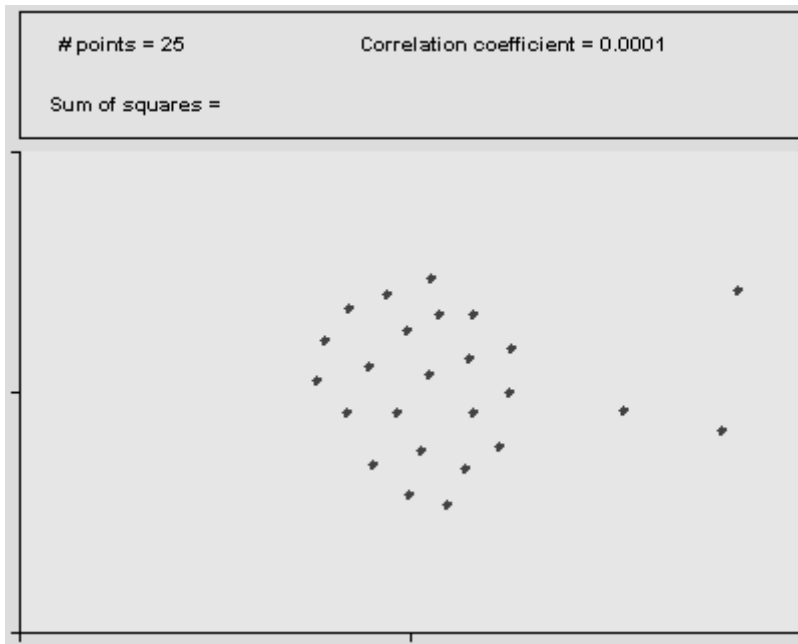
Εικόνα 9β: Διάγραμμα σκεδασμού για τη σχέση ωρών διαβάσματος και επίδοσης.



Εικόνα 10: Διάγραμμα σκεδασμού για την ύπαρξη θετικής σχέσης μεταξύ 2 μεταβλητών.



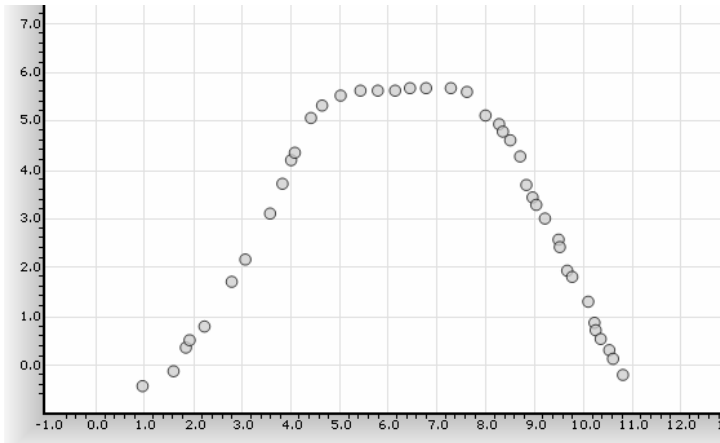
Εικόνα 11: Διάγραμμα σκεδασμού για την ύπαρξη αρνητικής σχέσης μεταξύ 2 μεταβλητών.



Εικόνα 12: Διάγραμμα σκεδασμού για την απουσία σχέσης μεταξύ 2 μεταβλητών.

2.2. Η προϋπόθεση της γραμμικότητας

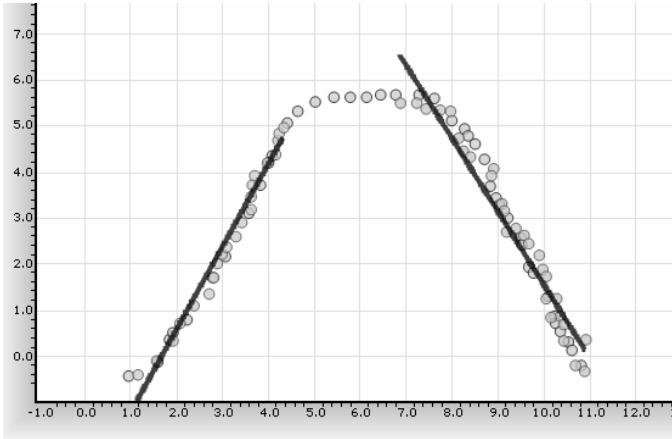
Όπως αναφέρθηκε, προκειμένου να υπολογιστεί ο συντελεστής συσχέτισης θα πρέπει να ισχύει η προϋπόθεση της γραμμικότητας. Με βάση αυτή την προϋπόθεση, θα πρέπει το μοντέλο της ευθείας γραμμής να ταιριάζει καλύτερα στα δεδομένα. Με άλλα λόγια, αν τραβήξουμε μια ευθεία γραμμή στα δεδομένα της Εικόνας 9, θα δούμε ότι τα περισσότερα από αυτά είναι πάνω, ή πολύ κοντά στη γραμμή. Επομένως φαίνεται και οπτικά ότι υπάρχει αναλογική (γραμμική) σχέση μεταξύ των δύο μεταβλητών. Μια από τις περιπτώσεις μη ύπαρξης γραμμικής σχέσης αναφέρεται στη σχέση μεταξύ άγχους και επίδοσης. Με βάση σχετικές θεωρίες (π.χ., Yerkes & Dowdson, 1908) χαμηλά επίπεδα άγχους σχετίζονται με χαμηλές επιδόσεις, μεσαία επίπεδα άγχους μπορεί να σχετίζονται με εξαιρετικές επιδόσεις (επειδή μπορεί να παρέχουν ιδανικά επίπεδα ενεργοποίησης), ενώ πολύ υψηλά επίπεδα άγχους είναι καταστροφικά για την επίδοση (εξαιτίας της ανικανότητας του ατόμου να αυτό-ρυθμίσει τη συμπεριφορά του και τα συναισθήματά του). Αυτό το είδος σχέσης φαίνεται να εκφράζεται καλύτερα με την ύπαρξη καμπυλόγραμμης σχέσης, όπου η σχέση άγχους και επίδοσης είναι γραμμική ως ένα σημείο. Στη συνέχεια, αύξηση των επιπέδων άγχους συνδυάζεται με μείωση της επίδοσης (βλέπε Εικόνα 13).



Εικόνα 13: Διάγραμμα σκεδασμού που εκφράζει την υποθετική σχέση μεταξύ άγχους και επίδοσης.

Όπως φαίνεται στην Εικόνα 13, υπάρχουν κάποια ιδανικά επίπεδα άγχους (περίπου στο μέσο του οριζόντιου άξονα) στα οποία επιτυγχάνονται οι υψηλότερες επιδόσεις. Αντίθετα, όταν ξεπεραστούν αυτά τα επίπεδα, τότε η επίδοση μειώνεται δραματικά. Η απουσία της προϋπόθεσης της γραμμικότητας γίνεται φανερή στην Εικόνα 14, όπου είναι σχεδόν αδύνατο να τραβήξουμε μια ευθεία γραμμή η οποία να ταιριάζει στα περισσότερα δεδομένα. Βλέπουμε ότι η πρώτη γραμμή ταιριάζει στο πρώτο μισό των παρατηρήσεων, ενώ η δεύτερη γραμμή στο δεύτερο μισό. Αν λοιπόν εφαρμόζαμε την πρώτη ευθεία γραμμή θα βγάζαμε το συμπέρασμα της θετικής συνάφειας, ενώ με τη δεύτερη το συμπέρασμα της αρνητικής. Επειδή λοιπόν η σχέση των δύο μεταβλητών αρχίζει να διαφοροποιείται (από θετική προς αρνητική) μετά από κάποιο επίπεδο της μιας μεταβλητής (μεσαία επίπεδα άγχους), δεν είναι κατάλληλη η εφαρμογή του συντελεστή συσχέτισης Pearson ο οποίος απαιτεί την ευθύγραμμη σχέση μεταξύ των δύο μεταβλητών.

Η παραπάνω καμπυλόγραμμη σχέση δεν είναι η μοναδική. Η διερεύνηση μη γραμμικών σχέσεων δεν θα μας απασχολήσει. Για όσους ενδιαφέρονται περισσότερο, το SPSS μπορεί πολύ εύκολα να μοντελοποιήσει μια σειρά μη γραμμικών σχέσεων όπως δεύτερη και τρίτη δύναμη, εκθετικές, κλπ. Αυτή η σειρά αναλύσεων ανήκει στο μενού "Regression-Curve Estimation". Εναλλακτικά, οι σχέσεις αυτές μπορούν να μοντελοποιηθούν απευθείας στο διάγραμμα σκεδασμού (fit functions).



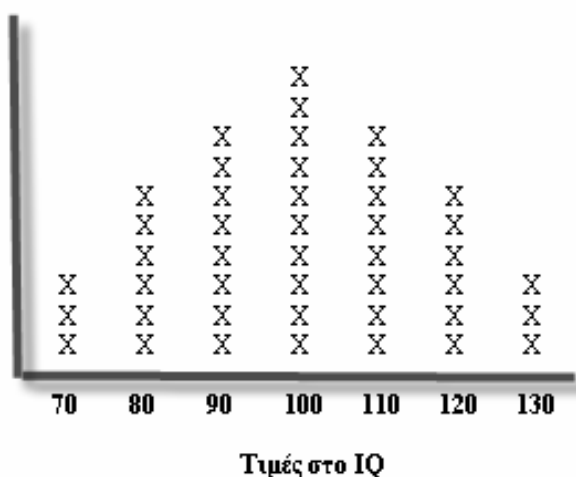
Εικόνα 14: Διάγραμμα σκεδασμού που εκφράζει τη σχέση μεταξύ άγχους και επίδοσης με την εφαρμογή προβλεπτικών γραμμών.

2.3. Η προϋπόθεση της κανονικής κατανομής

Προκειμένου να είναι έγκυρα τα αποτελέσματα του συντελεστή συσχέτισης θα πρέπει οι παρατηρήσεις (δηλ. οι επιμέρους τιμές των συμμετεχόντων) σε κάθε μια από τις δύο μεταβλητές να κατανέμονται κανονικά. Αν και η έννοια της κανονικής κατανομής είναι προαπαιτούμενη της κατανόησης της συσχέτισης ακολουθεί μια συνοπτική συζήτηση ώστε να γίνει κατανοητή η προϋπόθεση.

Η κανονική κατανομή αποτελεί ένα διάγραμμα συχνοτήτων και θα πρέπει να ακολουθεί ένα σχήμα σαν αυτό της "καμπάνας", του οποίου το περίγραμμα εκφράζει τον αριθμό των περιπτώσεων που εμπίπτουν σε κάθε επίπεδο της μεταβλητής που μελετάμε. Έτσι για παράδειγμα, αν θέλουμε να δούμε πως κατανέμεται ο δείκτης ευφυΐας¹⁴ θα περίμενε κανείς να δει μια εικόνα *περίπου* όπως την παρακάτω, γνωρίζοντας ότι ο μέσος όρος ευφυΐας με τη χρήση σταθμισμένων τεστ είναι 100 και η τυπική απόκλιση περίπου 15:

¹⁴ Με δεδομένο ότι αποδεχόμαστε τη σχετική με το δείκτη συζήτηση.



Εικόνα 15: Διάγραμμα συχνοτήτων όπου δείχνει μια υποθετική κανονική κατανομή.

Από την Εικόνα 15 φαίνεται ότι οι περισσότερες παρατηρήσεις (δηλ. συμμετέχοντες) βρίσκονται κοντά στο μέσο όρο (μεταξύ 90-100) και ότι όσο απομακρυνόμαστε από το μέσο όρο τόσο μειώνεται και ο αριθμός των παρατηρήσεων. Δηλαδή υπάρχουν λιγότερα άτομα που είτε είναι πάρα πολύ έξυπνα ή το αντίθετο, ενώ για τους περισσότερους από εμάς περιμένουμε (και ελπίζουμε) ότι (θα) είμαστε κοντά στο μέσο όρο. Η μη-τήρηση αυτής της προϋπόθεσης έχει επίσης σημαντικές (αρνητικές) επιπτώσεις στον υπολογισμό του συντελεστή συσχέτισης, οι οποίες γίνονται φανερές αν μειωθεί πολύ το εύρος κάποιας από τις δύο μεταβλητές που μελετώνται ή και των δύο. Για παράδειγμα, αν φανταστούμε στην Εικόνα 15 ότι όλες οι παρατηρήσεις είναι μεταξύ 120 και 130 (επειδή π.χ., κάναμε μετρήσεις IQ σε πολύ έξυπνα άτομα), τότε δεν θα υπάρχει αρκετό εύρος τιμών για να μοντελοποιηθεί και να αναδειχθεί πιθανή σχέση μεταξύ των δύο μεταβλητών. Το ίδιο θα συνέβαινε αν οι περισσότερες παρατηρήσεις είχαν ομαδοποιηθεί σε χαμηλά επίπεδα διαβάθμισης της ευφυΐας (π.χ., μεταξύ 70-80).

Για την εκτίμηση της κανονικότητας προτείνεται η χρήση του (στατιστικού) δείκτη Kolmogorov-Smirnov. Ο υπολογισμός του προκύπτει με χρήση των μη παραμετρικών τεστ του SPSS και εξηγείται αναλυτικότερα στο Παράρτημα Α. Στην παρούσα φάση, μια απλή οπτική αξιολόγηση των πινάκων συχνοτήτων μπορεί να προσφέρει μια σχετικά σαφή εικόνα για το πόσο η κατανομή που μας ενδιαφέρει προσεγγίζει την κανονική κατανομή.

2.4. Υπολογισμός και ερμηνεία του συντελεστή συσχέτισης Pearson

Τι πρέπει να γνωρίζουμε για το συντελεστή συσχέτισης Pearson r ;

Πρώτον ότι ανήκει στην κατηγορία των σταθμισμένων δεικτών, δηλαδή είναι στατιστικός δείκτης που έχει συγκεκριμένο εύρος με αποτέλεσμα να έχει νόημα από μόνος του και χωρίς τη σύνδεση του με επίπεδα σημαντικότητας ή άλλες πληροφορίες. Το εύρος που μπορεί να πάρει ο r είναι από -1 έως και $+1$. Το πρώτο μέγεθος μας δείχνει μια τέλεια αρνητική σχέση (αντιστρόφως ανάλογη σχέση), ενώ το δεύτερο μια τέλεια θετική (ευθέως ανάλογη) σχέση μεταξύ των δύο μεταβλητών. Το μηδέν αντιπροσωπεύει μηδενική σχέση μεταξύ δύο μεταβλητών. Είναι φυσικό ότι μια σχέση του τύπου 0.6 είναι ταυτόσημη σε μέγεθος με μια σχέση -0.6 , με τη διαφορά ότι η μια είναι θετική ενώ η άλλη αρνητική. Επίσης, η παρουσία σχέσεων μεγαλύτερων του 1.0 είναι κάτι αδύνατο/άτοπο (και άρα λάθος). Μεταξύ των 2 ορίων (-1 έως $+1$), έχουν προταθεί διάφορες κατηγοριοποιήσεις που αναφέρονται στην αξιολόγηση του μεγέθους του δείκτη. Μια από αυτές είναι και η παρακάτω:

0.00-0.20	Μηδενική σχέση
0.21-0.40	Μικρή σχέση
0.41-0.60	Μέτρια σχέση
0.61-0.80	Δυνατή σχέση
> 0.81	Εξαιρετικά δυνατή σχέση

Αν και στην πραγματικότητα σχέσεις της τάξης του 0.80 είναι δύσκολο να παρατηρηθούν, όταν συμβαίνει αυτό δημιουργείται και ένας θεωρητικός προβληματισμός: Μήπως αυτές οι δύο μεταβλητές που συσχετίζονται κατά 0.80 μετρούν το ίδιο θεωρητικά αντικείμενο/φαινόμενο; Επομένως, οι πολύ δυνατές σχέσεις θέλουν προσοχή στην αξιολόγηση και στην ερμηνεία τους.

Πριν την εποχή των υπολογιστών όλες οι πράξεις γίνονταν με το χέρι με αποτέλεσμα να απαιτείται και η κατανόηση των μαθηματικών μοντέλων. Σήμερα αυτό δεν είναι απαραίτητο αφού η χρήση στατιστικών πακέτων έχει αντικαταστήσει αυτή την ανάγκη. Ο τύπος για τον υπολογισμό του συντελεστή συσχέτισης Pearson r έχει ως εξής:

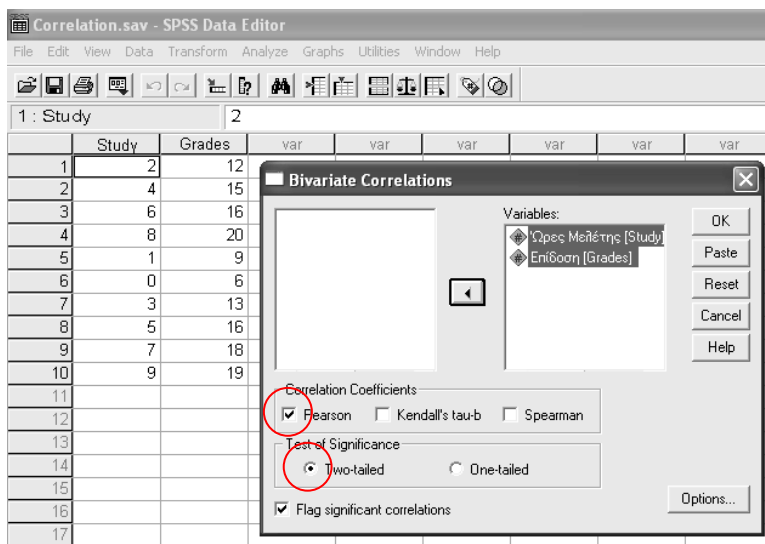
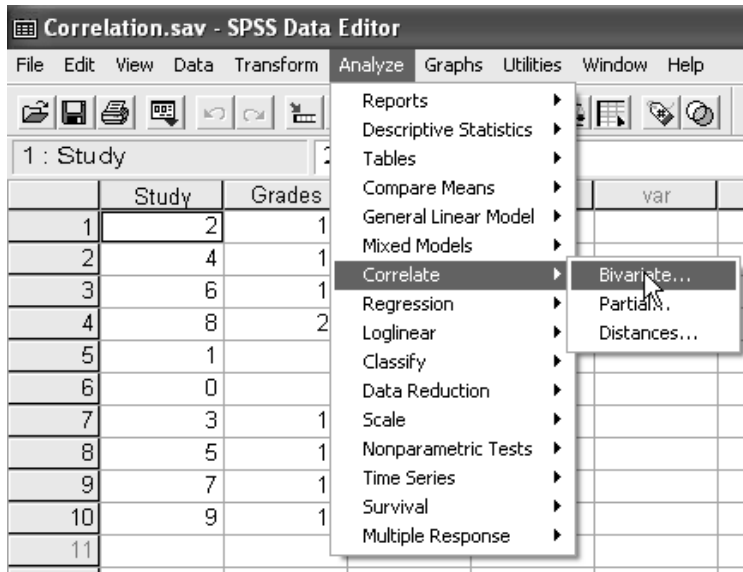
$$r = \frac{\sum(X - \mu_x)(Y - \mu_y)}{N\sigma_x\sigma_y}$$

Οι τιμές X και Y αναφέρονται στις τιμές κάθε συμμετέχοντα, ενώ οι τιμές μ_x και μ_y με τους μέσους όρους στις μεταβλητές X και Y . Επομένως η τιμή κάθε συμμετέχοντα στην

μεταβλητή X αφαιρείται από το μέσο όρο της μεταβλητής και όλες οι τιμές προστίθενται. Το ίδιο γίνεται και για τη μεταβλητή Y . Το γεγονός ότι αυτά τα ζευγάρια τιμών πολλαπλασιάζονται μεταξύ τους εκφράζει τη "συνδιακύμανση" των δύο μεταβλητών. Έτσι για παράδειγμα, αν υπάρχει αρνητική σχέση μεταξύ κατάθλιψης και επίδοσης στο σχολείο, τότε ένα άτομο που θα έχει τιμή μεγαλύτερη από το μέσο όρο στην κατάθλιψη, το πιθανότερο είναι ότι θα έχει μικρότερη τιμή από το μέσο όρο στη μεταβλητή επίδοση. Το γινόμενο θα είναι αρνητικό, και αν αυτό συμβαίνει για τους περισσότερους μαθητές τότε θα έχουμε αρνητική σχέση μεταξύ κατάθλιψης και επίδοσης. Ο παρανομαστής του κλάσματος εκφράζει το γινόμενο των δύο σφαλμάτων, το οποίο διορθώνεται για το μέγεθος του δείγματος. Υπάρχουν πολύ πιο εύκολες τυπολογίες για τον υπολογισμό του συντελεστή συσχέτισης με το χέρι, όμως η παραπάνω εκτιμήθηκε ως η καλύτερη για την εξήγηση του τι ακριβώς υπολογίζει ο συντελεστής συσχέτισης, που δεν είναι τίποτε άλλο από την ύπαρξη της συνδιακύμανσης μεταξύ δύο μεταβλητών.

Για τον υπολογισμό του συντελεστή συσχέτισης με το SPSS επιλέγουμε Analyze → Correlate → Bivariate. Στη σχετική οθόνη (βλ. εικόνες 16 & 17), επιλέγουμε τις μεταβλητές που θέλουμε να συσχετίσουμε, ενώ ο συντελεστής συσχέτισης Pearson r και η δίπλευρη εξέταση¹⁵ έχουν προεπιλεγεί. Για τη διερεύνηση μη-παραμετρικών σχέσεων μπορούν να επιλεγούν οι δείκτες των Spearman και Kendall.

¹⁵ Το μονόπλευρο τεστ χρησιμοποιείται στην περίπτωση που ο ερευνητής αξιολογεί μια συγκεκριμένη υπόθεση η οποία έχει κατεύθυνση. Για παράδειγμα η υπόθεση: «Υπάρχει σχέση μεταξύ προσκόλλησης και επίδοσης» δεν δηλώνει κάποια κατεύθυνση και ο ερευνητής οφείλει να διερευνήσει το είδος της σχέσης και στις δύο πλευρές της κατανομής (αν δηλ. η σχέση είναι θετική ή αρνητική). Στην διερεύνηση όμως για παράδειγμα της σχέσης άγχους και κατάθλιψης, ο ερευνητής είναι σχεδόν σίγουρος ότι η σχέση είναι θετική και αυτό που τον απασχολεί είναι το μέγεθος αυτής της σχέσης. Έτσι, μπορεί να επιλέξει τη χρήση του μονόπλευρου τεστ. Καλό όμως είναι να χρησιμοποιείται το δίπλευρο τεστ, αφού το μονόπλευρο εκφράζει κάποια προκατάληψη από την πλευρά του ερευνητή.



Εικόνες 16 & 17: Υπολογισμός του συντελεστή συσχέτισης Pearson r στο SPSS.

Correlations

		Ωρες Μελέτης	Επίδοση
Ωρες Μελέτης	Pearson Correlation	1	.964**
	Sig. (2-tailed)		.000
	N	10	10
Επίδοση	Pearson Correlation	.964**	1
	Sig. (2-tailed)	.000	
	N	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

Μέγεθος σχέσης
Πιθανότητα
Μέγεθος Δείγματος

Εικόνα 18: Αποτελέσματα από τη χρήση του συντελεστή συσχέτισης Pearson r στο SPSS.

Από τα παραπάνω φαίνεται πως η σχέση μεταξύ ωρών διαβάσματος και επίδοσης για τους 10 μαθητές που αξιολογήθηκαν ήταν 0.964 η οποία είναι στατιστικά σημαντική όπως φαίνεται

και από την παρατηρούμενη πιθανότητα: $p = .000$. Καθότι πιθανότητα ίση με το μηδέν δεν υφίσταται, η παραπάνω πιθανότητα είναι απλά μικρότερη από 1%. Αφού η πιθανότητα είναι μικρότερη από το συνηθισμένο επίπεδο στατιστικής σημαντικότητας $\alpha = 5\%$ (ή όποιο άλλο επίπεδο ο ερευνητής έχει υιοθετήσει), δεχόμαστε την εναλλακτική υπόθεση σύμφωνα με την οποία υπάρχει στατιστικά σημαντική συνάφεια μεταξύ “ωρών μελέτης” και της “επίδοσης”.

Η παρατηρούμενη σχέση στο παράδειγμα ήταν τόσο ισχυρή, ώστε εντοπίσθηκε παρόλο το μικρό μέγεθος του δείγματος. Χωρίς να μπορούμε σε μαθηματικές λεπτομέρειες τονίζουμε ότι τόσο τα πολύ μικρά (μικρότερα του 10 περίπου) όσο και τα πολύ μεγάλα δείγματα (μεγαλύτερα του 500 περίπου), δημιουργούν προβλήματα στην αξιοπιστία της στατιστικής σημαντικότητας του συντελεστή συσχέτισης. Σε παρόμοιες περιπτώσεις (οι πιο πολλές εκ των οποίων εμφανίζονται σε πολύ μεγάλα δείγματα) καλό είναι να περιοριζόμαστε σε μια ποιοτική ερμηνεία της τιμής του συντελεστή συσχέτισης (όπως αυτή που περιγράφεται στην αρχή της παραγράφου) και να αγνοείται η πιθανότητα. Πάντως, για την σωστή διερεύνηση των σχέσεων, ο ερευνητής θα πρέπει να επιλέξει έναν ιδανικό αριθμό συμμετεχόντων προκειμένου να μπορούν να επαληθευτούν εναλλακτικές υποθέσεις που είναι αληθείς.

2.5. Μερικές παρατηρήσεις για το συντελεστή συσχέτισης Pearson

Τι πρέπει να έχετε υπόψη σας/να προσέχετε όταν χρησιμοποιείτε τον συντελεστή συσχέτισης Pearson r :

1. Επηρεάζεται από ακραίες παρατηρήσεις. Αν έχετε άτομα που έχουν πολύ μεγάλες ή πολύ μικρές τιμές είναι πολύ πιθανό αυτές να (δια)στρεβλώνουν σημαντικά το μέγεθος του συντελεστή συσχέτισης.
2. Επηρεάζεται, όπως και οι περισσότεροι δείκτες, από το μέγεθος του δείγματος. Πολύ μικρά ή πολύ μεγάλα δείγματα δημιουργούν προβλήματα στην ερμηνεία της πιθανότητας και της στατιστικής σημαντικότητας.
3. Η αξιοπιστία του δείκτη "πλήττειται" από τη μη-τήρηση της προϋπόθεσης αναφορικά με τη γραμμικότητα της σχέσης. Εξετάστε τη γραμμικότητα (τουλάχιστον) από το διάγραμμα σκεδασμού για την ύπαρξη μη γραμμικών σχέσεων.
4. Επηρεάζεται, όπως και οι περισσότεροι δείκτες, από την αξιοπιστία των μετρήσεων. Αν οι μετρήσεις έχουν μεγάλες τιμές σε στατιστικό σφάλμα, το πιθανότερο είναι ότι το σφάλμα αυτό θα "φορτίσει" και τον συντελεστή συσχέτισης με απρόβλεπτες συνέπειες (συνήθως συμπιέζει τις τιμές του δείκτη προς τα κάτω¹⁶).
5. Είναι ακατάλληλος όταν οι μεταβλητές δεν είναι συνεχείς αλλά διακριτές. Στην δεύτερη περίπτωση συντελεστές όπως ο Φ ή ο point-biserial είναι καταλληλότεροι, ανάλογα με το αν η μία ή και οι δύο μεταβλητές είναι κατηγορικές.
6. Επηρεάζεται από την ύπαρξη στρεβλών κατανομών, ειδικά όταν αυτές αποκλίνουν κατά πολύ από την κανονικότητα. Αν καταπατείται κάποια από τις προϋποθέσεις εξετάστε¹⁷ την χρήση μη παραμετρικών κριτηρίων, τα οποία συζητώνται περιεκτικά παρακάτω.
7. Δεν μπορεί να χρησιμοποιηθεί για τη διερεύνηση αιτιωδών σχέσεων. Αυτό δεν σημαίνει ότι δεν υπάρχουν αιτιακές σχέσεις μεταξύ των μεταβλητών που μελετήθηκαν, αλλά ότι η χρήση του συντελεστή συσχέτισης δεν επιτρέπει αυτό το συμπέρασμα. Το μόνο συμπέρασμα που μπορεί να προκύψει από τη χρήση του συντελεστή συσχέτισης είναι ότι οι δύο μεταβλητές συνδιακυμαίνονται. Μόνο η μελλοντική χρήση

¹⁶ Υπάρχουν πολλοί λόγοι που αυξάνουν το σφάλμα της μέτρησης, άλλοι στατιστικοί και άλλοι μεθοδολογικοί. Το πιθανότερο είναι ότι αύξηση στο μέγεθος τυχαίων παραγόντων και αναξιοπιστίας του τεστ θα μειώσουν τον συντελεστή συσχέτισης, ενώ αύξηση του μεγέθους συστηματικού σφάλματος μπορεί να αυξήσει τον συντελεστή εξαιτίας της μείωσης των ατομικών διαφορών (τυπικής απόκλισης).

¹⁷ Η καταπάτηση των προϋποθέσεων δεν επιβάλλει την χρήση μη-παραμετρικών κριτηρίων. Η απόφαση αυτή θα πρέπει να είναι αποτέλεσμα προσεκτικής σκέψης μιας και η καταπάτηση των προϋποθέσεων μπορεί επίσης να οφείλεται (και) σε άλλους λόγους.

πειραματικών σχεδίων μπορεί να επιβεβαιώσει αν οι αρχικές αυτές συνάψεις έχουν αιτιακό χαρακτήρα.

2.6. Μη-παραμετρικοί συντελεστές συσχέτισης (Spearman)

Όταν οι παραμετρικές προϋποθέσεις δεν ικανοποιούνται (δηλ. η κανονικότητα και η γραμμικότητα, το εύρος των παρατηρήσεων και η ύπαρξη ισοδιαστημικής κλίμακας), τότε πρέπει να χρησιμοποιηθούν εναλλακτικοί στατιστικοί δείκτες για την ανίχνευση σχέσεων μεταξύ μεταβλητών. Ένας από αυτούς είναι και ο δείκτης συσχέτισης του Spearman, που υπολογίζεται μετατρέποντας τα δεδομένα σε "σειρές" με βάση το μέγεθος τους (δηλ. οι αρχικές τιμές μπαίνουν σε σειρά με βάση το μέγεθός τους: πρώτος, δεύτερος, τρίτος, κλπ.). Αυτό έχει σαν αποτέλεσμα οι αποστάσεις μεταξύ των παρατηρήσεων να χάνουν τη σημασία τους και να αξιολογείται η σειρά των συμμετεχόντων στην πρώτη μεταβλητή σε σχέση με τη σειρά που αυτοί έχουν στην δεύτερη μεταβλητή, κ.ό.κ. Το μέγεθος της συμφωνίας ή όχι της σειράς στις δύο μεταβλητές εκφράζει και το πρόσημο αλλά και το μέγεθος της σχέσης. Για παράδειγμα, αν κάποιος που είναι πρώτος στις επιδόσεις είναι και πρώτος στη δημοτικότητα και ακολουθείται από άτομα μικρότερων επιδόσεων και μικρότερης δημοτικότητας, θα αναδείξει μια θετική συνάφεια μεταξύ επίδοσης και δημοτικότητας.

Η εξίσωση για τον υπολογισμό του συντελεστή συσχέτισης Spearman έχει ως εξής:

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

όπου N είναι ο αριθμός των ζευγαριών και D η διαφορά στη σειρά μεταξύ πρώτης και δεύτερης μέτρησης (ζεύγη μετρήσεων). Στο SPSS, η χρήση του δείκτη Spearman είναι πολύ απλή αφού το μόνο που χρειάζεται είναι η επιλογή "Spearman" στο μενού Correlate → Bivariate. Άλλος σχετικός μη παραμετρικός δείκτης είναι ο "τ" του Kendall ο οποίος όμως δε θα συζητηθεί επί του παρόντος.

Στο παραπάνω παράδειγμα η εφαρμογή του συντελεστή Spearman έδωσε τα εξής αποτελέσματα:

Correlations

		Study	Grades
Spearman's rho	Study	1.000	.985**
	Correlation Coefficient		
	Sig. (2-tailed)	.	.000
	N	10	10
	Grades	.985**	1.000
	Correlation Coefficient		
	Sig. (2-tailed)	.000	.
	N	10	10

** . Correlation is significant at the 0.01 level (2-tailed).

Εικόνα 19: Συσχέτιση μεταξύ ωρών μελέτης και επίδοσης με το δείκτη Spearman.

Όπως φαίνεται στην Εικόνα 19, η συνάφεια μεταξύ των δύο μεταβλητών ήταν επίσης πολύ δυνατή $r = .985$ και στατιστικά σημαντική ($p < .001$). Δεν περιμένουμε να δούμε πολύ μεγάλες διαφορές μεταξύ των 2 δεικτών, ειδικά αν η καταπάτηση των προϋποθέσεων δεν έχει γίνει σε μεγάλο βαθμό. Στην παρούσα περίπτωση ο δείκτης Spearman είναι λίγο μεγαλύτερος από τον δείκτη Pearson, αν και συνήθως περιμένουμε πιο "συντηρητικά" ευρήματα από τη χρήση μη παραμετρικών δοκιμασιών¹⁸.

¹⁸ Συνήθως οι μη-παραμετρικοί δείκτες είναι πιο «αυστηροί» στα αποτελέσματά τους αφού στόχο έχουν να διορθώσουν για πιθανά προβλήματα που προκαλούνται από τις καταπατήσεις των προϋποθέσεων. Αυτό βέβαια δεν συμβαίνει συνέχεια όπως φάνηκε και παραπάνω.

3

ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

3. ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Μπορεί το οικογενειακό περιβάλλον να προβλέψει τις επιδόσεις, τις συμπεριφορές ή τις διαπροσωπικές σχέσεις των παιδιών στο σχολείο; Μπορεί ο χρόνος διαβάσματος να προβλέψει ή να ερμηνεύσει την επίδοση των μαθητών σε ένα συγκεκριμένο γνωστικό αντικείμενο; Οι απαντήσεις στα παραπάνω ερωτήματα είναι εξαιρετικά σημαντικές για τη σχολική εμπειρία και διαδρομή των παιδιών/μαθητών και των εκπαιδευτικών. Η σημασία που αποδίδεται στη χρήση των προβλεπτικών μοντέλων υπαγόρευσε την έρευνα και τη βελτίωσή τους¹⁹.

Το πιο απλό προβλεπτικό μοντέλο είναι η **ανάλυση παλινδρόμησης** (regression analysis), η οποία αποτελεί φυσική επέκταση του συντελεστή συσχέτισης. Οι διαφορές με το συντελεστή συσχέτισης είναι οι εξής:

- (α) χρησιμοποιείται για την πρόβλεψη των τιμών ορισμένων μεταβλητών και
- (β) παρέχεται η δυνατότητα συνεκτίμησης περισσότερων από δύο ανεξάρτητων μεταβλητών

Ας εξετάσουμε το ενδεχόμενο μιας απλής πρόβλεψης. Θέλουμε να εξετάσουμε αν οι ώρες διαβάσματος είναι σε θέση να προβλέψουν την επίδοση των μαθητών στο σχολείο. Ως εργαλείο μέτρησης επιλέγουμε να χρησιμοποιηθεί ένα κριτήριο αξιολόγησης μαθηματικών γνώσεων, το οποίο αποτελείται από 30 ερωτήσεις - ασκήσεις. Τα δεδομένα προέρχονται από 19 μαθητές μιας τάξης και οι υποθέσεις έχουν ως εξής²⁰:

- H_0 : Οι ώρες διαβάσματος **δεν** προβλέπουν την επίδοση
- H_e : Οι ώρες διαβάσματος προβλέπουν την επίδοση

Τα δεδομένα από τους 19 συμμετέχοντες και έχουν ως εξής:

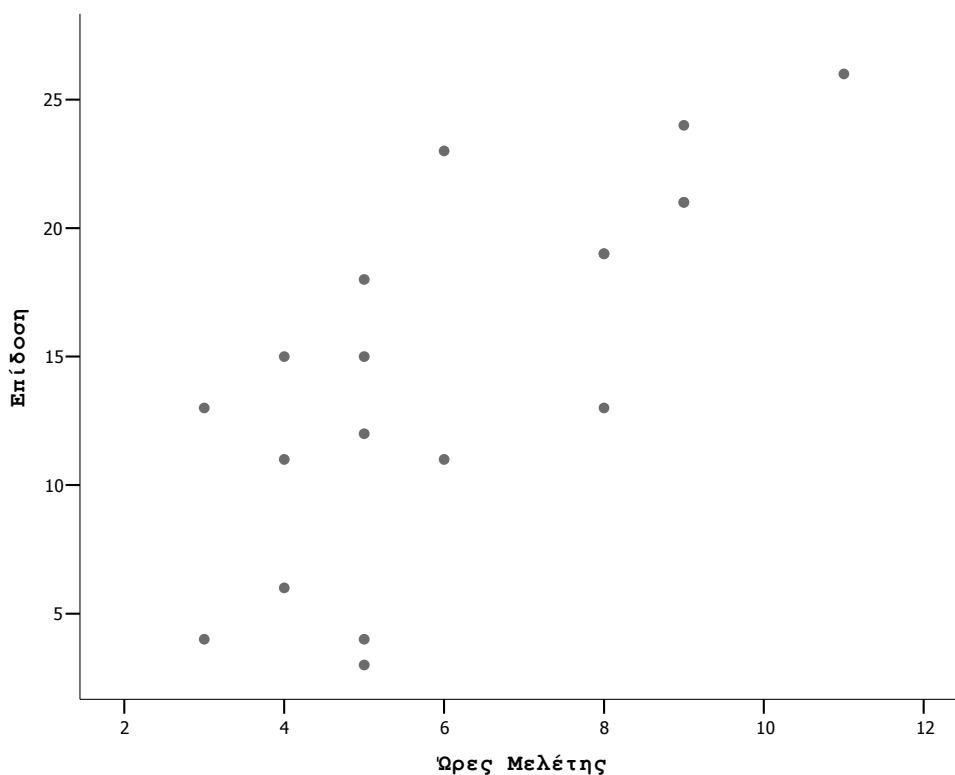
¹⁹ Η ανάλυση συσχέτισης για την οποία έγινε λόγος παραπάνω, δεν μπορεί να απαντήσει σε ερωτήματα αναφορικά με την πρόβλεψη, επισημαίνει απλώς τη συσχέτιση μεταξύ δύο μεταβλητών.

²⁰ Οι χειριστές των στατιστικών αναλύσεων θα πρέπει να συνηθίσουν στην ιδέα ότι η στατιστική είναι επικουρική και μόνο της ερευνητικής διαδικασίας και έχει νόημα μόνο για την απάντηση συγκεκριμένων ερευνητικών ερωτήσεων. Επομένως, πρώτα θα πρέπει να γίνει η διατύπωση των υποθέσεων και μετά η επιλογή του κατάλληλου στατιστικού κριτηρίου που θα δώσει απάντηση στη (συγκεκριμένη κάθε φορά) υπόθεση εργασίας.

Πίνακας 2: Δεδομένα από μια υποθετική μελέτη αναφορικά με την πρόβλεψη της επίδοσης στα μαθηματικά ως συνάρτηση των ωρών διαβάσματος

Συμμετέχοντες	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Ώρες διαβάσματος	11	9	9	9	8	8	8	6	6	5	5	5	5	5	4	4	4	3	3
Επίδοση	26	21	24	21	19	13	19	11	23	15	4	18	12	3	11	15	6	13	4

Το πρώτο βήμα, όπως και στην ανάλυση συσχέτισης, είναι η δημιουργία του διαγράμματος σκεδασμού το οποίο, πέρα από την οπτική ανίχνευση της ύπαρξης σχέσης, θα μας δώσει πληροφορίες και για την ύπαρξη ή όχι γραμμικότητας. Ακολουθώντας τα βήματα που περιγράφηκαν στην ανάλυση συσχέτισης κατά Pearson, δημιουργούμε το παρακάτω διάγραμμα σκεδασμού.



Εικόνα 20: Διάγραμμα σκεδασμού για την πρόβλεψη της σχολικής επίδοσης από τις ώρες διαβάσματος.

Επισημαίνεται ότι οι παραπάνω υποθέσεις υποδηλώνουν ποια είναι η εξαρτημένη μεταβλητή (αποτέλεσμα) και ποια η ανεξάρτητη²¹ (προβλεπτικός παράγοντας). Επομένως στη δημιουργία του

²¹ Όπως ειπώθηκε πριν, στην ανάλυση συσχέτισης η συνάφεια A-B είναι ίδιας "αξίας" με τη συνάφεια B-A από τη στιγμή που το ενδιαφέρον δεν είναι στο αν η μια μεταβλητή επιδρά στην άλλη, αλλά απλά στο αν οι δύο μεταβλητές "κινούνται" παράλληλα.

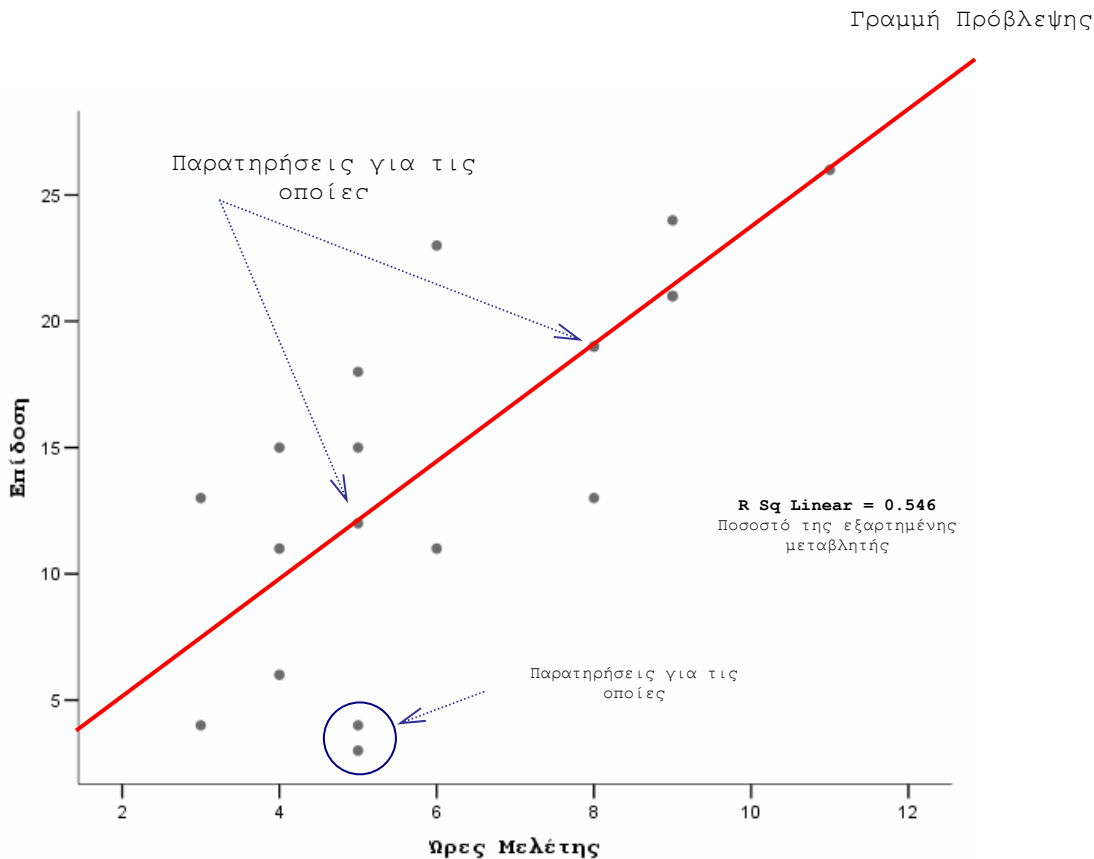
διαγράμματος σκεδασμού τοποθετούμε την εξαρτημένη μεταβλητή στον κάθετο άξονα (Y) και την ανεξάρτητη στον οριζόντιο (X).

Από την Εικόνα 15 είναι φανερό ότι υπάρχει θετική συνάφεια μεταξύ των δύο μεταβλητών από τη στιγμή που οι υψηλές τιμές της μιας φαίνεται να σχετίζονται με τις υψηλές τιμές της άλλης. Επομένως η προϋπόθεση της γραμμικότητας φαίνεται να ικανοποιείται. Αναφορικά με την "προϋπόθεση κανονικότητας", χρησιμοποιώντας το Kolmogorov-Smirnov test, διαπιστώνουμε ότι σε καμιά από τις δύο μεταβλητές (επίδοση ή ώρες διαβάσματος) δεν ακυρώνεται η συγκεκριμένη προϋπόθεση. Ο δείκτης K-S για τις ώρες μελέτης είναι 0.941 ενώ για την επίδοση 0.457 και κανένας τους δεν ξεπερνά τα επίπεδα στατιστικής σημαντικότητας (πιθανότητες 0.34 και 0.99, αντίστοιχα). Από τη στιγμή που καμιά πιθανότητα δεν είναι μικρότερη από το 5%, τότε αυτόματα δεχόμαστε τη μηδενική υπόθεση που υποστηρίζει ότι οι υπό εξέταση κατανομές (της μεταβλητής "επίδοση" και της μεταβλητής "ώρες διαβάσματος") δεν διαφέρουν από την (υποθετική) κανονική κατανομή. Έτσι λοιπόν μπορούμε να προχωρήσουμε στην ανάλυση παλινδρόμησης αφού ικανοποιούνται και οι δύο προϋποθέσεις.

Το επόμενο βήμα είναι να εφαρμόσουμε το μοντέλο της ευθείας γραμμής με πρόθεση την πρόβλεψη της σχέσης ανάμεσα στην επίδοση και τις ώρες διαβάσματος. Το μοντέλο αυτό συμβολίζεται με τη συνάρτηση:

$$\hat{Y} = bX + a$$

όπου \hat{Y} είναι η προβλεπτική τιμή επίδοσης, η οποία είναι συνάρτηση ενός συντελεστή κλίσης (b) και μιας σταθεράς (α). Τα σημεία α και b αποτελούν τα δύο αναγκαία σημεία προκειμένου να χαραχθεί μια και μοναδική ευθεία γραμμή η οποία ταιριάζει όσο το δυνατόν καλύτερα στα δεδομένα μας. Στην ιδεατή περίπτωση θα έπρεπε όλες οι παρατηρήσεις να είναι πάνω στην ευθεία γραμμή, αν όμως αυτό δε συμβαίνει, τότε θα έπρεπε (προκειμένου να ισχύει η ευθύγραμμη σχέση), όσο το δυνατόν περισσότερες παρατηρήσεις να είναι πάνω στη γραμμή και οι υπόλοιπες να βρίσκονται όσο το δυνατόν πιο κοντά της (προς τα πάνω ή προς τα κάτω). Το α (η σταθερά) της συνάρτησης είναι το σημείο στο οποίο η ευθεία γραμμή της πρόβλεψης τέμνει τον κάθετο άξονα (αναφέρεται δηλαδή στην τιμή του Y όταν η τιμή του X είναι ίση με μηδέν).



Εικόνα 21: Πρόβλεψη της επίδοσης από τις ώρες διαβάσματος.

Όπως φαίνεται στην Εικόνα 21 τέσσερις περίπου παρατηρήσεις εφάπτονται με τη γραμμή της πρόβλεψης, που σημαίνει ότι για αυτά τα άτομα η πρόβλεψη θα είναι εξαιρετικά ακριβής. Όσο μεγαλύτερη είναι η απόσταση κάποιων παρατηρήσεων από την ευθεία γραμμή, τόσο περισσότερο αποτυγχάνει το μοντέλο πρόβλεψης για τις (συγκεκριμένες) παρατηρήσεις (η πρόβλεψη βασίζεται στην "επιτυχία" της συγκεκριμένης ευθείας γραμμής να περιγράψει τα δεδομένα²²). Οι αποστάσεις των παρατηρήσεων από την ευθεία γραμμή αναφέρονται επίσης ως υπόλοιπα (residuals), αφού δεν αποτελούν τίποτε άλλο από την απόκλιση μεταξύ πραγματικών και αναμενόμενων (προβλεπόμενων) τιμών. Όσο μεγαλύτερες είναι οι αποστάσεις, τόσο μεγαλύτερη και η απόκλιση μεταξύ πραγματικότητας και πρόβλεψης. Για παράδειγμα για τον μαθητή/τρια που έχει διαβάσει 5 ώρες (σημείο A), προβλέπουμε ότι θα έχει επίδοση ίση με το 12 (αν υπολογίσουμε τη νοητή ευθεία της συγκεκριμένης παρατήρησης προς την γραμμή πρόβλεψης και από τη γραμμή πρόβλεψης στον άξονα Y). Είναι

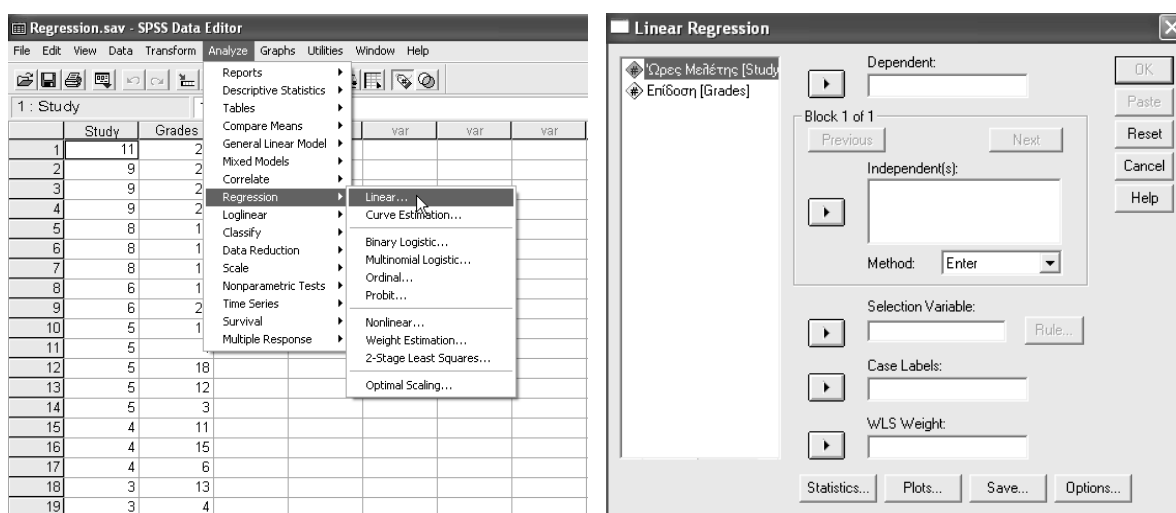
²² Κάποιος θα μπορούσε πολύ λογικά να ισχυριστεί ότι το πρώτο βήμα πριν την αξιολόγηση της πρόβλεψης είναι να διερευνηθεί αν υπάρχει σχέση μεταξύ των δύο μεταβλητών αφού η έλλειψη σχέσης δηλώνει και απουσία πρόβλεψης. Στο συγκεκριμένο παράδειγμα, η σχέση μεταξύ επίδοσης και ωρών διαβάσματος με τη χρήση του Pearson r ήταν 0.739. Το μέγεθος του συντελεστή συσχέτισης είναι αρκετά υψηλό και στατιστικά σημαντικό, ώστε να μπορεί να υπάρχει (και ουσιαστική) πρόβλεψη για τη σχέση των δύο μεταβλητών.

φανερó ότι για τον συγκεκριμένο μαθητή/τρια η πρόβλεψη θα είναι εξαιρετικά φτωχή.

Το παραπάνω παράδειγμα εξετάζει μία μόνο ανεξάρτητη μεταβλητή X (ώρες διαβάματος) σε σχέση με την εξαρτημένη μεταβλητή Y (επίδοση). Το μοντέλο αυτό ονομάζεται «απλή» γραμμική παλινδρόμηση (simple linear regression). Πιο ρεαλιστικά μοντέλα χρησιμοποιούν περισσότερες ανεξάρτητες μεταβλητές και ονομάζονται μοντέλα «πολλαπλής» γραμμικής παλινδρόμησης (multiple linear regression). Στις σημειώσεις θα επικεντρωθούμε στην απλή παλινδρόμηση αλλά η επέκταση της στην πολλαπλή περίπτωση είναι άμεση.

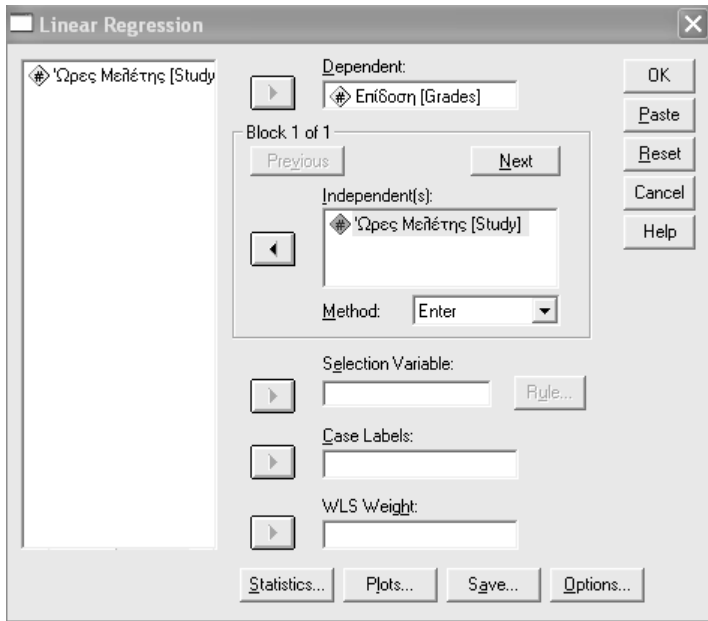
3.1. Υπολογισμός της γραμμής παλινδρόμησης με τη χρήση του SPSS

Για τον υπολογισμό της ανάλυσης παλινδρόμησης επιλέξτε: Analyze → Regression → Linear.



Εικόνες 22α & 22β

Επιλέξτε και τοποθετήστε την εξαρτημένη μεταβλητή (Επίδοση) στη γραμμή Dependent και την ανεξάρτητη μεταβλητή (Ωρες μελέτης) στη γραμμή Independent(s). Επιλέξτε το OK. Για λόγους απλοποίησης δεν απαιτούνται άλλες επιλογές για την βασική κατανόηση της απλής ανάλυσης παλινδρόμησης.



Εικόνα 23: Υπολογισμός της πρόβλεψης της σχολικής επίδοσης από τις ώρες διαβάσματος.

Όπως φαίνεται στην Εικόνα 24, ο πρώτος παραγόμενος πίνακας αναφέρει ορισμένα συγκεντρωτικά δεδομένα για την ύπαρξη πρόβλεψης.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.739 ^a	.546	.519	4.882

a. Predictors: (Constant), Ώρες Μελέτης

Εικόνα 24: Αποτελέσματα από την ανάλυση παλινδρόμησης (1^{ος} Πίνακας).

Έτσι, αναφέρεται ο δείκτης συνάφειας των δύο μεταβλητών Pearson r , ο οποίος στη συγκεκριμένη περίπτωση είναι 0.739, δηλώνοντας μια (ισχυρή) θετική συνάφεια ανάμεσα στις ώρες μελέτης και στην επίδοση των μαθητών. Ο συντελεστής συσχέτισης υπολογισμένος στη δεύτερη δύναμη (τετράγωνο), ορίζει τον "συντελεστή προσδιορισμού" (R Square) και εξηγεί το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής που οφείλεται στην επίδραση της ανεξάρτητης. Το ποσοστό αυτό στο συγκεκριμένο παράδειγμα είναι 0.546, δηλ. ~55%. Με απλά λόγια το συγκεκριμένο εύρημα υποδηλώνει ότι αν γνωρίζουμε πόσες ώρες διαβάζει κάποιος μαθητής/τρια μπορούμε να κατανοήσουμε κατά 55% το φαινόμενο της σχολικής επίδοσης²³. Δηλαδή, η γνώση του

²³ Αυτό **δεν** είναι ισοδύναμο του γεγονότος ότι για το 55% των συμμετεχόντων θα έχουμε ακριβή πρόβλεψη, ενώ για τους υπόλοιπους συμμετέχοντες η πρόβλεψη θα είναι ατυχή.

χρόνου μελέτης φαίνεται να είναι καθοριστικός παράγοντας εκτίμησης της επίδοσης, με την επιφύλαξη (ασφαλώς), ότι η αιτία της επίδοσης δεν είναι απαραίτητα ο χρόνος μελέτης κάποιου μαθητή/τριας. Οι ώρες μελέτης εκτιμάται ότι συμμετέχουν στην εξέταση και ερμηνεία της επίδοσης, αλλά θα πρέπει να δοθεί προσοχή σε δύο παράγοντες: (α) στο γεγονός ότι η διερεύνηση αιτιακών σχέσεων (causal relationships) απαιτεί τη χρήση πειραματικών ερευνητικών σχεδίων, και (β) στο το ότι η μια μεταβλητή εξηγεί το 55% της διακύμανσης μιας άλλης μεταβλητής, δεν σημαίνει ότι αποτελεί και την αιτία αλλαγής της. Στο συγκεκριμένο παράδειγμα οι πιθανές αιτιολογήσεις μπορεί να είναι πολλές. Επομένως, αν διερευνούσαμε την προβλεπτική ικανότητα και άλλων μεταβλητών θα ήταν πιθανό να αξιολογηθεί σε μεγαλύτερο βάθος η σημαντικότητα και η αξία της μεταβλητής "ώρες διαβάσματος" ως προβλεπτικός παράγοντας της επίδοσης, αφού οι ερευνητές θα μπορούσαν να αξιολογήσουν αυτή τη συνεισφορά σαν συνάρτηση της "ταυτόχρονης" συμμετοχής - συνεισφοράς και άλλων μεταβλητών. Για παράδειγμα, η εισαγωγή μιας δεύτερης προβλεπτικής μεταβλητής (π.χ, κίνητρα) θα μπορούσε να μειώσει τα επίπεδα πρόβλεψης της μεταβλητής επίδοση (π.χ., από 55% σε 45%), αφού ένα 10% της διακύμανσης της επίδοσης θα μπορούσε να οφείλεται στα κίνητρα για μάθηση ή επίτευξη.

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	487.252	1	487.252	20.444	.000 ^a
	Residual	405.169	17	23.833		
	Total	892.421	18			

a. Predictors: (Constant), Ωρες Μελέτης

b. Dependent Variable: Επίδοση

Εικόνα 25: Αποτελέσματα από την ανάλυση παλινδρόμησης: Τεστ σημαντικότητας (2^{ος} Πίνακας).

Το Σχήμα 19 αξιολογεί το μέγεθος των δεικτών του Μοντέλου (Σχήμα 18) για το αν (αυτοί) ξεπερνούν ή όχι τα επίπεδα στατιστικής σημαντικότητας. Με άλλα λόγια αξιολογεί την υπόθεση ότι η πρόβλεψη είναι σημαντική ή τυχαία. Στον πίνακα της "ανάλυσης διακύμανσης" (ANOVA) αξιολογείται η πιθανότητα ο συντελεστής συσχέτισης r ή ο συντελεστής R^2 να είναι στατιστικώς σημαντικά διαφορετικοί από το μηδέν. Όπως φαίνεται στο Σχήμα 19, το μέγεθος τους τεστ F^{24} είναι 20.444, το οποίο

²⁴ Το ίδιο τεστ (F) χρησιμοποιείται συνήθως για την ανίχνευση διαφορών δύο ή και περισσότερων μέσων όρων. Το ίδιο ουσιαστικά μαθηματικό μοντέλο χρησιμοποιείται και για τον υπολογισμό της πρόβλεψης.

για 1 και 17 βαθμούς ελευθερίας²⁵ (στον αριθμητή και στον παρανομαστή, αντίστοιχα), σχετίζεται με μια πιθανότητα που είναι μικρότερη από 1%). Γνωρίζοντας ότι ένα στατιστικά σημαντικό εύρημα σημαίνει αυτόματα την υιοθέτηση της εναλλακτικής υπόθεσης, συμπεραίνει κανείς ότι υπάρχει στατιστικά σημαντική συσχέτιση/πρόβλεψη της επίδοσης από τις ώρες μελέτης.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.938	3.229		.290	.775
	Ώρες Μελέτης	2.224	.492	.739	4.522	.000

a. Dependent Variable: Επίδοση

Εικόνα 26: Αποτελέσματα από την ανάλυση παλινδρόμησης: Συντελεστές και προβλεπτική εξίσωση (3^{ος} Πίνακας από το συνολικό output του SPSS).

Στον Πίνακα του Σχήματος 20 εντοπίζουμε τους συντελεστές που ορίζουν την εξίσωση της πρόβλεψης. Αξιοποιούμε την εξίσωση για να προβλέψουμε τις τιμές συγκεκριμένων συμμετεχόντων ή γενικότερα για να κάνουμε προβλέψεις για δεδομένες τιμές X (βλέπε παρακάτω την ενότητα αναφορικά με την “εφαρμογή της προβλεπτικής εξίσωσης”). Από το Σχήμα 20 εντοπίζουμε δύο στατιστικά τεστ που αναφέρονται στη χρήση του t-test²⁶. Έχουμε ήδη επισημάνει ότι ένα στατιστικό τεστ θα πρέπει να εξετάζει μια συγκεκριμένη ερευνητική υπόθεση (μηδενική και εναλλακτική). Το πρώτο τεστ αξιολογεί τις παρακάτω υποθέσεις αναλογικά με την παράμετρο της σταθεράς (constant, α στην εξίσωση), η οποία όπως είπαμε αναφέρεται στην πρόβλεψη της τιμής του Y όταν η τιμή του X είναι μηδέν. Οι υποθέσεις λοιπόν που αξιολογούνται έχουν ως εξής:

²⁵ Σε μια προσπάθεια απλοποίησης ισχύει ότι οι βαθμοί ελευθερίας του αριθμητή αντιστοιχούν στον αριθμό των προβλεπτικών μεταβλητών, ενώ αυτοί του παρανομαστή, περίπου στον αριθμό των παρατηρήσεων (συμμετεχόντων). Φυσικά το παραπάνω αποτελεί υπεραπλούστευση για το τι ακριβώς είναι και τι ρόλο διαδραματίζουν οι βαθμοί ελευθερίας.

²⁶ Το t-test, όπως και το F-test ανήκει στα τεστ διαφορών στη βάση των οποίων αξιολογείται εάν ένας μέσος όρος είναι διαφορετικός από κάποια παράμετρο του πληθυσμού ή αν διαφέρουν μεταξύ τους δύο μέσοι όροι.

H_0 : Η τιμή της σταθεράς είναι ίση με το μηδέν

H_e : Η τιμή της σταθεράς είναι διαφορετική από το μηδέν

Επί τις ουσίας αξιολογείται το αν η τιμή 0.938 είναι ίση με το μηδέν. Από την αξιολόγηση της πιθανότητας, η οποία είναι 77.5%, φαίνεται ότι πρέπει να δεχτούμε τη μηδενική υπόθεση (θα δεχόμασταν την εναλλακτική αν η πιθανότητα ήταν μικρότερη από 5%), η οποία μας λέει ότι η σταθερά δεν είναι στατιστικώς σημαντικά διαφορετική από το μηδέν. Ποια είναι λοιπόν η συνεισφορά του τεστ; Σχεδόν καμία αφού η σταθερά, (αν και συνήθως είναι στατιστικώς σημαντική) δεν συνεισφέρει ουσιαστικές πληροφορίες για το μέγεθος ή την κατεύθυνση της πρόβλεψης. Αντίθετα, η κλίση b μας πληροφορεί για το μέγεθος της πρόβλεψης και το είδος της σχέσης μεταξύ των δύο μεταβλητών. Η κλίση δίνεται από τη στήλη B του Σχήματος 20 μετά τη σταθερά (2,224). Επομένως το δεύτερο τεστ αναφέρεται στις εξής υποθέσεις:

H_0 : Η κλίση b είναι ίση με το μηδέν

H_e : Η κλίση b είναι στατιστικώς σημαντικά διαφορετική από το μηδέν

Επειδή ο στατιστικός δείκτης t με μέγεθος 4.522 σχετίζεται με μια πιθανότητα που είναι μικρότερη από το 5%, δεχόμαστε αυτόματα την εναλλακτική υπόθεση που μας λέει ότι η κλίση είναι στατιστικώς σημαντικά διαφορετική από το μηδέν. Σημαντική κλίση σημαίνει (φυσικά) και σημαντική πρόβλεψη. Επομένως συμπεραίνουμε (και από το συγκεκριμένο τεστ, όπως και από το F-test παραπάνω), ότι υπάρχει μια στατιστικά σημαντική πρόβλεψη της επίδοσης σε συνάρτηση με τις ώρες μελέτης.

Τι σημαίνει ο δείκτης b ; Ο συντελεστής b μας δείχνει πόσο μεταβάλλεται το αποτέλεσμα (στο παράδειγμά μας η επίδοση) για κάθε μια μονάδα αλλαγής της ανεξάρτητης μεταβλητής (δηλαδή των ωρών μελέτης). Η τιμή του b είναι 2.224 που σημαίνει ότι η αύξηση κατά μια και μόνο ώρα μελέτης την ημέρα, σχετίζεται με αύξηση της επίδοσης στα μαθηματικά κατά 2 περίπου μονάδες.

3.2. Εφαρμογή της προβλεπτικής εξίσωσης

Η έννοια της πρόβλεψης έχει αξία αν με πληροφορίες που έχουμε σήμερα μπορούμε να προβλέψουμε, με σχετικά μεγάλη ακρίβεια, ότι κάτι θα συμβεί στο μέλλον. Με τη χρήση λοιπόν της εξίσωσης πρόβλεψης μπορούμε να προβλέψουμε τον αριθμό των ασκήσεων στα μαθηματικά που θα λυθούν αν π.χ., κάποιος μαθητής/τρια διαβάζει 5 ώρες την ημέρα. Με βάση την παρακάτω εξίσωση και εφαρμόζοντας όπου $x = 5$, έχουμε:

$$\hat{Y} = \beta x + \alpha \Rightarrow$$

$$\hat{Y} = .938x + 2.224 \Rightarrow$$

$$\hat{Y} = .938(5) + 2.224 \Rightarrow$$

$$\hat{Y} = 6.914, \sim 7 \text{ ασκήσεις}$$

Δηλαδή, περίπου 7 ασκήσεις στα μαθηματικά είναι συνάρτηση διαβάσματος 5 ωρών την ημέρα. Δεδομένου ότι και η αξιολόγηση στα μαθηματικά δεν ήταν σταθμισμένη, είναι λίγο δύσκολο να αξιολογηθεί αν το 7 είναι μικρός ή μεγάλος αριθμός. Ας δούμε λοιπόν πόσες μαθηματικές ασκήσεις θα λύσει κάποιος μαθητής/τρια που διαβάζει 0 ώρες την ημέρα για τα μαθηματικά. Εφαρμόζοντας την παραπάνω προβλεπτική εξίσωση για $X = 0$, βλέπουμε ότι οι μαθητές αυτοί θα λύσουν μόνο 2 μαθηματικές ασκήσεις. Η διαφορά λοιπόν των 2 από τις 7 ασκήσεις είναι εμφανής. Από την πρόβλεψη αυτή φαίνεται πόσο σημαντικό παράγοντα της μαθηματικής επίδοσης αποτελεί η μελέτη. Όπως αναφέρθηκε και παραπάνω, η εικόνα σχετικά με την προβλεπτική ικανότητα των ωρών διαβάσματος μπορεί να αλλάξει δραματικά από την αξιολόγηση (και) άλλων ανεξάρτητων μεταβλητών.

4

ΕΛΕΓΧΟΣ ΜΕΣΟΥ ΟΡΟΥ

4. ΕΛΕΓΧΟΣ ΓΙΑ ΤΟ ΜΕΣΟ ΟΡΟ

4.1. Εισαγωγή

Ένα από τα πιο ενδιαφέροντα μέτρα θέσης που χαρακτηρίζει έναν πληθυσμό ή ένα δείγμα είναι, χωρίς αμφιβολία, ο μέσος όρος. Στην πλειονότητα όμως των ποσοτικών εμπειρικών ερευνών δεν υπάρχουν στοιχεία από όλο τον πληθυσμό (εκτός και εάν έχουμε απογραφικά δεδομένα). Έτσι, αν μας ενδιαφέρει να εξετάσουμε την τιμή του μέσου όρου ορισμένων χαρακτηριστικών σε έναν ή περισσότερους πληθυσμούς, αυτό μπορεί να γίνει μόνο με βάση τα στοιχεία του δείγματος. Ο σκοπός της παρούσας στατιστικής μεθοδολογίας είναι να παρουσιάσει τις τεχνικές μέσω των οποίων ο ερευνητής μπορεί να εξάγει από το δείγμα ασφαλή συμπεράσματα για τη μέση τιμή ορισμένων χαρακτηριστικών του πληθυσμού. Πριν προχωρήσουμε, θα παρουσιάσουμε ορισμένα παραδείγματα από το χώρο της Εκπαιδευτικής Έρευνας (και όχι μόνο) όπου η παραπάνω μεθοδολογία είναι χρήσιμη.

Παράδειγμα 1α: Γνωρίζουμε ότι το μέσο ποσό που ξοδεύει μηνιαίως για φροντιστήρια μία οικογένεια που έχει ένα παιδί στην Γ' Λυκείου είναι 2000 ευρώ (το νούμερο είναι υποθετικό). Επιθυμούμε να διαπιστώσουμε εάν οι οικογένειες των τελειόφοιτων μαθητών μίας συγκεκριμένης πόλης ξοδεύουν κατά μέσο όρο το ίδιο ή διαφορετικό πόσο για φροντιστήρια. Η σύγκριση αφορά το μηνιαίο μέσο όρο του ποσού που ξοδεύει μία οικογένεια που κατοικεί στη συγκεκριμένη πόλη και έχει ένα παιδί στην Γ' Λυκείου με την προκαθορισμένη γνωστή τιμή των 2000 ευρώ.

Παράδειγμα 2α: Μας ενδιαφέρει να εξετάσουμε την επίδραση που έχει η χρήση των Τεχνολογιών Πληροφορίας & Επικοινωνιών στην εκπαιδευτική διαδικασία. Πιο συγκεκριμένα, επιθυμούμε να διαπιστώσουμε εάν οι μαθητές που διδάσκονται ορισμένα μαθήματα με τη βοήθεια των Τεχνολογιών Πληροφορίας & Επικοινωνιών (χρήση εκπαιδευτικού λογισμικού, αξιοποίηση υπηρεσιών διαδικτύου, κλπ) παρουσιάζουν διαφορές ως προς τη βαθμολογία τους σε σχέση με τους μαθητές που διδάσκονται τα ίδια μαθήματα με τον "συμβατικό" τρόπο. Στην προκειμένη περίπτωση, η σύγκριση γίνεται μεταξύ των μέσων όρων βαθμολογίας δύο διαφορετικών μαθητικών πληθυσμών (ο πληθυσμός που χρησιμοποιεί Τεχνολογίες Πληροφορίας & Επικοινωνιών σε σχέση με τον πληθυσμό που διδάσκεται κάποιο γνωστικό αντικείμενο με τον "συμβατικό" τρόπο διδασκαλίας).

Παράδειγμα 3α: Τα φαινόμενα σχολικής βίας αρχίζουν να εμφανίζονται σε ολοένα και μικρότερη ηλικία. Οι ερευνητές θεωρούν ότι η προβολή βίαιων εικόνων στα παιδιά ηλικίας 12 ετών συμβάλλει σημαντικά σε αυτό. Για το σκοπό αυτό αναπτύχθηκε οπτικοακουστικό υλικό στο οποίο υπήρχαν ποικίλα

ερεθίσματα και εικόνες βίας. Πριν την έκθεση παιδιών ηλικίας 12 ετών στο συγκεκριμένο υλικό καταγράφηκαν τα κρούσματα σχολικής βίας των υπό εξέταση παιδιών κατά το τελευταίο δίμηνο. Στη συνέχεια, μέσω ειδικής εκδήλωσης (προβολή του οπτικοακουστικού υλικού), οι μαθητές ήρθαν σε επαφή με το υλικό βίας και καταγράφηκαν τα κρούσματα βίας των ίδιων μαθητών για το επόμενο δίμηνο. Οι ερευνητές ενδιαφέρονται να εξετάσουν εάν ο μέσος όρος των κρουσμάτων βίας αυξήθηκε μετά την προβολή της ταινίας. Η σύγκριση αφορά το μέσο όρο κρουσμάτων βίας του ίδιου πληθυσμού σε δύο διαφορετικές χρονικές στιγμές²⁷.

Τονίζουμε ότι όλη η διαδικασία που θα περιγραφεί στις επόμενες ενότητες αφορά σε μέσους όρους που προέρχονται από *συνεχή αριθμητικά δεδομένα* (π.χ. βάρος, εισόδημα στις αρχικές μονάδες μέτρησης). Στην περίπτωση των κατηγορικών μεταβλητών για τις οποίες δεν έχει νόημα η διαβάθμιση μεταξύ των τιμών (πχ. φύλο 1=άνδρας, 2=γυναίκα) ο μέσος όρος δεν έχει νόημα²⁸. Σε διακριτά δεδομένα όπου η διαβάθμιση μεταξύ των τιμών έχει νόημα²⁹, ο μέσος όρος χρησιμοποιείται, αλλά υπάρχουν αρκετές επιφυλάξεις για την ερμηνεία του και τη χρήση του συγκεκριμένου ελέγχου³⁰.

4.2. Στατιστική μεθοδολογία

Στα προηγούμενα παραδείγματα η σύγκριση ήταν κατά σειρά:

- μεταξύ του μέσου όρου και μίας προκαθορισμένης τιμής
- μεταξύ δύο διαφορετικών μέσων όρων και
- μεταξύ του ίδιου μέσου όρου σε δύο διαφορετικές χρονικές στιγμές.

²⁷ Παρατήρηση στο τελευταίο (τρίτο) παράδειγμα: Ο αναγνώστης μπορεί εύκολα και σωστά να διατυπώσει ερωτήματα και ενστάσεις σχετικά με την "ηθική" βάση της προτεινόμενης διαδικασίας. Με ποιο σκεπτικό οδηγούμε παιδιά ηλικίας 12 ετών σε ενδεχόμενη αύξηση της βίαιης συμπεριφοράς τους; Οι επιφυλάξεις είναι απολύτως κατανοητές. Το συγκεκριμένο (δυνητικό) ερευνητικό παράδειγμα όμως δεν είναι το μόνο για το οποίο θα μπορούσαν να διατυπωθούν αντιρρήσεις. Οι δεοντολογικές παράμετροι του ερευνητικού σχεδίου οφείλουν να απασχολούν συστηματικά τους ερευνητές. Ο αναγνώστης που ενδιαφέρεται μπορεί να μελετήσει τη σχετική βιβλιογραφία. Για ένα "ένοχο" ερευνητικό πείραμα βλ. ενδεικτικά την υπόθεση Stanley Milgram's Experiment (<http://www.cba.uri.edu/Faculty/dellabitta/mr415s98/EthicEtcLinks/Milgram.htm>)

²⁸ Οι συγκεκριμένες μεταβλητές έχουν οργανώσει τις τιμές τους σε δεδομένα ονομαστικής κλίμακας (nominal data).

²⁹ Δεδομένα τακτικής κλίμακας (ordinal data). Τέτοιου τύπου δεδομένα αναπτύσσονται σε μεταβλητές όπως το εισόδημα στη μορφή 1=Χαμηλής Εισοδηματικής Τάξης, 2=Μεσαίας Εισοδηματικής Τάξης, 3=Υψηλής Εισοδηματικής Τάξης.

³⁰ Βλ. σχετικά και την ενότητα 1.2 Κλίμακες Μέτρησης. Βλ. επίσης συμπληρωματικά <http://www.cmh.edu/stats/definitions.asp>

Η πρώτη περίπτωση αφορά στην εξέταση του μέσου όρου από έναν πληθυσμό: οι άλλες δύο θα εξετασθούν στην περίπτωση των δύο πληθυσμών (η τελευταία περίπτωση εντάσσεται σε αυτή την κατηγορία γιατί έχουμε δύο δείγματα έστω και αν τυπικά αυτά προέρχονται από τον ίδιο πληθυσμό). Σε όλες τις περιπτώσεις η συνηθέστερη στατιστική μεθοδολογία είναι ο έλεγχος t (t -test). Στη συνέχεια θα περιγράψουμε (χωρίς να αναφερθούμε ιδιαίτερα σε μαθηματικές εξισώσεις) την παραπάνω τεχνική, αλλά και διάφορους εναλλακτικούς τρόπους ελέγχου στη περίπτωση όπου το t -test δεν είναι κατάλληλο.

4.2.1 Έλεγχος για το μέσο όρο ενός πληθυσμού

4.2.1.1 Το σκεπτικό του ελέγχου

Προτού προχωρήσουμε στους τρόπους ελέγχου, θεωρούμε χρήσιμο να παραθέσουμε τον τρόπο που η Στατιστική θεωρία προσεγγίζει το πρόβλημα του ελέγχου για το μέσο όρο ενός πληθυσμού. Πολλές από τις παρακάτω ιδέες θα χρησιμοποιηθούν (με την αναγκαία προσαρμογή) και στην περίπτωση των δύο πληθυσμών.

Όταν θέλουμε να συγκρίνουμε το μέσο όρο μιας μεταβλητής του πληθυσμού με μία προκαθορισμένη τιμή και επειδή δεν έχουμε δεδομένα από όλο τον πληθυσμό (αν υπήρχαν, τότε ο έλεγχος θα ήταν περιττός, θα μπορούσαμε να υπολογίσουμε άμεσα τον πληθυσμιακό μέσο όρο) χρησιμοποιούμε τα αντίστοιχα δεδομένα του δείγματος. Πιο συγκεκριμένα, αυτό το οποίο ουσιαστικά κάνουμε είναι να δούμε πόσο "κοντά" είναι ο μέσος όρος του δείγματος για τη συγκεκριμένη μεταβλητή με την τιμή που έχουμε καθορίσει. Εφόσον είναι "αρκετά κοντά", τότε θεωρούμε ότι δεν υπάρχουν διαφορές και οδηγούμαστε σε σημείο να ισχυριστούμε κατ' επέκταση ότι ο μέσος όρος του πληθυσμού για τη συγκεκριμένη μεταβλητή είναι ίσος με την τιμή που είχαμε (προ)καθορίσει. Σε αντίθετη περίπτωση εντοπίζουμε τυχόν διαφορές και λέμε ότι ο μέσος όρος του πληθυσμού για τη συγκεκριμένη μεταβλητή είναι διαφορετικός από την προκαθορισμένη τιμή. Στην τελευταία περίπτωση είναι δυνατό να εξεταστεί εάν ο μέσος όρος είναι μικρότερος ή μεγαλύτερος της συγκεκριμένης τιμής.

4.2.1.2 Το κριτήριο ελέγχου

Ένα ενδιαφέρον ζήτημα είναι πώς καθορίζεται το "κοντά" για το οποίο έγινε λόγος παραπάνω. Ως μέτρο απόστασης λαμβάνεται αρχικά η διαφορά της προκαθορισμένης τιμής από το μέσο όρο του δείγματος. Όμως αυτή η διαφορά εξαρτάται από τις μονάδες του προβλήματος κάθε φορά. Για παράδειγμα, διαφορά 5 μονάδων σε εικοσάβαθμη βαθμολογική κλίμακα ή 50 μονάδων στο δείκτη ευφυΐας σημαίνει πολύ μεγάλη απόκλιση μεταξύ των δύο ποσοτήτων, ενώ η διαφορά των 500 ευρώ στο ετήσιο οικογενειακό

εισόδημα δεν αντιστοιχεί στην ίδια ποιοτική διαφορά. Επίσης, η ερμηνεία αλλάζει και μεταξύ ομοειδών διαφορών. Ας εξετάσουμε τη διαφορά των 500 ευρώ στο ετήσιο οικογενειακό εισόδημα. Αν ο υπό εξέταση πληθυσμός είναι μία κοινωνική ομάδα χαμηλών εισοδημάτων, η παραπάνω διαφοροποίηση είναι πολύ πιο έντονη από το αν εξετάζαμε έναν πληθυσμό με πιο ισόρροπη οικονομική κατανομή. Για το λόγο αυτό, ως τελικό μέτρο απόστασης του ελέγχου παίρνουμε τη διαφορά της προκαθορισμένης τιμής από το μέσο όρο του δείγματος διαιρεμένης με την τυπική απόκλιση του μέσου όρου του δείγματος που αποτελεί ένα μέτρο της "ποικιλίας"/διασποράς των τιμών του δείγματος γύρω από το μέσο όρο. Έχουμε δηλαδή περιληπτικά:

$$\text{Κριτήριο ελέγχου} = \frac{(\text{Μέσος όρος δείγματος} - \text{καθορισμένη τιμή})}{\text{Τυπική απόκλιση μέσου}}$$

Το κριτήριο αυτό έχει, υπό προϋποθέσεις, κάποιες γνωστές στατιστικές ιδιότητες. Πιο συγκεκριμένα ακολουθεί μία γνωστή στατιστική κατανομή, την κατανομή t (ή αλλιώς κατανομή Student), για αυτό και ονομάζεται έλεγχος t (t -test). Η κατανομή αυτή μας βοηθά στον υπολογισμό της τιμής p (p -value), στη σύγκριση με το επίπεδο στατιστικής σημαντικότητας (*statistical significance level*) και τέλος στην εξαγωγή των τελικών συμπερασμάτων όπως θα περιγράψουμε παρακάτω.

Ενδιαφέρον παρουσιάζει η προϋπόθεση που πρέπει να ισχύει για την εφαρμογή του παραπάνω κριτηρίου ελέγχου. Αυτή αναφέρεται στο γεγονός ότι τα δεδομένα πρέπει να προέρχονται από πληθυσμό που ακολουθεί κανονική κατανομή. Η κανονική κατανομή είναι η πιο γνωστή κατανομή της Στατιστικής και έχει τη μορφή συμμετρικής καμπύλης. Όπως αναφέρθηκε και σε προηγούμενα κεφάλαια, για να εξακριβώσουμε την ισχύ της παραπάνω προϋπόθεσης καλό είναι να σχεδιάσουμε ένα ιστόγραμμα, προκειμένου να διαπιστώσουμε εάν υπάρχουν σοβαρές αποκλίσεις από την κανονικότητα. Στις εφαρμογές, αλλά και στο Παράρτημα Α, θα εξεταστούν με τη βοήθεια στατιστικών ελέγχων οι προϋποθέσεις κανονικότητας. Σε περίπτωση κατά την οποία θα διαπιστωθεί ότι υπάρχουν σοβαρές αποκλίσεις από την κανονικότητα ο στατιστικός έλεγχος t δεν είναι (πλέον) κατάλληλος. Ο πιο αποδεκτός και καθιερωμένος εναλλακτικός έλεγχος στην περίπτωση αυτή είναι το *Wilcoxon signed rank* που βασίζεται στην κατάταξη των παρατηρήσεων.

4.2.1.3 Εξαγωγή συμπερασμάτων

Σε όλες τις περιπτώσεις το τελικό συμπέρασμα για τον πληθυσμό βασίζεται στα δειγματικά δεδομένα που αποτελούν ένα μικρό υποσύνολο του πληθυσμού. Υπάρχει λοιπόν η περίπτωση *λάθους συμπεράσματος*, δηλ. το τυχαίο δείγμα να αποτελείται από τέτοια δεδομένα που να δίνουν ένα λανθασμένο συμπέρασμα σχετικά με τον πληθυσμό. (Προσοχή: Αυτό δεν σημαίνει αναγκαστικά ότι το δείγμα δεν είναι αντιπροσωπευτικό. Υπάρχει πάντα πιθανότητα (έστω και μικρή) ένα τυχαίο δείγμα που συλλέχθηκε με τη σωστή τεχνική δειγματοληψίας να δίνει αποτελέσματα που δεν συμβαδίζουν με τη γενική εικόνα του πληθυσμού. Αυτή είναι εξάλλου και η έννοια του "τυχαίου" στη δειγματοληψία). Δυστυχώς τέτοιες περιπτώσεις δεν μπορούν να αποκλεισθούν. Υπάρχει όμως δυνατότητα συρρίκνωσης της πιθανότητας. Αναλυτικότερα:

Αρχικά υποθέτουμε ότι δεν υπάρχει (ουσιαστική) διαφορά μεταξύ του άγνωστου μέσου όρου του *πληθυσμού* και της προκαθορισμένης τιμής. Είναι δηλαδή αρκετά κοντά αυτές οι δύο τιμές. Αυτή ονομάζεται αρχική ή *μηδενική υπόθεση* (*null hypothesis*, συμβολίζεται με H_0) και εξετάζεται σε σχέση με την εναλλακτική υπόθεση H_E ως προς το εάν υπάρχει διαφορά.

Στη συνέχεια, με τη βοήθεια του κριτηρίου ελέγχου υπολογίζουμε την πιθανότητα η οποιαδήποτε διαφορά του μέσου όρου να οφείλεται σε τυχαίους παράγοντες ή όχι.

Αν η παραπάνω πιθανότητα είναι μικρότερη από ένα επιτρεπτό όριο, τότε απορρίπτουμε την αρχική υπόθεση και θεωρούμε ότι υπάρχει στατιστικώς σημαντική διαφορά (*statistically significant difference*) μεταξύ του μέσου όρου του πληθυσμού και της καθορισμένης τιμής. Στην αντίθετη περίπτωση θεωρούμε ότι δεν υπάρχει στατιστικώς σημαντική διαφορά.

Η πιθανότητα που περιγράψαμε παραπάνω υπολογίζεται από το στατιστικό έλεγχο t και ονομάζεται *τιμή p* (*p-value*). Το επιτρεπτό όριο με το οποίο συγκρίνεται η τιμή p ονομάζεται *επίπεδο στατιστικής σημαντικότητας* (*statistical significance level*) και καθορίζεται από τον ερευνητή. Το επίπεδο σημαντικότητας είναι ουσιαστικά η πιθανότητα λανθασμένης απόρριψης της αρχικής (μηδενικής) υπόθεσης. Η πιο συνηθισμένη τιμή είναι 5% (ή 0.05) ενώ άλλες τιμές που χρησιμοποιούνται είναι 1% (ή 0.01) και 10% (ή 0.10) για περισσότερο "αυστηρούς" και "χαλαρούς" ελέγχους αντιστοίχως. Είναι λοιπόν φανερό ότι σε κάθε έλεγχο είναι απαραίτητο για τον ερευνητή να έχει καθορίσει ένα επίπεδο σημαντικότητας. Τότε, σε περίπτωση όπου η τιμή p είναι μικρότερη ή ίση με το επίπεδο σημαντικότητας δηλώνουμε ότι υπάρχει στατιστικά σημαντική διαφορά μεταξύ του μέσου όρου του πληθυσμού και της προκαθορισμένης τιμής στο συγκεκριμένο επίπεδο σημαντικότητας.

Σχηματικά έχουμε τα ακόλουθα βήματα για την εξαγωγή συμπερασμάτων μέσω ενός ελέγχου t :

1. Έλεγχος της κανονικότητας των δεδομένων (απαραίτητη προϋπόθεση για κάθε έλεγχο t).
2. Καθορισμός της αρχικής (μηδενικής) υπόθεσης ότι δεν υπάρχει διαφορά μεταξύ πληθυσμιακού μέσου όρου και προκαθορισμένης τιμής. Αντιστοίχως καθορίζεται και η εναλλακτική υπόθεση.
3. Ορισμός του επιπέδου σημαντικότητας.
4. Πραγματοποίηση του ελέγχου t .
5. Εύρεση της τιμής p .
6. Σύγκριση της τιμής p με το επίπεδο σημαντικότητας:
 - 6(α) Αν η τιμή $p \leq$ επίπεδο σημαντικότητας τότε απορρίπτουμε την αρχική υπόθεση, δηλαδή υπάρχει στατιστικά σημαντική διαφορά μεταξύ του μέσου όρου του πληθυσμού και της προκαθορισμένης τιμής στο συγκεκριμένο επίπεδο σημαντικότητας, αλλιώς
 - 6(β) Αν τιμή $p >$ επίπεδο σημαντικότητας τότε δεν απορρίπτουμε την αρχική υπόθεση, δηλαδή πράγματι δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ του μέσου όρου του πληθυσμού και της προκαθορισμένης τιμής στο συγκεκριμένο επίπεδο σημαντικότητας.

4.3. Σχόλια για τον έλεγχο του μέσου όρου ενός πληθυσμού

Είναι φανερό ότι (σχεδόν) όλα τα παραπάνω μπορεί να υλοποιηθούν με τη χρήση ηλεκτρονικού υπολογιστή για μεγαλύτερη ταχύτητα και ευκολία. Το SPSS μας παρέχει την τιμή p για τον παραπάνω έλεγχο και ο ερευνητής, αφού έχει θέσει το επίπεδο σημαντικότητας που είναι κατάλληλο για την έρευνα, καταλήγει στο σχετικό συμπέρασμα.

Αξίζει επίσης να αναφερθεί ότι η παραπάνω ανάλυση εξέτασε την περίπτωση ύπαρξης διαφοράς μεταξύ του μέσου όρου ενός πληθυσμού και μίας προκαθορισμένης τιμής. Δεν έχουμε ασχοληθεί με την πιο πρακτική περίπτωση, στην οποία ο ερευνητής ενδιαφέρεται εάν ο μέσος όρος είναι μεγαλύτερος ή μικρότερος από την τιμή και όχι απλά διαφορετικός. Η αντιμετώπιση όμως του παραπάνω προβλήματος είναι ακριβώς η ίδια, με την ιδιαιτερότητα ότι το στατιστικό λογισμικό που θα χρησιμοποιήσουμε δεν μας δίνει απευθείας την τιμή p αλλά θα πρέπει να την υπολογίσουμε με βάση ορισμένες τιμές που αναφέρονται στα αποτελέσματα.

Ο έλεγχος όπου εξετάζεται απλά η διαφορά μεταξύ του μέσου όρου και της καθορισμένης τιμής λέγεται *δίπλευρος έλεγχος* (two-sided test), ενώ οι άλλοι λέγονται *μονόπλευροι έλεγχοι* (one-sided test).

Σε περιπτώσεις που το δείγμα είναι πολύ μεγάλο, τότε κατά πάσα πιθανότητα ο έλεγχος θα καταδείξει στατιστικά σημαντικές διαφορές έστω και αν οι αριθμητικές διαφορές είναι μικρές. Αυτό είναι ένα χαρακτηριστικό πολλών τέτοιων ελέγχων. Παρόμοιος προβληματισμός υπάρχει και στην ανάλυση συσχέτισης αλλά και στο κριτήριο χ^2 . Χωρίς να μπορούμε σε περισσότερες λεπτομέρειες, καλό είναι να αποφεύγονται τα πολύ μεγάλα ή τα πολύ μικρά δείγματα όταν έχουμε στατιστικούς ελέγχους.

4.4. Εφαρμογές του ελέγχου για το μέσο όρο ενός πληθυσμού

Οι παρακάτω εφαρμογές θα αναλύσουν τη μεθοδολογία που περιγράψαμε παραπάνω.

A.

Παράδειγμα 1β: Θεωρούμε ότι οι μαθητές ηλικίας 12 ετών ασχολούνται 8 ώρες κατά μέσο όρο εβδομαδιαίως με εξωσχολικές δραστηριότητες (ξένες γλώσσες, αθλητισμός, κ.λπ). Τα παρακάτω στοιχεία αφορούν τις ώρες που αφιερώνουν την εβδομάδα δέκα τυχαία επιλεγμένοι μαθητές αυτής της ηλικίας για εξωσχολικές δραστηριότητες. Μας ενδιαφέρει να εξετάσουμε εάν ο αρχικός συλλογισμός είναι σωστός.

Ώρες για εξωσχολικές δραστηριότητες την εβδομάδα 10 μαθητών

12 17 9,5 7,5 6,5 9,5 11 14 8,5 9,5

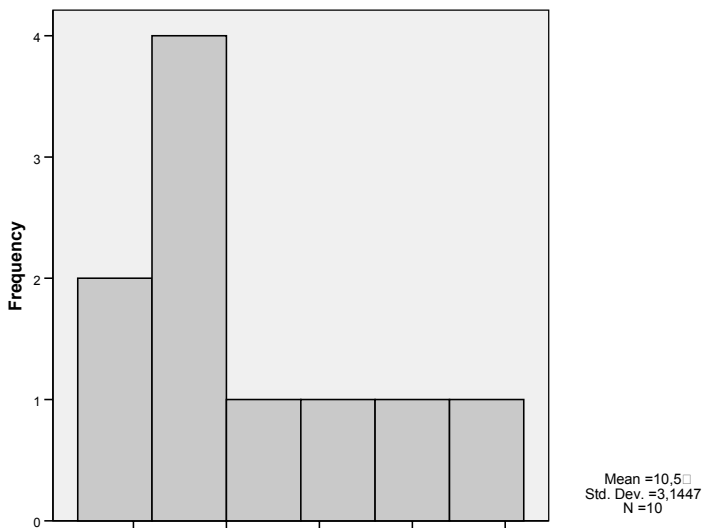
Θα ακολουθήσουμε τα βήματα που αναφέρθηκαν προηγουμένως για την εξαγωγή συμπερασμάτων.

Βήμα 1: Έλεγχος κανονικότητας

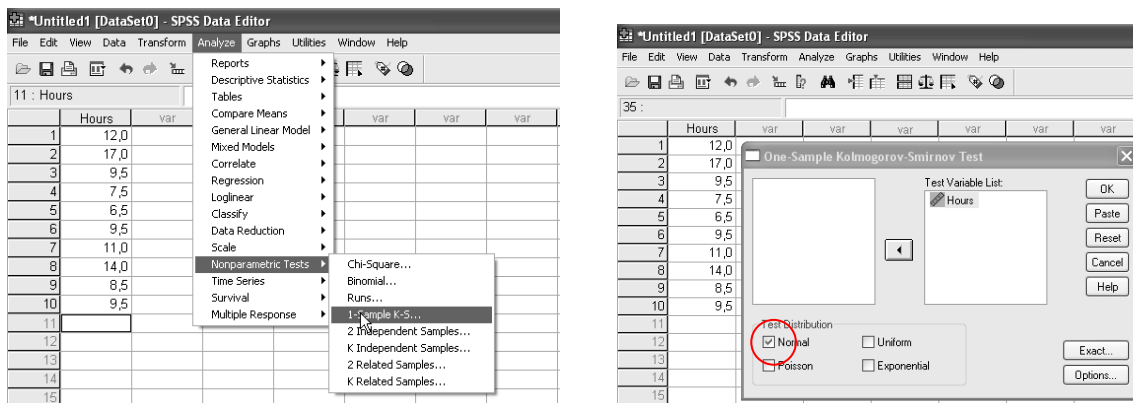
Αφού εισάγουμε τα δεδομένα στο SPSS εξετάζουμε αν προέρχονται από κανονική κατανομή.

Αρχικά είναι σκόπιμο να επιχειρήσουμε τη δημιουργία ενός ιστογράμματος (**Graphs → Histogram** τοποθετώντας το όνομα της μεταβλητής στη θέση **Variables**) (Εικόνα 27).

Ο έλεγχος κανονικότητας προκύπτει με τη βοήθεια των εντολών **Analyze → Nonparametric tests-1 Sample K-S**. (Εικόνα 28). Στο μενού που θα εμφανισθεί εισάγουμε το όνομα της μεταβλητής στο μέρος που λέει **Test Variable List** φροντίζοντας να υπάρχει κλικ στην επιλογή **Test Distribution Normal**. Τα αποτελέσματα εμφανίζονται αφού πατήσουμε OK (βλέπε Εικόνες 28 και 29).



Εικόνα 27: Ιστόγραμμα για τη μεταβλητή HOURS



Εικόνες 28 & 29: Οι ενέργειες για τον έλεγχο κανονικότητας

Από το ιστόγραμμα παρατηρούμε ότι δεν υπάρχουν σοβαρές αποκλίσεις από την κανονικότητα. Αυτό επιβεβαιώνεται από τα αποτελέσματα του ελέγχου κανονικότητας που παρουσιάζονται παρακάτω.

One-Sample Kolmogorov-Smirnov Test

		Hours
N		10
Normal Parameters ^{a,b}	Mean	10,500
	Std. Deviation	3,1447
Most Extreme Differences	Absolute	,225
	Positive	,225
	Negative	-,102
Kolmogorov-Smirnov Z		,711
Asymp. Sig. (2-tailed)		,693

a. Test distribution is Normal.

b. Calculated from data.

Εικόνα 30: Τα αποτελέσματα του ελέγχου κανονικότητας.

Επικεντρώνουμε την προσοχή μας στην ένδειξη: **Asymp. Sig. (2-tailed)**: πρόκειται για την τιμή p του ελέγχου για την κανονικότητα. Η αρχική ή μηδενική υπόθεση αναφέρει ότι τα δεδομένα προέρχονται από κανονική κατανομή. Αν η τιμή p είναι μικρότερη του επιπέδου σημαντικότητας (έστω 5%), τότε απορρίπτουμε τη μηδενική υπόθεση. Στην περίπτωσή μας έχουμε ότι $0,693 > 0,05$, άρα εξακολουθεί να ισχύει η υπόθεση της κανονικότητας και μπορούμε να προχωρήσουμε στον έλεγχο t (t-test).

Βήμα 2: Καθορισμός της αρχικής ή μηδενικής υπόθεσης (H_0) ότι δεν υπάρχει διαφορά μεταξύ πληθυσμιακού μέσου όρου και προκαθορισμένης τιμής. Αντιστοίχως καθορίζεται και η εναλλακτική υπόθεση (H_1).

H_0 : Οι μαθητές απασχολούνται κατά μέσο όρο 8 ώρες την εβδομάδα σε εξωσχολικές δραστηριότητες.

H_1 : Οι μαθητές δεν απασχολούνται κατά μέσο όρο 8 ώρες την εβδομάδα σε εξωσχολικές δραστηριότητες.

Βήμα 3: Ορισμός του επιπέδου σημαντικότητας

Ως επίπεδο σημαντικότητας ορίζεται το 5% (ή 0,05).

Βήμα 4: Πραγματοποίηση του ελέγχου t

Για το σκοπό αυτό επιλέγουμε κατά σειρά τις εξής εντολές:

Analyze → Compare Means → One Sample T Test

Στη συνέχεια εισάγουμε στη θέση **Test Variable(s)** τη μεταβλητή της οποίας θα ελέγξουμε το μέσο όρο (hours στο SPSS). Η προκαθορισμένη τιμή (στην προκειμένη περίπτωση 8) εισάγεται στη θέση **Test Value** και στη συνέχεια επιλέγουμε OK. Τα αποτελέσματα της ανάλυσης εμφανίζονται παρακάτω (Εικόνες 31 & 32):

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Hours	10	10,500	3,1447	,9944

Εικόνα 31

One-Sample Test

	Test Value = 8					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Hours	2,514	9	,033	2,5000	,250	4,750

Εικόνα 32

Βήμα 5: Εύρεση της τιμής p

Στα περιγραφικά στοιχεία της μεταβλητής Hours διαπιστώνουμε ότι ο δειγματικός μέσος όρος είναι 10.50, αρκετά διαφορετικός από το 8. Η στήλη Std. Error Mean δίνει το μέτρο της διασποράς των δεδομένων που χρησιμοποιείται στον παρονομαστή του κριτηρίου ελέγχου. Πράγματι μπορούμε να διαπιστώσουμε ότι: $t = (10,5 - 8) / 0,944 = 2,514$ που εμφανίζεται στο δεύτερο πίνακα. Η στήλη Sig. (2-tailed) δίνει την τιμή p για ένα δίπλευρο έλεγχο που είναι 0,033.

Βήμα 6: Σύγκριση της τιμής p με το επίπεδο σημαντικότητας

Διαπιστώνουμε ότι $0,033 < 0,05$, κατά συνέπεια υπάρχει στατιστικά σημαντική διαφορά σε σχέση με το 8. Το συμπέρασμα είναι ότι δεν ισχύει η αρχική υπόθεση, δηλαδή οι μαθητές ηλικίας 12 ετών δεν αφιερώνουν 8 ώρες εβδομαδιαίως για εξωσχολικές δραστηριότητες.

Αξίζει να προσέξουμε ότι εάν ήμασταν πιο αυστηροί στον καθορισμό του επιπέδου σημαντικότητας και επιλέγαμε το 1% τότε το συμπέρασμα θα ήταν το ακριβώς αντίθετο. Αυτό δεν είναι παράδοξο, αφού το στατιστικό συμπέρασμα για τον πληθυσμό με βάση το δείγμα περιέχει πάντα την πιθανότητα λάθους. Αν

Θέλουμε να είμαστε εξαιρετικά αυστηροί, θα πρέπει τα δειγματικά δεδομένα να είναι πολύ διαφορετικά από την προκαθορισμένη τιμή, κάτι το οποίο δεν συμβαίνει σε τόσο μεγάλο βαθμό εδώ.

B.

Ας υποθέσουμε τώρα ότι μας ενδιαφέρει να εξετάσουμε την εναλλακτική υπόθεση ότι οι μαθητές απασχολούνται υπερβολικά με εξωσχολικές δραστηριότητες, εξετάζουμε δηλαδή εάν ο μέσος όρος είναι μεγαλύτερος του 8. Σε αυτή την περίπτωση οι υποθέσεις παρουσιάζονται ως εξής:

H_0 : Οι μαθητές απασχολούνται κατά μέσο όρο 8 ώρες την εβδομάδα σε εξωσχολικές δραστηριότητες. (δεν αλλάζει)

H_1 : Οι μαθητές απασχολούνται κατά μέσο όρο *περισσότερο από* 8 ώρες την εβδομάδα σε εξωσχολικές δραστηριότητες.

Για να απαντήσουμε στο ερώτημα αυτό επαναλαμβάνουμε τη διαδικασία του SPSS που περιγράψαμε προηγουμένως και στη συνέχεια αφού επιλέξουμε και πάλι το 5% ως επίπεδο σημαντικότητας θα πρέπει να βρούμε την τιμή p . Αυτό γίνεται ως εξής:

1. Βρίσκουμε την τιμή p του δίπλευρου ελέγχου (0,033)

2. Βρίσκουμε την τιμή του κριτηρίου ελέγχου ($t=2,514$).

2α. Αν η τιμή του κριτηρίου είναι θετική και μας ενδιαφέρει ο μονόπλευρος έλεγχος για τιμές μεγαλύτερες της προκαθορισμένης τιμής, τότε η τιμή p του μονόπλευρου ελέγχου είναι το μισό της τιμής p του δίπλευρου ελέγχου.

2β. Αν η τιμή του κριτηρίου είναι θετική και μας ενδιαφέρει ο μονόπλευρος έλεγχος για τιμές μικρότερες της προκαθορισμένης τιμής, τότε η τιμή p του μονόπλευρου ελέγχου είναι:

1-το μισό της τιμής p του δίπλευρου ελέγχου.

2γ. Αν η τιμή του κριτηρίου είναι αρνητική και μας ενδιαφέρει ο μονόπλευρος έλεγχος για τιμές μικρότερες της προκαθορισμένης τιμής, τότε η τιμή p του μονόπλευρου ελέγχου είναι το μισό της τιμής p του δίπλευρου ελέγχου.

2δ. Αν η τιμή του κριτηρίου είναι αρνητική και μας ενδιαφέρει ο μονόπλευρος έλεγχος για τιμές μεγαλύτερες της προκαθορισμένης τιμής, τότε η τιμή p του μονόπλευρου ελέγχου είναι:

1-το μισό της τιμής p του δίπλευρου ελέγχου.

Στο παράδειγμά μας ισχύει η περίπτωση 2α και η τιμή p του μονόπλευρου ελέγχου είναι 0.016. Άρα το συμπέρασμα είναι ότι ο μέσος όρος εξωσχολικών δραστηριοτήτων είναι μεγαλύτερος του 8

σε επίπεδο στατιστικής σημαντικότητας 5%. Αξίζει να τονίσουμε ότι όπως και στην περίπτωση του δίπλευρου ελέγχου, το συμπέρασμα αντιστρέφεται σε επίπεδο σημαντικότητας 1%.

Τέλος αν στο παράδειγμα μάς ενδιέφερε μόνο εάν ο μέσος όρος είναι μικρότερος του 8, η τιμή p δίνεται από την περίπτωση 2β και είναι $1-0,016=0,984>0,05$. Το συμπέρασμα είναι ότι κατά μέσο όρο οι εξωσχολικές δραστηριότητες δεν είναι λιγότερες των 8 ωρών την εβδομάδα σε επίπεδο στατιστικής σημαντικότητας 5% (ή και οποιοδήποτε άλλο επίπεδο που χρησιμοποιείται στην πράξη).

4.5. Έλεγχος για τη σύγκριση μέσων όρων δύο πληθυσμών

Θα επεκτείνουμε την παραπάνω διαδικασία για την περίπτωση όπου έχουμε δείγματα από δύο πληθυσμούς. Όπως είδαμε και στην αρχική συζήτηση της ενότητας, στην κατηγορία των δύο πληθυσμών εξετάζουμε περιπτώσεις όπως του Παραδείγματος 2 (όπου πράγματι έχουμε δύο διαφορετικούς πληθυσμούς) αλλά και του Παραδείγματος 3 (όπου τα δύο δείγματα προέρχονται από τον ίδιο πληθυσμό). Στην πρώτη περίπτωση τα δείγματα είναι ανεξάρτητα, ενώ στη δεύτερη τα δείγματα είναι εξαρτημένα, αφού οι παρατηρήσεις τους προέρχονται από τα ίδια άτομα. Θα ακολουθήσουμε αυτή τη διαφοροποίηση και εδώ.

4.5.1. Ανεξάρτητα δείγματα

Ο έλεγχος t που περιγράφηκε προηγουμένως μπορεί άνετα να επεκταθεί και στους δύο πληθυσμούς. Το σκεπτικό ανάπτυξης του κριτηρίου ελέγχου παραμένει το ίδιο με την διαφορά ότι στον αριθμητή έχουμε τη διαφορά των δειγματικών μέσων όρων και στον παρονομαστή έχουμε μία συνολική τυπική απόκλιση μεταξύ των μέσων όρων των δύο δειγμάτων. Πιο συγκεκριμένα έχουμε:

$$\text{Κριτήριο ελέγχου} = (\text{Διαφορά μέσων όρων}) / \text{Συνολική τυπική απόκλιση}$$

Ισχύει η ίδια προϋπόθεση όπως και πριν για την εφαρμογή του παραπάνω κριτηρίου ελέγχου. Αυτή αναφέρεται στο γεγονός ότι τα δεδομένα από τις δύο ομάδες πρέπει να προέρχονται από πληθυσμούς με κανονικές κατανομές. Για να εξακριβώσουμε την ισχύ της παραπάνω (προ)ϋπόθεσης είναι σκόπιμο σε πρώτη φάση να σχεδιάσουμε δύο ιστογράμματα για την κατανομή των τιμών στα αντίστοιχα δείγματα και να δούμε αν υπάρχουν σοβαρές αποκλίσεις από την κανονικότητα ή να εφαρμόσουμε τον στατιστικό έλεγχο που προτάθηκε. Σε περίπτωση που διαπιστώσουμε ότι υπάρχουν σοβαρές αποκλίσεις από την

κανονικότητα, ο στατιστικός έλεγχος t δεν είναι πλέον κατάλληλος και για το λόγο αυτό χρησιμοποιούμε το κριτήριο Mann-Whitney, (το οποίο δεν προϋποθέτει κανονικότητα). Το κριτήριο Mann-Whitney ταξινομεί τα δεδομένα και χρησιμοποιεί τη συγκεκριμένη κατάταξη για την εξαγωγή του τελικού συμπεράσματος. Για να αποφύγουμε την περαιτέρω μαθηματική εξειδίκευση δεν θα παρουσιάσουμε λεπτομερώς το κριτήριο αυτό, αλλά θα δούμε πως εφαρμόζεται στο SPSS.

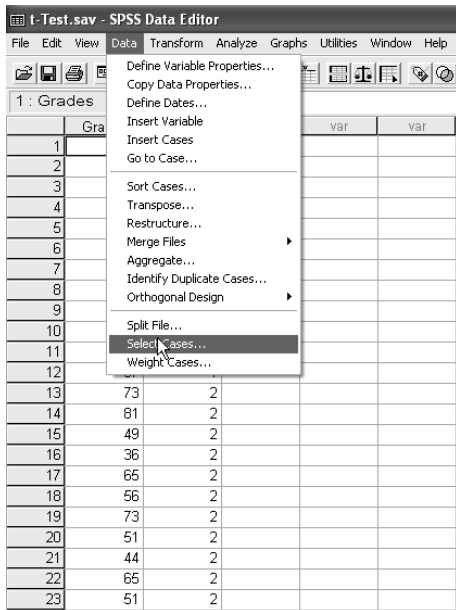
4.5.1.1 Εφαρμογές για ανεξάρτητα δείγματα

Τα παρακάτω στοιχεία αφορούν στη βαθμολογία (κλίμακα 0-100) των μαθητών μιας σχολικής τάξης γυμνασίου στο μάθημα των μαθηματικών και στην εξάρτηση της επίδοσης από τη μεταβλητή του φύλου (1-Αγόρι, 2-Κορίτσι). Εξετάζεται το ερώτημα εάν τα αγόρια έχουν υψηλότερο μέσο όρο βαθμολογίας στην εξέταση από τα κορίτσια.

Βαθμολογία	Φύλο	Βαθμολογία	Φύλο
83	1	73	2
75	1	81	2
67	1	49	2
44	1	36	2
28	1	65	2
56	1	56	2
91	1	73	2
39	1	51	2
71	1	44	2
54	1	65	2
38	1	51	2
57	2		

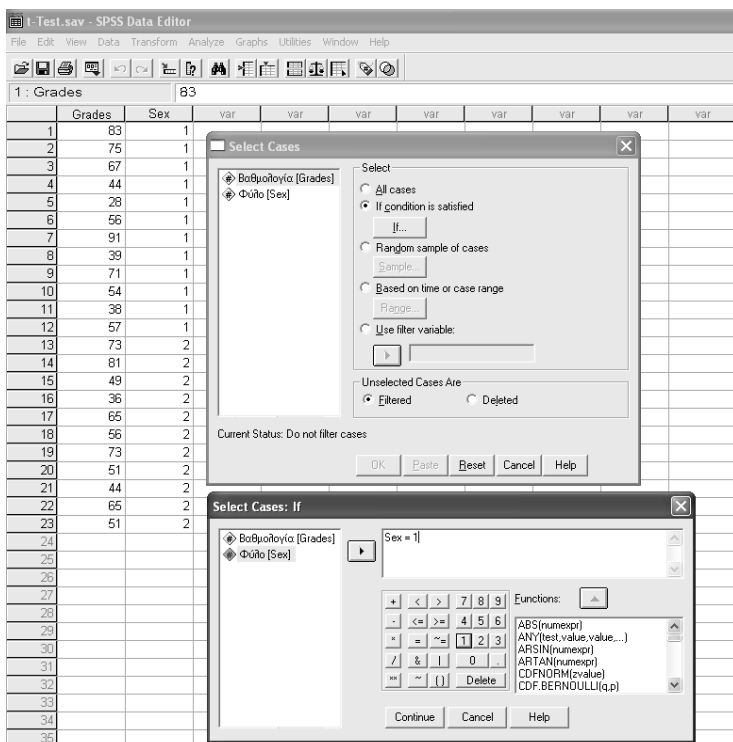
Αρχικά εισάγουμε τα δεδομένα σε δύο στήλες (Grades -Sex) όπως είναι και στον Πίνακα. Μπορούμε να διαπιστώσουμε ότι τα δεδομένα για τον κάθε πληθυσμό ακολουθούν την κανονική κατανομή. Αυτό γίνεται με την ίδια μεθοδολογία όπως και πριν, με τη διαφορά ότι ο έλεγχος θα πρέπει να γίνει χωριστά για αγόρια και κορίτσια.

Data → Select Cases



Εικόνα 33

Στη συνέχεια επιλέγουμε **If condition is satisfied** τοποθετώντας μέσα στο διαλογικό κουτί που εμφανίζεται **Sex=1**.



Εικόνα 34

Με τη διαδικασία αυτή επιλέγουμε αρχικά μόνο τα αγόρια, εφαρμόζοντας τους απαραίτητους ελέγχους κανονικότητας.

Στη συνέχεια επαναλαμβάνουμε τη διαδικασία μόνο για τα κορίτσια (**Sex=2**). Αφού έχουμε τελειώσει με τους ελέγχους κανονικότητας³¹, επαναλαμβάνουμε για μια τελευταία φορά τη διαδικασία επιλέγοντας **All cases**, ώστε να προχωρήσουμε στον έλεγχο των μέσων όρων και για τα δύο δείγματα.

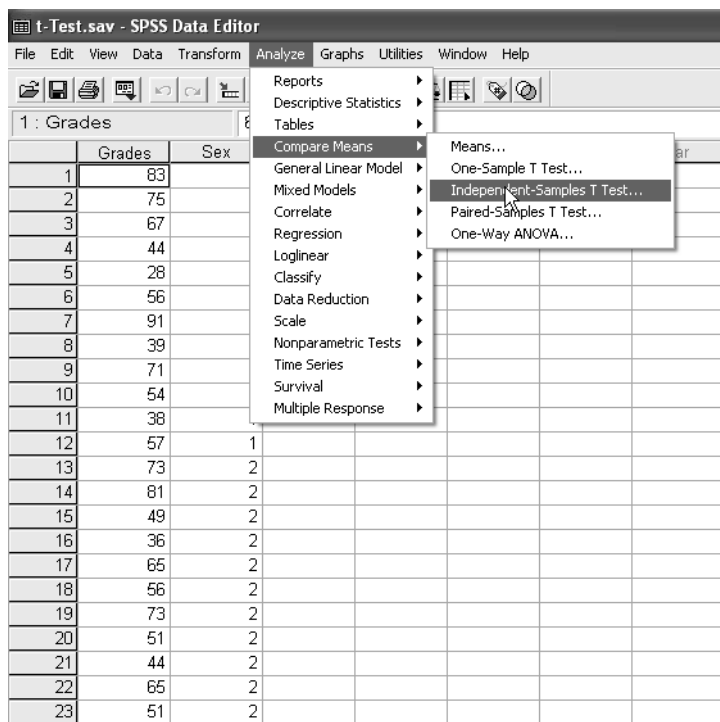
Οι υποθέσεις διαμορφώνονται ως εξής:

H_0 : ΔΕΝ υπάρχει διαφορά στο μέσο όρο βαθμολογίας ανάμεσα στα δύο φύλα

H_1 : Τα αγόρια έχουν υψηλότερο μέσο όρο βαθμολογίας από τα κορίτσια

Ο έλεγχος t γίνεται με την παρακάτω ακολουθία εντολών:

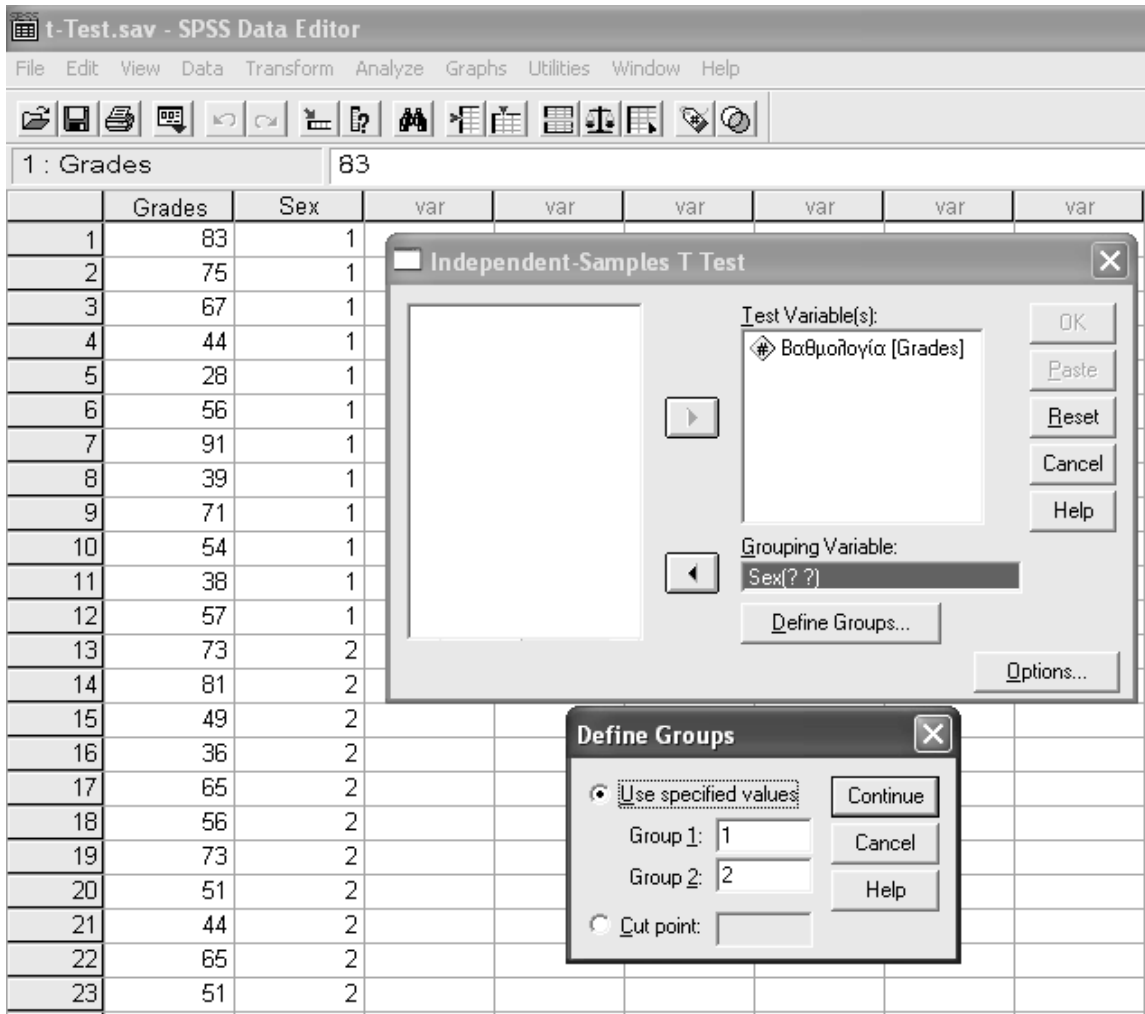
Analyze → Compare means → Independent Samples T Test



Εικόνα 35

Στη συνέχεια εισάγουμε το όνομα της μεταβλητής Grades στο χώρο **Test Variable(s)** και το όνομα της μεταβλητής Sex στο χώρο **Grouping Variable**. Κάνουμε κλικ στην επιλογή **Define Groups** βάζοντας τις τιμές που αντιστοιχούν στα αγόρια και στα κορίτσια (1 και 2 στα κουτιά **Group 1** και **Group 2** αντιστοίχως).

³¹ Στη συγκεκριμένη περίπτωση διαπιστώνουμε ότι οι κατανομές είναι κανονικές.



Εικόνα 36

Επιλέγουμε Continue και OK για να πάρουμε τα παρακάτω αποτελέσματα.

Group Statistics

	Φύλο	N	Mean	Std. Deviation	Std. Error Mean
Βαθμολογία	1	11	58.73	20.308	6.123
	2	12	58.42	13.276	3.833

Εικόνα 37

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Βαθμολογία	Equal variances assumed	3.068	.094	.044	21	.965	.311	7.093	-14.440	15.061
	Equal variances not assumed			.043	16.999	.966	.311	7.224	-14.930	15.551

Εικόνα 38

Τα αποτελέσματα **Group Statistics** δείχνουν μία πολύ μικρή διαφορά μεταξύ των δύο μέσων όρων (58.73 για τα αγόρια έναντι 58.42 για τα κορίτσια). Αυτό καταρχήν αποτελεί μία ένδειξη ότι δεν πρέπει να υπάρχει στατιστική διαφορά. Επιλέγουμε το 5% ως επίπεδο σημαντικότητας. Εξετάζοντας τα αποτελέσματα του ελέγχου t στη στήλη Sig (2-tailed) παρατηρούμε την ύπαρξη 2 τιμών p . Όπως και στην περίπτωση του ενός πληθυσμού αυτές αντιστοιχούν στην τιμή p του *δίπλευρου ελέγχου*. Ποια όμως από τις δύο θα επιλέξουμε (.965 ή .966);

Η απάντηση θα δοθεί από τον έλεγχο των διασπορών που γίνεται στις αρχικές στήλες. Πιο συγκεκριμένα ο έλεγχος του Levene που εμφανίζεται στην τρίτη στήλη θεωρεί αρχικά ότι οι διασπορές μεταξύ των δύο πληθυσμών είναι ίσες. Η τιμή p που δίνεται από τη στήλη Sig συγκρίνεται με το επίπεδο σημαντικότητας που έχουμε επιλέξει. Εάν η τιμή p (.094) είναι μικρότερη του 5%, τότε ο έλεγχος συμπεραίνει ότι οι διασπορές δεν είναι ίσες και θα επιλεγόταν η δεύτερη τιμή p του δίπλευρου ελέγχου (.966). Κάτι τέτοιο όμως δε συμβαίνει αφού $.094 > .05$, άρα οι διασπορές είναι στατιστικά ίσες και επιλέγουμε την τιμή .995 ως τιμή του *δίπλευρου ελέγχου*.

Όμως ο έλεγχος μεταξύ των μέσων όρων είναι μονόπλευρος, αφού μας ενδιαφέρει αν τα αγόρια έχουν μεγαλύτερο μέσο όρο από τα κορίτσια. Για να βρούμε την τιμή p του μονόπλευρου ελέγχου ακολουθούμε το παρακάτω σκεπτικό (παρόμοιο όπως και στην περίπτωση ενός πληθυσμού):

1. Βρίσκουμε την τιμή p του δίπλευρου ελέγχου (.996).
2. Βρίσκουμε την τιμή του κριτηρίου ελέγχου που αντιστοιχεί στην παραπάνω τιμή ($t = .005$).
- 2α. Αν η τιμή του κριτηρίου είναι θετική και μας ενδιαφέρει ο μονόπλευρος έλεγχος, όπου ο πρώτος πληθυσμός (Group 1=Αγόρια) έχει μέσο όρο *μεγαλύτερο* από το δεύτερο πληθυσμό (Group 2=Κορίτσια), τότε η τιμή p του μονόπλευρου ελέγχου, είναι το μισό ($\frac{1}{2}$) της τιμής p του δίπλευρου ελέγχου.
- 2β. Αν η τιμή του κριτηρίου είναι θετική και μας ενδιαφέρει ο μονόπλευρος έλεγχος, όπου ο πρώτος πληθυσμός (Group 1=Αγόρια) έχει μέσο όρο *μικρότερο* από το δεύτερο πληθυσμό (Group 2=Κορίτσια,) τότε η τιμή p του μονόπλευρου ελέγχου είναι:

1 - το μισό ($\frac{1}{2}$) της τιμής p του δίπλευρου ελέγχου.

- 2γ. Αν η τιμή του κριτηρίου είναι αρνητική και μας ενδιαφέρει ο μονόπλευρος έλεγχος, όπου ο πρώτος πληθυσμός (Group 1=Αγόρια) έχει μέσο όρο *μικρότερο* από το δεύτερο πληθυσμό (Group 2=Κορίτσια), τότε η τιμή p του μονόπλευρου ελέγχου είναι το μισό ($\frac{1}{2}$) της τιμής p του δίπλευρου ελέγχου.

2δ. Αν η τιμή του κριτηρίου είναι αρνητική και μας ενδιαφέρει ο μονόπλευρος έλεγχος, όπου ο πρώτος πληθυσμός (Group 1=Αγόρια) έχει μέσο όρο μεγαλύτερο από το δεύτερο πληθυσμό (Group 2=Κορίτσια), τότε η τιμή p του μονόπλευρου ελέγχου είναι:

1- το μισό ($\frac{1}{2}$) της τιμής p του δίπλευρου ελέγχου.

Στις παραπάνω περιπτώσεις με τον όρο "πρώτος πληθυσμός" εννοούμε αυτόν ο οποίος ορίστηκε πρώτα (στο συγκεκριμένο παράδειγμα: αγόρι) και όχι κάποιου άλλου είδους διαβάθμιση ή ταξινόμηση. Στο παράδειγμα ισχύει η περίπτωση 2α και η τιμή p του μονόπλευρου ελέγχου είναι .47. Άρα το συμπέρασμα είναι ότι ο μέσος όρος βαθμολογίας δεν είναι μεγαλύτερος στα αγόρια απ' ό,τι στα κορίτσια, σε επίπεδο στατιστικής σημαντικότητας 5%.

4.5.2. Εξαρτημένα δείγματα: Κριτήριο ελέγχου και εφαρμογή

Ο έλεγχος αυτός εξετάζει για κάθε στοιχείο του δείγματος τις διαφορές μεταξύ των μέσων όρων δύο διαφορετικών χρονικών στιγμών. Αν αυτή η διαφορά (καταλλήλως διαιρεμένη) είναι αρκετά κοντά στο 0 τότε λέμε ότι δεν υπάρχει στατιστική διαφορά μεταξύ των μέσων όρων για τις δύο διαφορετικές χρονικές στιγμές. Έχουμε λοιπόν το εξής κριτήριο ελέγχου:

$$\text{Κριτήριο ελέγχου} = \frac{\text{Μέσος όρος διαφορών}}{\text{Τυπική απόκλιση του μέσου}}$$

Ο έλεγχος αυτός ονομάζεται έλεγχος t για ζευγαρωτά δείγμα (*paired samples t-test*). Για να γίνει αυτός ο έλεγχος πρέπει να ισχύει η προϋπόθεση της κανονικότητας για τις τιμές των διαφορών μεταξύ των δύο διαφορετικών χρονικών στιγμών. Η διαπίστωση γίνεται με τις μεθόδους που αναφέρθηκαν και προηγουμένως. Σε περίπτωση που υπάρχει σοβαρή απόκλιση από την κανονικότητα τότε χρησιμοποιείται το κριτήριο *Wilcoxon signed rank*.

Εφαρμογή: Μας ενδιαφέρει να εξετάσουμε εάν μία συγκεκριμένη ψυχολογική θεραπεία συμβάλλει στην αύξηση της αυτο-εκτίμησης παιδιών που είχαν διαγνωσθεί ότι βρίσκονται στα αρχικά στάδια κατάθλιψης. Στα παιδιά αυτά δόθηκαν κατάλληλα ερωτηματολόγια πριν και μετά τη θεραπεία. Για κάθε φάση δημιουργήθηκε ο σχετικός δείκτης αυτο-εκτίμησης (κλίμακα 0-100). Τα στοιχεία παρουσιάζονται παρακάτω:

Αυτο-εκτίμηση πριν τη θεραπεία	Αυτο-εκτίμηση μετά τη θεραπεία	Διαφορά (Πριν-Μετά)
12	23	-11
34	45	-11
67	73	-6
43	54	-11
81	76	5
54	56	-2
56	66	-10
76	78	-2
65	79	-14
56	63	-7

Αφού περάσετε τα στοιχεία των 2 πρώτων στηλών (before και after) για να δημιουργήσετε στο SPSS τη στήλη της διαφοράς (Diff=before-after) ακολουθείτε τις εντολές: **Transform-Compute**. Στη συνέχεια στο κουτί **Target variable** βάζετε το όνομα της μεταβλητής της διαφοράς και στο κουτί **Numeric expression** δημιουργείτε τη διαφορά. Είναι εύκολο να δούμε ότι τα στοιχεία της διαφοράς ακολουθούν την κανονική κατανομή. Ο έλεγχος ανάγεται στην αρχική περίπτωση μέσου όρου ενός πληθυσμού για τη μεταβλητή diff σε σχέση με την προκαθορισμένη τιμή 0. Πιο συγκεκριμένα οι υποθέσεις είναι οι εξής:

H_0 : ΔΕΝ υπάρχει διαφορά κατά μέσο όρο στην αυτο-εκτίμηση των παιδιών πριν και μετά τη θεραπεία

H_1 : Μετά τη θεραπεία, η αυτο-εκτίμηση των παιδιών έχει αυξηθεί κατά μέσο όρο

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Before	54,40	10	20,533	6,493
	After	61,30	10	17,588	5,562

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Before & After	10	,966	,000

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 Befor - After	-6,900	5,782	1,828	-11,036	-2,764	-3,774	9	,004

Εικόνα 39

Εναλλακτικά, κάποιος μπορεί να χρησιμοποιήσει τις εντολές:

Analyze → Compare Means → Paired samples T Test

Στη συνέχεια επιλέγουμε τις δύο μεταβλητές, εισάγουμε τη σχηματιζόμενη διαφορά στο κουτί **Paired variables** και πατάμε OK. Τα αποτελέσματα με τη δεύτερη προσέγγιση φαίνονται παρακάτω (προφανώς το συμπέρασμα θα είναι το ίδιο με οποιοδήποτε τρόπο και να γίνει η ανάλυση).

Από το πρώτο μέρος των αποτελεσμάτων παρατηρούμε ότι ο μέσος όρος της αυτο-εκτίμησης των μαθητών έχει αυξηθεί αρκετά (54,4 πριν τη θεραπεία έναντι 61,3 μετά). Αυτό είναι μία πρώτη ένδειξη ότι πράγματι υπάρχει στατιστικά σημαντική αύξηση. Η εξαγωγή των συμπερασμάτων γίνεται με βάση τα αποτελέσματα στο μέρος Paired Sample Test. Η στήλη Sig (2-tailed) δίνει την τιμή p για τον δίπλευρο έλεγχο. Για να πάρουμε την τιμή p του μονόπλευρου ελέγχου ακολουθούμε τη διαδικασία που είχαμε αναφέρει και στην περίπτωση του ενός μέσου όρου για τη μεταβλητή diff. Σύμφωνα με την εφαρμογή μας ενδιαφέρει να εξετάσουμε το μονόπλευρο έλεγχο για τιμές μικρότερες της προκαθορισμένης τιμής, αφού η βελτίωση της αυτο-εκτίμησης σημαίνει $diff < 0$ (περίπτωση 2Γ). Άρα η τιμή p του μονόπλευρου ελέγχου είναι 0,002 και μπορούμε να πούμε ότι σε επίπεδο σημαντικότητας 5% η συγκεκριμένη θεραπεία βελτιώνει την αυτο-εκτίμηση.

4.6.Σχόλια για τον έλεγχο μεταξύ των μέσων όρων δύο πληθυσμών

Σε όλα τα παραπάνω παραδείγματα είδαμε ότι ισχύει η υπόθεση της κανονικότητας. Για το λόγο αυτό χρησιμοποιήσαμε κατάλληλους ελέγχους t. Όταν δεν ισχύει η υπόθεση της κανονικότητας, τότε χρησιμοποιούμε μία σειρά άλλων ελέγχων που δεν έχουν ως προϋπόθεση τα δεδομένα να ακολουθούν μία κανονική κατανομή. Αυτοί οι έλεγχοι ονομάζονται μη παραμετρικοί (non-parametric tests). Προφανώς η κατηγορία ελέγχων t, στους οποίους αναφερθήκαμε προηγουμένως αναφέρονται και ως παραμετρικοί έλεγχοι (parametric tests).

Οι εντολές για κάθε κατηγορία ελέγχου είναι οι ακόλουθες.

Για 2 ανεξάρτητα δείγματα έχουμε τον έλεγχο Mann Whitney:

Analyze → Non-Parametric Tests → 2 Independent Samples

→ **Grades** στο test variable list, sex στο group variable και Mann Whitney U στο test type

Για εξαρτημένα δείγματα έχουμε τον έλεγχο Wilcoxon signed rank:

Analyze → Non-Parametric Tests → 2 related Samples

before και after στο test pair list και Wilcoxon στο test type.

Το SPSS δεν έχει ενσωματωμένους μη παραμετρικούς ελέγχους για την περίπτωση του ενός πληθυσμού.

Μπορούμε όμως να δημιουργήσουμε τον έλεγχο αυτό βασιζόμενοι στον έλεγχο Wilcoxon signed rank για εξαρτημένα δείγματα στην περίπτωση δύο πληθυσμών. Πιο συγκεκριμένα, εκτός από τη μεταβλητή, το μέσο όρο της οποίας εξετάζουμε (Hours για το παράδειγμά μας προηγουμένως) δημιουργούμε μία νέα μεταβλητή (έστω New το όνομα της) η οποία έχει τιμή 0 παντού. Στη συνέχεια εφαρμόζουμε τον έλεγχο Wilcoxon signed rank για εξαρτημένα δείγματα βάζοντας hours και new στο test pair list και επιλέγοντας Wilcoxon στο test type. Η δημιουργία της new γίνεται με την ακόλουθη διαδικασία: **Transform-Compute** και μετά new στο Target Variable και 0 στο Numeric Expression.

Όπως αναφέραμε και στην ανάπτυξη των ελέγχων οι αρχικές υποθέσεις είναι ότι δεν υπάρχει διαφορά:

- (α) μεταξύ του μέσου όρου και της προκαθορισμένης τιμής
- (β) μεταξύ των μέσων όρων από δύο πληθυσμούς και
- (γ) μεταξύ των μέσων όρων του ίδιου πληθυσμού σε δύο διαφορετικές χρονικές στιγμές.

Η αρχική ή μηδενική υπόθεση για κάθε περίπτωση εξετάζεται με βάση το κριτήριο ελέγχου και το επίπεδο στατιστικής σημαντικότητας ώστε να διαπιστώσουμε αν τα δεδομένα την απορρίπτουν ή όχι. Όπως είδαμε, η μηδενική υπόθεση εξετάζεται για να διαπιστώσουμε αν από τα δεδομένα μπορούμε να πούμε ότι ισχύει τελικά αυτή η υπόθεση ή απορρίπτεται προς όφελος μιας άλλης υπόθεσης. Δηλαδή, η αρχική υπόθεση εξετάζεται σε σχέση με μία άλλη υπόθεση που λέει ότι ο μέσος όρος είναι μεγαλύτερος/μικρότερος/διαφορετικός από την προκαθορισμένη τιμή ή από τον άλλο μέσο όρο. Αυτή η δεύτερη υπόθεση όπως έχουμε πει ονομάζεται εναλλακτική υπόθεση ή ερευνητική υπόθεση ή υπόθεση εργασίας (alternative hypothesis-research hypothesis). Στην ουσία η υπόθεση εργασίας είναι αυτό που προσπαθούμε να αποδείξουμε και με βάση τα δεδομένα, το κριτήριο ελέγχου και το επίπεδο σημαντικότητας εξετάζουμε αν υπάρχουν επαρκή στοιχεία για να απορρίψουμε την αρχικώς θεωρούμενη ως σωστή μηδενική υπόθεση προς όφελος της εναλλακτικής υπόθεσης ή υπόθεσης εργασίας.

5

ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

5. ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

5.1. Εισαγωγή

Η ανάλυση διακύμανσης (Analysis of Variance - ANOVA) αποτελεί τη φυσική επέκταση της διαδικασίας ελέγχου του μέσου όρου ενός ή δύο πληθυσμών. Η διαφορά έγκειται στις επιπλέον δυνατότητες που προσφέρει, σε σχέση με τον έλεγχο t . Στην ανάλυση διακύμανσης δεν υφίσταται περιορισμός στη σύγκριση των μέσων όρων. Ενώ εφαρμόζοντας t -test ο ερευνητής ήταν υποχρεωμένος να εξετάσει έναν ή δύο μέσους όρους, στην ανάλυση διακύμανσης παρέχεται η δυνατότητα συνεξέτασης μέσων όρων που προέρχονται από τρεις, τέσσερις ή και περισσότερους πληθυσμούς. Η συγκεκριμένη τεχνική επιτρέπει μάλιστα στον ερευνητή να εξετάσει την επίδραση περισσότερων από μία μεταβλητών στους υπό εξέταση πληθυσμούς. Για την καλύτερη κατανόηση των διαφορών σε σχέση με τον έλεγχο t μελετήστε το παρακάτω παράδειγμα:

Παράδειγμα: Ένας ερευνητής επιθυμεί να εξετάσει τις διαφορές στους μέσους όρους εισοδήματος των αποφοίτων τριών (3) διαφορετικών τμημάτων. Είναι επίσης πιθανό, εκτός από τις διαφορές στους μέσους όρους, (για τους αποφοίτους των τριών τμημάτων), ο ερευνητής να ενδιαφέρεται για την εξέταση και ανάλυση των διαφορών στο εισόδημα με κριτήριο το φύλο (άνδρες, γυναίκες) των αποφοίτων.

Ο έλεγχος με το κριτήριο t προσφέρει πολύ λίγες δυνατότητες για τη διερεύνηση του παραπάνω παραδείγματος. Εφαρμόζοντας t -test ο έλεγχος μπορεί να γίνει μόνο μεταξύ δύο (2) πληθυσμών. Στην προκειμένη περίπτωση έχουμε τρία (3) διαφορετικά τμήματα. Ενδεχομένως μια αρχική σκέψη θα ήταν ο ερευνητής να κάνει διαδοχικούς ελέγχους t για όλους τους πιθανούς συνδυασμούς ζευγών μεταξύ των τριών (3) τμημάτων. Η συγκεκριμένη πρόταση, εκτός από τις πρακτικές της δυσκολίες (στην περίπτωση των οκτώ (8) τμημάτων θα έπρεπε να γίνουν συνολικά 28 έλεγχοι t) συνοδεύεται και από στατιστικά ολισθήματα: πολλαπλοί έλεγχοι t αυξάνουν σημαντικά την πιθανότητα στρέβλωσης του αποτελέσματος.

Υπάρχει όμως ακόμη ένας περιορισμός. Το κριτήριο t ελέγχει την επίδραση στο εισόδημα ενός μόνο παράγοντα κάθε φορά, (π.χ. ή του τμήματος ή του φύλου) και όχι ταυτόχρονα και των δύο. Στο παράδειγμά μας έχουμε δημιουργήσει 6 διαφορετικά επίπεδα εξέτασης του μέσου όρου (για τα 3 τμήματα εξετάζουμε χωριστά άνδρες και γυναίκες). Μία προσέγγιση θα ήταν να κάναμε διαδοχικούς ελέγχους t μεταξύ των 6 επιπέδων αλλά τότε επανερχόμαστε στον αρχικό προβληματισμό.

Η μεταβλητή της οποίας εξετάζουμε το μέσο όρο (συγκεκριμένη περίπτωση το εισόδημα) ονομάζεται *εξαρτημένη μεταβλητή (dependent variable)*, ενώ οι παράγοντες βάσει των οποίων δημιουργούμε τα επίπεδα εξέτασης (εδώ τμήμα φοίτησης και φύλο) ονομάζονται *ανεξάρτητες μεταβλητές (independent variables)*. Σε περίπτωση όπου έχουμε μόνο μία ανεξάρτητη μεταβλητή η μεθοδολογία ονομάζεται *ανάλυση διακύμανσης μίας (ή μονής) κατεύθυνσης (one-way analysis of variance, one-way ANOVA)*. Σε περίπτωση δύο ανεξάρτητων μεταβλητών, έχουμε ανάλυση διακύμανσης διπλής κατεύθυνσης. Είναι προφανές ότι η one-way ANOVA με 2 επίπεδα εξέτασης είναι ισοδύναμη διαδικασία με το γνωστό μας έλεγχο t.

Η διαδικασία ANOVA στηρίζεται (όπως και ο έλεγχος t) στην *κανονικότητα* των δεδομένων για κάθε επίπεδο πληθυσμού. Θα πρέπει λοιπόν πριν εφαρμοσθεί η συγκεκριμένη τεχνική να εξετάζεται αν υπάρχουν σοβαρές αποκλίσεις από την κανονικότητα. Η μεθοδολογία ελέγχου είναι όμοια με τον έλεγχο κανονικότητας στην περίπτωση του t-test. Σε περίπτωση σοβαρών αποκλίσεων από την κανονικότητα υπάρχουν διάφορες μη-παραμετρικές στατιστικές διαδικασίες (όπως ο έλεγχος Kruskal-Wallis). Σημειώνεται ότι τα δεδομένα της εξαρτημένης μεταβλητής πρέπει να είναι *συνεχή*.

Στη συνέχεια, η συζήτηση της σχετικής μεθοδολογίας επικεντρώνεται στην περίπτωση της one-way ANOVA, με ορισμένες αναφορές στην περίπτωση διπλής κατεύθυνσης. Σε όλες τις περιπτώσεις τα δείγματα είναι ανεξάρτητα.

5.2. Μεθοδολογία ANOVA

Στη συγκεκριμένη ενότητα θα παρουσιάσουμε το σκεπτικό ανάπτυξης του κριτηρίου ANOVA, τη μεθοδολογία που ακολουθείται (για να διαπιστώσουμε τις διαφοροποιήσεις) και τέλος ορισμένες προσεγγίσεις αναφορικά με την ανάλυση διακύμανσης διπλής κατεύθυνσης.

5.2.1 Το σκεπτικό για τη διαδικασία ANOVA

Η ANOVA αναλύει τη συνολική μεταβλητότητα όλων των τιμών. Δηλαδή αναλύει πόσο "διάσπαρτες" είναι όλες οι τιμές σε σχέση με ένα συνολικό μέσο όρο όλων των τιμών, ανεξάρτητα από το δείγμα στο οποίο ανήκουν. Αυτή η συνολική μεταβλητότητα παραμένει σταθερή και μπορεί να επιμερισθεί σε δύο (όχι κατ' ανάγκη ίσα) μέρη:

(α) τη μεταβλητότητα *μεταξύ των δειγμάτων*, δηλαδή πόσο διαφέρει ο μέσος όρος κάθε δείγματος από το συνολικό μέσο όρο και

(β) τη μεταβλητότητα *μέσα στα δείγματα*, δηλαδή πόσο διαφέρει (σε κάθε δείγμα) κάθε παρατήρηση από το μέσο όρο του συγκεκριμένου δείγματος.

Σε περίπτωση που υπάρχουν μεγάλες διαφορές μεταξύ των μέσων όρων, τότε η μεταβλητότητα μεταξύ των δειγμάτων είναι ιδιαίτερα μεγάλη και άρα η μεταβλητότητα μέσα στα δείγματα είναι μικρή (αφού η συνολική μεταβλητότητα παραμένει σταθερή). Έτσι, το κριτήριο ελέγχου στην ANOVA είναι ο λόγος των δύο πηγών μεταβλητότητας που αναφέρθηκαν παραπάνω, δηλαδή

$$\text{Κριτήριο ελέγχου} = (\text{μεταβλητότητα μεταξύ των δειγμάτων}) / \text{μεταβλητότητα μέσα στα δείγματα}$$

Όσο μεγαλύτερες τιμές λαμβάνει το κλάσμα, τόσο πιθανότερο είναι να υπάρχουν διαφορές μεταξύ των επιπέδων. Ο λόγος αυτός ακολουθεί μία γνωστή στατιστική κατανομή F και ισχύουν όλα όσα είχαμε αναφέρει στον έλεγχο t για την πιθανότητα εξαγωγής λάθους συμπεράσματος. Και σε αυτή την περίπτωση ανιχνεύεται τιμή p και επίπεδο στατιστικής σημαντικότητας.

5.2.2 Post hoc κριτήρια

Σε περίπτωση που διαπιστωθεί η ύπαρξη στατιστικά σημαντικών διαφορών ανάμεσα στους μέσους όρους, το επόμενο βήμα είναι να εντοπίσουμε τα επίπεδα μεταξύ των οποίων υπάρχουν διαφορές. Για παράδειγμα στο εισαγωγικό παράδειγμα της ενότητας, μας ενδιαφέρει να εξετάσουμε μεταξύ ποιων τμημάτων υπάρχει διαφορά ως προς το μέσο εισόδημα και συγκεκριμένα σε ποια ζεύγη των τριών τμημάτων εντοπίζονται οι διαφορές.

Επειδή η ανάλυση αυτή γίνεται μετά τη διαπίστωση στατιστικά σημαντικών διαφορών μεταξύ των μέσων όρων ονομάζεται post hoc ανάλυση (εκ των υστέρων ανάλυση ή και a posteriori ανάλυση) και τα κριτήρια post hoc κριτήρια. Στη σχετική βιβλιογραφία προτείνεται ποικιλία post hoc κριτηρίων. Θα αναφερθούμε περιληπτικά στα τέσσερα από αυτά, τα οποία ως πιο σημαντικά χρησιμοποιούνται συνηθέστερα:

-Το κριτήριο LSD (Least Squares Differences). Ουσιαστικά πρόκειται για πολλαπλούς ελέγχους t , χωρίς όμως να αντιμετωπίζεται επαρκώς το πρόβλημα της αύξησης στην πιθανότητα εξαγωγής λάθους συμπεράσματος. Είναι λιγότερο επικίνδυνο για μικρό αριθμό ελέγχων, σε κάθε περίπτωση πάντως καλό είναι να αποφεύγεται.

-Το κριτήριο Bonferroni. Πρόκειται για ένα "αυστηρό" κριτήριο όσον αφορά το επίπεδο σημαντικότητας που αντιμετωπίζει ικανοποιητικά το πρόβλημα της πιθανότητας εξαγωγής λάθους συμπεράσματος.

-Το κριτήριο Tukey HSD (Honestly Significant Difference). Θεωρείται το ασφαλέστερο κριτήριο. Είναι ένα "συντηρητικό" κριτήριο, με την έννοια ότι κάποιες στατιστικά σημαντικές

διαφορές με άλλα κριτήρια είναι πιθανό να μην εμφανίζονται εδώ.

-Το κριτήριο Scheffe. Θεωρείται επίσης "συντηρητικό" κριτήριο. Εφαρμόζεται συχνά και καλύπτει πλήθος περιπτώσεων. Μαζί με το Tukey είναι από τα πιο χρήσιμα στην πράξη.

5.3. Σχόλια για την ANOVA

Η μεθοδολογία ANOVA μπορεί να επεκταθεί και σε παραπάνω από μία ανεξάρτητες μεταβλητές. Το σκεπτικό δεν αλλάζει, απλά οι υπολογισμοί γίνονται πιο δύσκολοι, κάτι το οποίο δεν αποτελεί πρόβλημα στο SPSS (ή και σε οποιοδήποτε άλλη εφαρμογή στατιστικής επεξεργασίας & ανάλυσης δεδομένων). Ένα σημαντικό ζήτημα, το οποίο δεν αντιμετωπίστηκε στις μέχρι τώρα αναλύσεις, είναι η ενδεχόμενη αλληλεπίδραση μεταξύ των ανεξάρτητων μεταβλητών. Αυτή η αλληλεπίδραση υπολογίζεται ως μία επιπλέον ανεξάρτητη μεταβλητή.

Τα κριτήρια της post hoc ανάλυσης δίνουν συνήθως (αλλά όχι πάντα) ταυτόσημα συμπεράσματα. Καλό είναι να επιλέγουμε κριτήρια με βάση τις στατιστικές τους ιδιότητες και όχι με βάση το ποιο δίνει στατιστικά σημαντικές διαφορές.

Όπως και στον έλεγχο t , ισχύει το φαινόμενο της "συχνής" και όχι αναγκαστικά πραγματικής, στατιστικά σημαντικής διαφοράς όταν τα δείγματα είναι μεγάλα σε μέγεθος.

5.4. Εφαρμογή της ANOVA

Μας ενδιαφέρει να συγκρίνουμε την απόδοση τριών σχολικών τάξεων στη μαθηματική ικανότητα με βάση τις εξετάσεις που διεξάγει κάθε τρία χρόνια σε παιδιά ηλικίας 15 ετών ο ΟΟΣΑ μέσω του προγράμματος PISA (Program for the International Student Assessment³²).

³² Βλ. σχετικά <http://www.pisa.oecd.org>

Τα αποτελέσματα παρουσιάζονται παρακάτω³³:

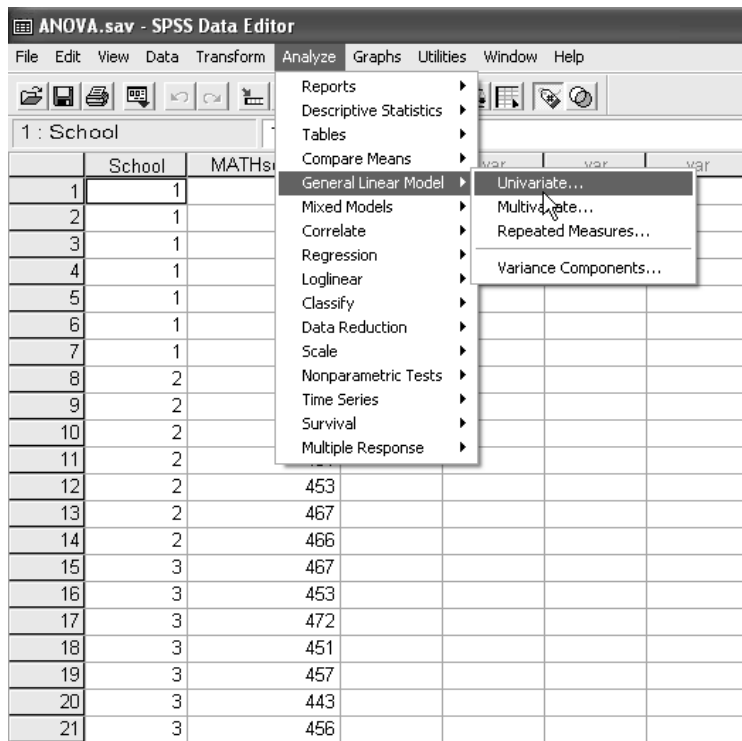
Σχολείο 1: 469, 474, 478, 465, 459, 489, 478

Σχολείο 2: 462, 465, 447, 431, 453, 467, 466

Σχολείο 3: 467, 453, 472, 451, 457, 443, 456

Εισάγουμε τα δεδομένα στο SPSS. Εφαρμόζοντας τους σχετικούς ελέγχους κανονικότητας³⁴, διαπιστώνουμε ότι δεν παραβιάζεται η προϋπόθεση της κανονικότητας. Επιλέγεται ο στατιστικός έλεγχος one-way ANOVA και θέτουμε το 5% ως επίπεδο στατιστικής σημαντικότητας.

Από το μενού Analyze επιλέγουμε General Linear Model και Univariate.



Εικόνα 40

Ως **Dependent Variable** αναγνωρίζουμε τη μεταβλητή MATHSscores και ως **Fixed Factor(s)** τη μεταβλητή school. (Σημείωση: εάν είχαμε two-way ANOVA, θα τοποθετούσαμε και τις δύο ανεξάρτητες μεταβλητές στη συγκεκριμένη θέση). Επιλέγουμε **Post hoc** και αφού εντάξουμε τη μεταβλητή school στο μενού **Post hoc tests for** επιλέγουμε τα τεστ Scheffe, Tukey. Επιλέγουμε Continue και OK και έχουμε τα παρακάτω αποτελέσματα.

³³ Πρόκειται για υποθετικές (και όχι πραγματικές) επιδόσεις

³⁴ Βλ. σχετικά Παράρτημα 1

Univariate Analysis of Variance

Between-Subjects Factors

		N
Σχολική Μονάδα	1	7
Φοίτησης	2	7
	3	7

Εικόνα 41

Tests of Between-Subjects Effects

Dependent Variable: Επίδοση στα Μαθηματικά

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1308,286 ^a	2	654,143	5,329	,015
Intercept	4482324,000	1	4482324,000	36512,337	,000
School	1308,286	2	654,143	5,329	,015
Error	2209,714	18	122,762		
Total	4485842,000	21			
Corrected Total	3518,000	20			

a. R Squared = ,372 (Adjusted R Squared = ,302)

Εικόνα 42

Multiple Comparisons

Dependent Variable: Επίδοση στα Μαθηματικά

	(I) Σχολική Μονάδα Φοίτησης	(J) Σχολική Μονάδα Φοίτησης	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	1	2	17,29*	5,922	,024	2,17	32,40
		3	16,14*	5,922	,035	1,03	31,26
	2	1	-17,29*	5,922	,024	-32,40	-2,17
		3	-1,14	5,922	,980	-16,26	13,97
	3	1	-16,14*	5,922	,035	-31,26	-1,03
		2	1,14	5,922	,980	-13,97	16,26
Scheffe	1	2	17,29*	5,922	,031	1,49	33,08
		3	16,14*	5,922	,045	,35	31,93
	2	1	-17,29*	5,922	,031	-33,08	-1,49
		3	-1,14	5,922	,982	-16,93	14,65
	3	1	-16,14*	5,922	,045	-31,93	-,35
		2	1,14	5,922	,982	-14,65	16,93

Based on observed means.

*. The mean difference is significant at the ,05 level.

Εικόνα 43

Το κριτήριο ελέγχου F και η αντίστοιχη τιμή p δίνονται από το δεύτερο μέρος των αποτελεσμάτων και ειδικότερα στις στήλες F και Sig. Οι τιμές αυτές είναι 5,329 και 0,015 αντιστοίχως. Άρα σε επίπεδο στατιστικής σημαντικότητας 5% (όχι όμως και αν είχαμε επιλέξει το 1%) υπάρχει στατιστικά σημαντική διαφορά μεταξύ των τριών σχολικών μονάδων. Αυτό είναι μάλλον αναμενόμενο αν κοιτάξουμε και τους 3 μέσους όρους

(σχολείο 1=473, σχολείο 2=455,9 και σχολείο 3=457) ενώ οι τυπικές αποκλίσεις τους δεν διαφέρουν σημαντικά (γύρω στο 10).

Η post hoc ανάλυση συγκρίνει, με τα κριτήρια Tukey και Scheffe, τις διαφορές ανά σχολείο και εμφανίζεται στην ενότητα με τον τίτλο Multiple Comparisons. Για κάθε μέθοδο, κάθε σχολείο (I) συγκρίνεται με τα άλλα 2 σχολεία (J). Έτσι με τη μέθοδο Tukey ο δίπλευρος έλεγχος του σχολείου 1 με το σχολείο 2 έδωσε μία τιμή $p = 0,024$. Αν μας ενδιαφέρει ο μονόπλευρος έλεγχος για το αν το σχολείο 1 έχει μέσο όρο μεγαλύτερο από το σχολείο 2 (όπως φαίνεται και από τις αριθμητικές τους τιμές), τότε σύμφωνα με την τεχνική που αναπτύχθηκε στην περίπτωση ελέγχου t για το μέσο όρο δύο πληθυσμών για ανεξάρτητα δείγματα η τιμή p είναι 0,012. Αυτό σημαίνει ότι το σχολείο 1 υπερτερεί κατά μέσο όρο του σχολείου 2 στη μέση μαθηματική επίδοση στο PISA σε επίπεδο στατιστικής σημαντικότητας 5% (όχι όμως και για 1%). Παρόμοια συμπεράσματα βγαίνουν από τη σύγκριση μεταξύ των σχολείων 1 και 3, ενώ δεν υπάρχουν διαφορές μεταξύ των σχολείων 2 και 3. Τέλος τα συμπεράσματα είναι ταυτόσημα και με τις 2 μεθόδους.

Οι εντολές για το μη-παραμετρικό έλεγχο Kruskal Wallis είναι οι ακόλουθες: **Analyze → Non-Parametric Tests-K Independent Samples** και μετά εισάγουμε το pisa στο test variable list και school στο group variable.

Πίνακας 3: Σύνοψη των κυριότερων ελέγχων μέσων όρων

Περίπτωση	Παραμετρικοί έλεγχοι	Μη-παραμετρικοί
Ένας μέσος όρος	t-test για ένα πληθυσμό	Wicoxon signed rank
Δύο μέσοι όροι (ανεξάρτητα δείγματα)	t-test για δύο πληθυσμούς	Mann-Whitney
Δύο μέσοι όροι (εξαρτημένα δείγματα)	Paired t-test για δύο πληθυσμούς	Wicoxon signed rank
Περισσότεροι από δύο μέσοι όροι (ανεξάρτητα δείγματα)	ANOVA	Kruskal-Wallis (για one-way)

6

ΤΟ ΚΡΙΤΗΡΙΟ χ^2

6. ΤΟ ΚΡΙΤΗΡΙΟ χ^2

Το κριτήριο χ^2 είναι γνωστό ως κριτήριο ελέγχου ανεξαρτησίας (chi square test of independence) ή κριτήριο ελέγχου πινάκων συνάφειας (contingency tables). Οι πίνακες συνάφειας ή πίνακες διπλής εισόδου αναπτύσσονται σε (έστω k) γραμμές και (έστω l) στήλες. Το κριτήριο εξετάζει αν οι δύο μεταβλητές που απαρτίζουν τον πίνακα διπλής εισόδου είναι ανεξάρτητες ή όχι. Πρόκειται ουσιαστικά για στατιστικό έλεγχο ο οποίος βασίζεται στη χρήση της στατιστικής κατανομής χ^2 σε επίπεδο σημαντικότητας α . Το σκεπτικό του ελέγχου είναι το εξής: Αρχικά υποθέτουμε ότι δεν υπάρχει σχέση μεταξύ των μεταβλητών (δηλαδή είναι ανεξάρτητες οι μεταβλητές μεταξύ τους). Μπορούμε να υπολογίσουμε τότε τις αναμενόμενες συχνότητες (δηλαδή τις συχνότητες που αναμένουμε αν ισχύει η αρχική υπόθεση ότι δεν υπάρχει σχέση). Τις συγκρίνουμε με τις πραγματικές συχνότητες. Αν διαφέρουν πολύ τότε απορρίπτουμε την υπόθεση της ανεξαρτησίας και λέμε ότι υπάρχει σχέση μεταξύ των 2 μεταβλητών.

Οι προϋποθέσεις εφαρμογής του κριτηρίου χ^2 είναι οι ακόλουθες:

- Το εξεταζόμενο δείγμα θα πρέπει να έχει επιλεγεί τυχαία και να έχει πλήθος $n > 25$ ³⁵
- Οι παρατηρήσεις πρέπει να είναι ανεξάρτητες
- Στις περιπτώσεις των κατηγοριοποιημένων μεταβλητών οι επιμέρους κατηγορίες θα πρέπει να μην επικαλύπτονται
- Δεν πρέπει πολλές (συνήθως πάνω από το 25%) από τις αναμενόμενες συχνότητες να είναι μικρότερες του πέντε
- Το κριτήριο δεν είναι αξιόπιστο όταν το συνολικό δείγμα είναι μεγάλο ($n > 250$), γιατί τότε υπάρχει το ενδεχόμενο η στατιστικά σημαντική σχέση μεταξύ των δύο μεταβλητών να είναι πλαστή καθώς προκύπτει από τη διόγκωση³⁶ της τιμής του χ^2 .

6.1. Εφαρμογή του χ^2

Ο ερευνητής Α επιχειρεί να ελέγξει κατά πόσο το φύλο επηρεάζει τη δήλωση των υποκειμένων αναφορικά με το βαθμό πολιτικής τους ενημέρωσης (ή διαφορετικά εάν ο βαθμός πολιτικής ενημέρωσης επηρεάζεται από το φύλο)³⁷.

³⁵ Στην περίπτωση που το $n < 25$ προτείνεται η επιλογή της μεθόδου Monte Carlo.

³⁶ Στην περίπτωση που το $n > 250$ προτείνεται η επιλογή τυχαίου υποσυνόλου (εφαρμογή στο SPSS της διαδικασίας random sample) ή η επιλογή της μεθόδου Monte Carlo.

³⁷ Ενδεικτική σχετική ερώτηση: Πόσο νομίζεις ότι είσαι ενημερωμένος/ος για τα πολιτικά ζητήματα; [Πάρα πολύ=1, Πολύ=2, Αρκετά=3, Λίγο=4, Πολύ λίγο=5, Καθόλου=6]

Προκειμένου να ελεγχθεί η υπόθεση εργασίας, διατυπώνεται η μηδενική της υπόθεση και εξετάζεται η ανεξαρτησία των δύο μεταβλητών με εφαρμογή του ελέγχου ανεξαρτησίας (chi square test of independence). Οι υποθέσεις διαμορφώνονται ως εξής:

H_0 : Ο βαθμός πολιτικής ενημέρωσης ΔΕΝ επηρεάζεται από το φύλο

H_1 : Ο βαθμός πολιτικής ενημέρωσης επηρεάζεται από το φύλο

Επιλέγεται το στατιστικό κριτήριο χ^2 , προκειμένου ο ερευνητής να συγκρίνει (υπό όρους) τις παρατηρούμενες συχνότητες με τις συχνότητες που θα εμφανίζονταν εάν η μηδενική υπόθεση ήταν αληθής (αναμενόμενες συχνότητες). Αρχικά υπολογίζονται οι συχνότητες για τις τιμές κάθε μεταβλητής. Ο υπολογισμός των συχνοτήτων με τη βοήθεια του SPSS έχει ως εξής:

Πίνακας 4: Υπολογισμός συχνοτήτων για τις τιμές της μεταβλητής του

Φύλο

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Κορίτσι	83	41.5	41.5	41.5
Αγόρι	117	58.5	58.5	100.0
Total	200	100.0	100.0	

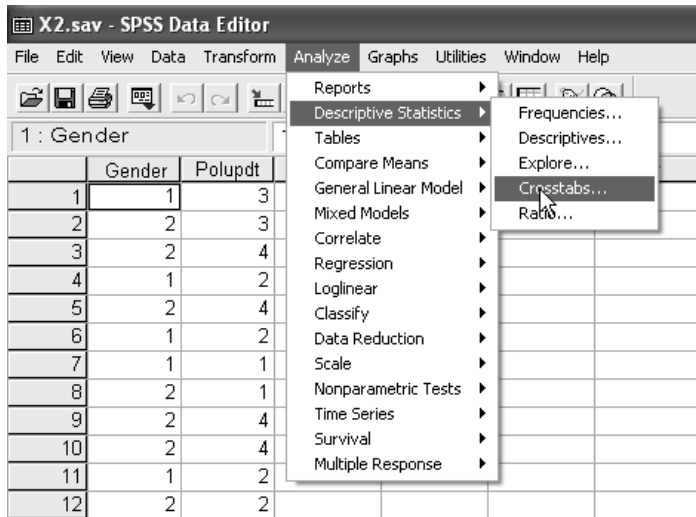
Πίνακας 5: Υπολογισμός συχνοτήτων για τις τιμές της μεταβλητής ΒΑΘΜΟΣ ΠΟΛΙΤΙΚΗΣ ΕΝΗΜΕΡΩΣΗΣ

Βαθμός Πολιτικής Ενημέρωσης

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Πάρα πολύ	37	18.5	18.8	18.8
Πολύ	40	20.0	20.3	39.1
Αρκετά	31	15.5	15.7	54.8
Λίγο	69	34.5	35.0	89.8
Πολύ λίγο	11	5.5	5.6	95.4
Καθόλου	9	4.5	4.6	100.0
Total	197	98.5	100.0	
Missing Missing	3	1.5		
Total	200	100.0		

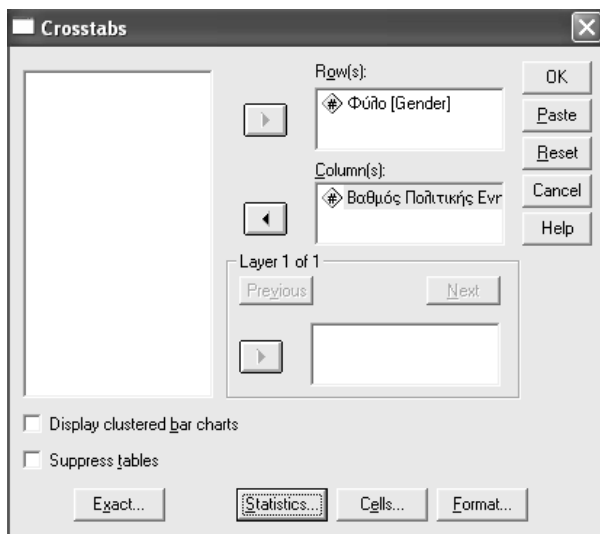
Αυτό βέβαια που ενδιαφέρει είναι η συνεξέταση των μεταβλητών, γεγονός που επιτυγχάνεται με τη διασταύρωσή τους. Στο SPSS η διασταύρωση των μεταβλητών προκύπτει με τη βοήθεια της επιλογής Crosstabs ως εξής:

- από το Menu Analyse επιλέγεται η εντολή Descriptive Statistics και στη συνέχεια Crosstabs.



Εικόνα 44: Εντολές του SPSS για το χ^2

Στο πλαίσιο διαλόγου που εμφανίζεται οι μεταβλητές που ενδιαφέρουν τοποθετούνται σε γραμμές (rows) και στήλες (columns).



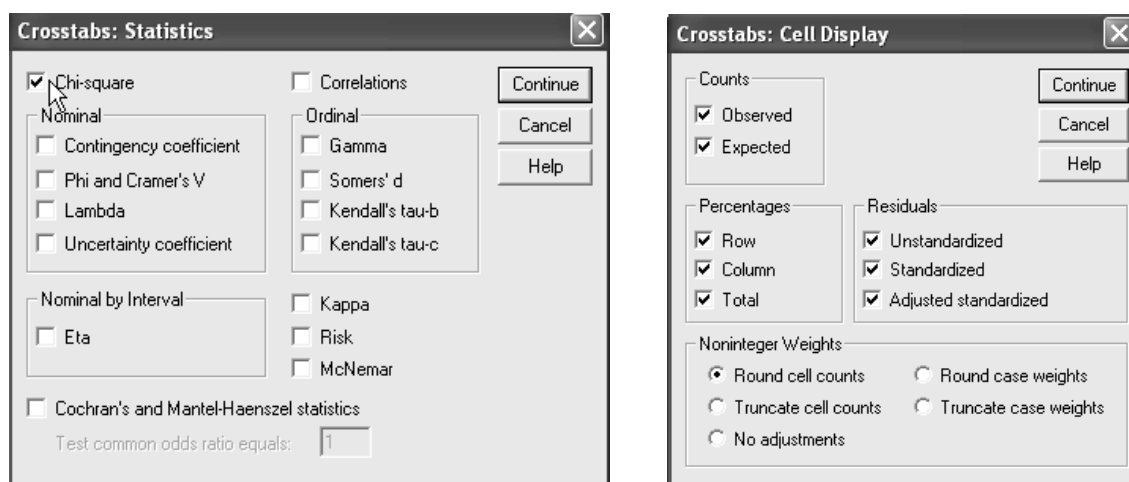
Εικόνα 45: Εντολές του SPSS για το κριτήριο χ^2

Φύλο * Βαθμός Πολιτικής Ενημέρωσης Crosstabulation

Count		Βαθμός Πολιτικής Ενημέρωσης						Total
		Πάρα πολύ	Πολύ	Αρκετά	Λίγο	Πολύ λίγο	Καθόλου	
Φύλο	Κορίτσι	24	19	17	17	2	2	81
	Αγόρι	13	21	14	52	9	7	116
Total		37	40	31	69	11	9	197

Εικόνα 46: Διασταύρωση της μεταβλητής Φύλο με τη μεταβλητή Βαθμός Πολιτικής Ενημέρωσης

Στον παραπάνω πίνακα συνάφειας οι τιμές στα κελιά του πίνακα είναι οι παρατηρούμενες συχνότητες. Για το στατιστικό έλεγχο είναι απαραίτητος ο προσδιορισμός των αναμενόμενων συχνοτήτων. Οι αναμενόμενες συχνότητες αναφέρονται στις συχνότητες που ο ερευνητής θα διαπίστωνε εάν ίσχυε η μηδενική υπόθεση. Για το SPSS η σχετική διαδικασία είναι ιδιαίτερα απλή. Ως συνέχεια της διαδικασίας που προτάθηκε παραπάνω (Analyse – Descriptive Statistics – Crosstabs), στην επιλογή Statistics (βλ. σχετική εικόνα) να επιλέξουμε το στατιστικό κριτήριο που μας ενδιαφέρει (στη συγκεκριμένη περίπτωση Chi-square)³⁸ και να ζητηθεί η εμφάνιση των παρατηρούμενων και αναμενόμενων συχνοτήτων σε απόλυτες τιμές και ποσοστά.



Εικόνα 47 & 48: Επιλογές του SPSS για το κριτήριο χ^2

Τα αποτελέσματα έχουν ως εξής:

³⁸ Οι δυνατές επιλογές είναι όσες εμφανίζονται στο σχετικό πλαίσιο διαλόγου. Για την επιλογή είναι απαραίτητο να εξετάζεται η τήρηση των σχετικών προδιαγραφών.

Φύλο * Βαθμός Πολιτικής Ενημέρωσης Crosstabulation

			Βαθμός Πολιτικής Ενημέρωσης					Total		
			Πάρα πολύ	Πολύ	Αρκετά	Λίγο	Πολύ λίγο		Καθόλου	
Παρατηρηθείσες Συχνότητες	Φύλο Κορίτσι	Count	24	19	17	17	2	2	81	
		Expected Count	15.2	16.4	12.7	28.4	4.5	3.7	81.0	
	Αναμενόμενες Συχνότητες	% within Φύλο	29.6%	23.5%	21.0%	21.0%	2.5%	2.5%	100.0%	
		% within Βαθμός Πολιτικής Ενημέρωσης	64.9%	47.5%	54.8%	24.6%	18.2%	22.2%	41.1%	
		% of Total	12.2%	9.6%	8.6%	8.6%	1.0%	1.0%	41.1%	
		Υπόλοιπα	Residual	8.8	2.6	4.3	-11.4	-2.5	-1.7	
	Αγόρι	Αγόρι	Count	13	21	14	52	9	7	116
			Expected Count	21.8	23.6	18.3	40.6	6.5	5.3	116.0
		Αναμενόμενες Συχνότητες	% within Φύλο	11.2%	18.1%	12.1%	44.8%	7.8%	6.0%	100.0%
			% within Βαθμός Πολιτικής Ενημέρωσης	35.1%	52.5%	45.2%	75.4%	81.8%	77.8%	58.9%
% of Total			6.6%	10.7%	7.1%	26.4%	4.6%	3.6%	58.9%	
Υπόλοιπα			Residual	-8.8	-2.6	-4.3	11.4	2.5	1.7	
Total		Total	Std. Residual	2.3	.6	1.2	-2.1	-1.2	-.9	
			Adjusted Residual	3.3	.9	1.7	-3.5	-1.6	-1.2	
		Count	37	40	31	69	11	9	197	
		Expected Count	37.0	40.0	31.0	69.0	11.0	9.0	197.0	
	% within Φύλο	18.8%	20.3%	15.7%	35.0%	5.6%	4.6%	100.0%		
	% within Βαθμός Πολιτικής Ενημέρωσης	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%		
% of Total	18.8%	20.3%	15.7%	35.0%	5.6%	4.6%	100.0%			

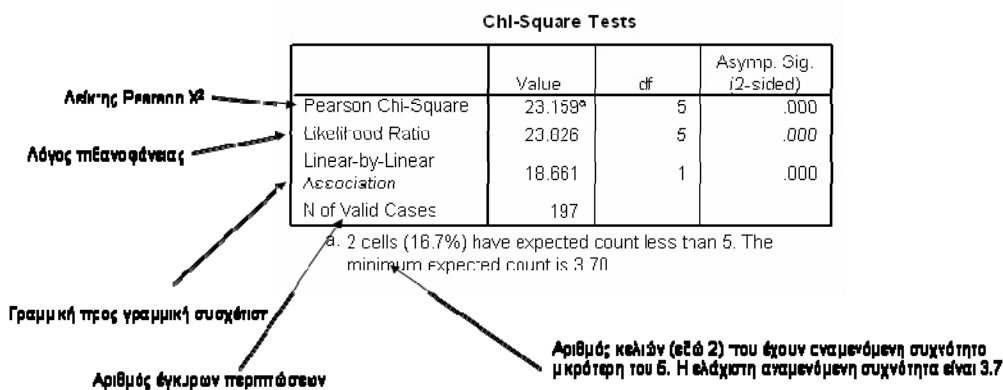
Πίνακας 7: Διασταύρωση της μεταβλητής Φύλο με τη μεταβλητή Βαθμός Πολιτικής Ενημέρωσης με εμφάνιση των παρατηρούμενων και αναμενόμενων συχνοτήτων

Φύλο * Βαθμός Πολιτικής Ενημέρωσης Crosstabulation

			Βαθμός Πολιτικής Ενημέρωσης					Total	
			Πάρα πολύ	Πολύ	Αρκετά	Λίγο	Πολύ λίγο		Καθόλου
Φύλο	Κορίτσι	Count	24	19	17	17	2	2	81
		Expected Count	15.2	16.4	12.7	28.4	4.5	3.7	81.0
		Residual	8.8	2.6	4.3	-11.4	-2.5	-1.7	
	Αγόρι	Count	13	21	14	52	9	7	116
		Expected Count	21.8	23.6	18.3	40.6	6.5	5.3	116.0
		Residual	-8.8	-2.6	-4.3	11.4	2.5	1.7	
Total	Count	37	40	31	69	11	9	197	
	Expected Count	37.0	40.0	31.0	69.0	11.0	9.0	197.0	

Εικόνα 49: Διασταύρωση της μεταβλητής Φύλο με τη μεταβλητή Βαθμός Πολιτικής Ενημέρωσης με εμφάνιση των παρατηρούμενων και αναμενόμενων συχνοτήτων.

Στον πίνακα εμφανίζονται οι παρατηρούμενες συχνοότητες, οι αναμενόμενες συχνοότητες, οι ποσοστιαίες αναγωγές και τα υπόλοιπα. Οι ενδείξεις υπολοίπων προκύπτουν από τη διαφορά παρατηρούμενων - αναμενόμενων συχνοτήτων. Θετική τιμή υπολοίπου σημαίνει ότι το ισοζύγιο παρατηρούμενων - αναμενόμενων συχνοτήτων βαρύνει υπέρ των παρατηρούμενων (οι παρατηρούμενες συχνοότητες είναι περισσότερες από τις αναμενόμενες). Όσο μεγαλύτερη είναι η διαφορά (θετική ή αρνητική), τόσο πιθανότερο είναι να προκύπτει στατιστικά σημαντικό αποτέλεσμα. Με τις κατάλληλες επιλογές η εφαρμογή επιστρέφει και τον πίνακα (Εικόνα 50):



Εικόνα 50: Αναλυτική παρουσίαση των αποτελεσμάτων του SPSS

- Ο δείκτης χ^2 είναι ο στατιστικός δείκτης που χρησιμοποιείται για την εκτίμηση της στατιστικής σημαντικότητας.
- Ο λόγος πιθανοφάνειας στα μεγάλα δείγματα συμπίπτει με την τιμή του χ^2 .
- Ο δείκτης γραμμικής προς γραμμική συσχέτιση παρέχει μια ένδειξη γραμμικής συσχέτισης των μεταβλητών.
- Τα δύο κελιά με μικρό αριθμό παρατηρήσεων συνιστούν πρόβλημα (περισσότερο εάν το ποσοστό υπερβαίνει το 25%). Στην τελευταία περίπτωση προτείνεται η επανακωδικοποίηση (recode) της μεταβλητής.

Από τον πίνακα συνάφειας των δύο μεταβλητών διαπιστώνουμε ότι οι παρατηρούμενες τιμές διαφέρουν συστηματικά από τις αναμενόμενες³⁹. Είναι όμως και στατιστικά σημαντική η διαφορά; Η τιμή του Pearson Chi-Square είναι ενδεικτική της εξάρτησης. Η βασική παραδοχή είναι ότι όσο η τιμή αυξάνει, τόσο αυξάνει και η πιθανότητα οι μεταβλητές να μην είναι ανεξάρτητες. Έτσι τιμές της πιθανότητας κοντά στο 1 (υπό)δηλώνουν ανεξαρτησία αφού απορρίπτεται η αρχική υπόθεση.

Για χ^2 (Pearson Chi-Square) 564.786, p-value (Significance level) 0.000 και επίπεδο σημαντικότητας 5% (σύνηθες στις έρευνες των Επιστημών της Αγωγής) η μηδενική υπόθεση απορρίπτεται, άρα οι δύο μεταβλητές είναι εξαρτημένες (βλ. ενδεικτικά και <http://home.clara.net/sisa/two2hlp.htm>, <http://www.graphpad.com/quickcalcs/contingency1.cfm>).

³⁹ Σημειώστε τις διαφορές ανάμεσα στις παρατηρούμενες και τις αναμενόμενες συχνότητες στη γραμμή των υπολοίπων.

6.2. Τυπολόγιο ανάγνωσης ενδείξεων για τους πίνακες του SPSS

Στη συνέχεια, παραθέτουμε μία μικρή εξήγηση για τις πιο σημαντικές ενδείξεις που «βγάζει» το SPSS σε αυτό το κριτήριο. Αρκετές από αυτές τις ενδείξεις συνήθως προϋποθέτουν αρκετά καλή γνώση της Στατιστικής. Για το λόγο αυτό ο αναγνώστης-κοινωνικός ερευνητής ας τις διαβάσει «ενημερωτικά» χωρίς να θεωρεί απαραίτητη την πλήρη κατανόηση όλων των όρων.

Count (καταμέτρηση περιπτώσεων) = πρόκειται για τον αριθμό που δηλώνει το πλήθος των παρατηρήσεων για κάθε κατηγορία.

Exp. count (αναμενόμενη καταμέτρηση περιπτώσεων) = πρόκειται για το πλήθος που θα εμφανίζονταν εάν οι δύο μεταβλητές ήταν εντελώς ανεξάρτητες μεταξύ τους.

Residual (υπόλοιπο περιπτώσεων) = αριθμός ο οποίος προκύπτει από τη διαφορά του πλήθους των παρατηρούμενων τιμών από τις αναμενόμενες.

Row Total (σύνολο γραμμής) = το συνολικό πλήθος τιμών για κάθε γραμμή.

Column Total (σύνολο στήλης) = το συνολικό πλήθος τιμών για κάθε στήλη.

Chi Square Pearson & Likelihood Ratio (X^2 και λόγος πιθανοφάνειας).

Degrees of Freedom (βαθμοί ελευθερίας) = πρόκειται για τον αριθμό που προκύπτει ως γινόμενο του αριθμού των κατηγοριών της μιας μεταβλητής -1 επί τον αριθμό των κατηγοριών της δεύτερης μεταβλητής -1, δηλαδή $(k-1)(\lambda-1)$

Significance (σημαντικότητα) = ο αριθμός δηλώνει την πιθανότητα τα αποτελέσματα που προέκυψαν να είναι τυχαία. Τιμές μεγαλύτερες (>) του 0.05 δηλώνουν ότι οι παρατηρούμενες τιμές δεν διαφέρουν με τρόπο στατιστικά σημαντικό από τις αναμενόμενες τιμές.

Linear by Linear association (γραμμική συσχέτιση) = με τη χρήση του συγκεκριμένου στατιστικού μέτρου εξετάζεται εάν υφίσταται (γραμμική) συσχέτιση ανάμεσα στις δύο μεταβλητές.

Minimum Expected Count (ελάχιστη αναμενόμενη εκτίμηση).

Cells (.%) have expected count less than 5 = Τιμές άνω του 25% καθιστούν την τιμή του X^2 προβληματική και πιθανώς μη-έγκυρη.

Phi (ϕ) = πρόκειται για τιμή η οποία αναφέρεται στην εκτίμηση της έντασης στη σχέση δύο ονομαστικών μεταβλητών. Ο συντελεστής Phi είναι συνάρτηση του χ^2 , του μεγέθους του δείγματος και ανεξάρτητος από την ταξινόμηση των τιμών των μεταβλητών και το πλήθος των τιμών. Αν οι μεταβλητές είναι ανεξάρτητες, τότε το ϕ τείνει στο μηδέν.

Cramer's V = η τιμή του συντελεστή Cramer's V αναφέρεται (επίσης) στην ισχύ της σχέσης ανάμεσα στις δύο (ονομαστικές) μεταβλητές. Εμπεριέχει τον συντελεστή ϕ και η διαφορά ως προς τον ϕ έγκειται στο γεγονός ότι η ϕ μπορεί να πάρει τιμές >1 , ενώ το εύρος τιμών του Cramer's V είναι ανάμεσα στο 0 και το 1.

Τέλος να σημειώσουμε ότι η δημοσίευση των αποτελεσμάτων σύμφωνα με το APA Style θα μπορούσε να έχει ως εξής: $\chi^2(5, N=200) = 23.159, p < 0.05$ ⁴⁰.

⁴⁰ American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: American Psychological Association, p.113

7

ΠΑΡΑΡΤΗΜΑΤΑ

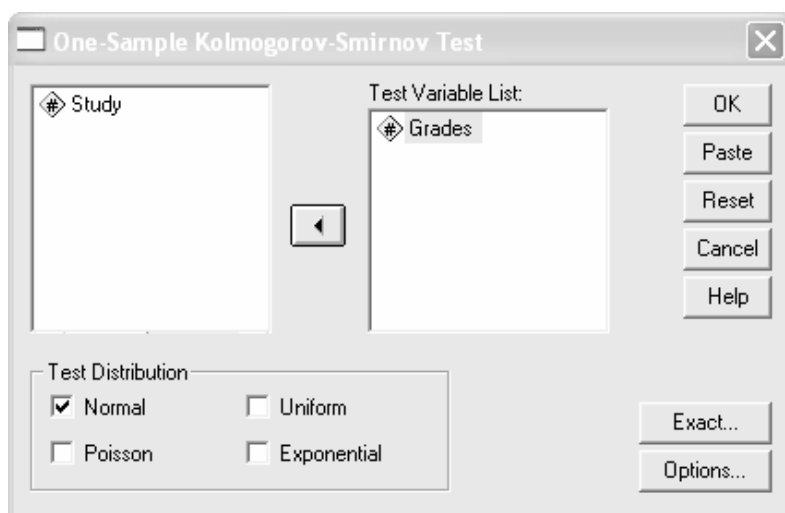
7.1 Έλεγχος κανονικότητας

Συνήθως η στατιστική σημαντικότητα είναι ένα επιθυμητό αποτέλεσμα μιας και οι ερευνητές προσπαθούν να επαληθεύσουν εναλλακτικές υποθέσεις (δεν έχει νόημα η επαλήθευση μηδενικών υποθέσεων—δεν ψάχνει ο ερευνητής τι ΔΕΝ ισχύει αλλά τι ισχύει). Εξαίρεση αποτελούν τα περισσότερα τεστ προϋποθέσεων (όχι όλα) για τα οποία επιθυμητό αποτέλεσμα είναι η υιοθέτηση της μηδενικής υπόθεσης. Αυτό φαίνεται ξεκάθαρα μόλις κάποιος διατυπώσει τις δύο υποθέσεις.

H_0 : Η κατανομή ΔΕΝ είναι στατιστικά σημαντικά διαφορετική από την κανονική

H_a : Η κατανομή είναι στατιστικά σημαντικά διαφορετική από την κανονική

Από τις παραπάνω υποθέσεις γίνεται φανερό ότι για να ικανοποιηθεί η προϋπόθεση της κανονικής κατανομής πρέπει να ισχύει η μηδενική υπόθεση. Για την διερεύνηση της προϋπόθεσης στη μεταβλητή επίδοση (grades) πηγαίνουμε στο μενού του SPSS και επιλέγουμε: Statistics → NonParametric tests → 1-Sample K-S.



Εικόνα 1: Εντολές του SPSS για το τεστ Kolmogorov-Smirnov

Με επιλογή της κανονικής κατανομής (Normal) ως υποθετική κατανομή, βλέπουμε ότι σχετικά με τη μεταβλητή επίδοση, η πιθανότητα να κατανέμεται αυτό το φαινόμενο κανονικά είναι 97%. Επομένως γίνεται δεκτή η μηδενική υπόθεση που επιβεβαιώνει την ύπαρξη της προϋπόθεσης της κανονικότητας.

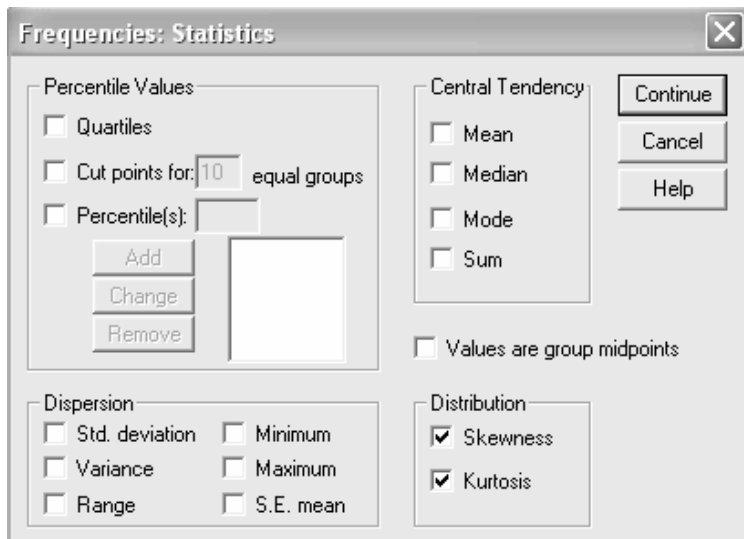
One-Sample Kolmogorov-Smirnov Test

		Grades
N		10
Normal Parameters ^{a,b}	Mean	14.4000
	Std. Deviation	4.45222
Most Extreme Differences	Absolute	.154
	Positive	.104
	Negative	-.154
Kolmogorov-Smirnov Z		.486
Asymp. Sig. (2-tailed)		.972

a. Test distribution is Normal.
b. Calculated from data.

Εικόνα 2: Αποτελέσματα του SPSS για το τεστ Kolmogorov-Smirnov

Ένας εναλλακτικός τρόπος αξιολόγησης της κανονικότητας είναι με τη χρήση της περιγραφικής στατιστικής και των δεικτών κυρτότητας και στρέβλωσης. Ο λόγος λ της κύρτωσης/στρέβλωσης προς το στατιστικό τους σφάλμα εκφράζει την πιθανότητα να κατανέμεται κανονικά το φαινόμενο που μελετάμε. Οι δείκτες αυτοί υπολογίζονται στο μενού συχνοτήτων. Έτσι επιλέγουμε: **Statistics** → **Descriptive Statistics** → **Frequencies** → Βάζετε στο παράθυρο "Variables" τη/τις μεταβλητή(ές) που θέλετε να αξιολογήσετε και κάντε κλικ στο "statistics".



Εικόνα 3: Εντολές του SPSS για εναλλακτικό τρόπο

Αφού επιλέξτε Skewness/Kurtosis, επιλέξτε «continue» → OK. Τα αποτελέσματα φαίνονται παρακάτω:

Statistics		
Grades		
N	Valid	10
	Missing	0
Skewness		-.690
Std. Error of Skewness		.687
Kurtosis		-.185
Std. Error of Kurtosis		1.334

Εικόνα 4: Αποτελέσματα του SPSS για εναλλακτικό τρόπο

Όταν ο λόγος λ ξεπερνά το ± 2 τότε υπάρχει ένδειξη ότι μια δεδομένη κατανομή αποκλίνει από την κανονικότητα είτε σε επίπεδο στρέβλωσης είτε σε επίπεδο κύρτωσης. Σχετικά με τα

παραπάνω δεδομένα, ο λόγος για την στρέβλωση ήταν $\lambda = \frac{Skew}{Skew_{Error}}$

είναι $-.690/.687 = -1.004$, ενώ για την κύρτωση $-.185/1.334 = -.1386$. Μιας και κανένας λόγος λ δεν είναι μεγαλύτερος από ± 2 τότε δεχόμαστε την υπόθεση ότι η μεταβλητή «επίδοση» κατανέμεται κανονικά και επομένως ικανοποιείται αυτή η προϋπόθεση. Αξιολόγηση της κανονικότητας μπορεί να γίνει και για υπο-ομάδες με το μενού «Explore».

7.2 Επιλογή κριτηρίου ελέγχου

7.2.1 Σύγκριση της ίδιας μεταβλητής σε δύο δείγματα

Όταν η μεταβλητή είναι κατηγορική και τα δείγματα είναι ανεξάρτητα.

Test χ^2 ομοιογένειας.

Προϋποθέσεις:

- (α) ο έλεγχος να γίνεται σε δείγματα μεγέθους > 30 ,
- (β) οι θεωρητικές τιμές που υπολογίζονται στον πίνακα συνάφειας να είναι > 1 και
- (γ) το 80% να είναι $>$ του 5

Όταν η μεταβλητή είναι κατηγορική και τα δείγματα είναι εξαρτημένα.

Test Mc Nemar.

Το test ελέγχει αν είναι στατιστικά σημαντικές οι αλλαγές που επήλθαν από την επίδραση κάποιου εξωγενούς παράγοντα.

Προϋποθέσεις:

- (α) οι μεταβλητές να είναι δίτιμες (dichotomous)
- (β) οι μεταβλητές να έχουν τις ίδιες τιμές (value labels)

Όταν η μεταβλητή είναι διατακτική και τα δείγματα είναι ανεξάρτητα.

Test Mann & Whitney

Το test ελέγχει εάν οι μέσες τιμές των δύο πληθυσμών μπορούν να θεωρηθούν ίσες συγκρίνοντας τις διαμέσους των δύο δειγμάτων.

Όταν η μεταβλητή είναι διατακτική και τα δείγματα είναι εξαρτημένα.

Test προσήμου (Sign test) ή test Wilcoxon. Το Sign test προτείνεται για τη διερεύνηση του ποιοτικού μέρους (ποιο δείγμα υπερέχει), ενώ το Wilcoxon test μετρά πόσο υπερέχει. Η βιβλιογραφία εκτιμά ότι το Sign test είναι αυστηρότερο του test Wilcoxon.

Όταν η μεταβλητή είναι ποσοτική και τα δείγματα είναι ανεξάρτητα.

T - test

Προϋπόθεση:

(α) τα δύο δείγματα θα πρέπει να έχουν κανονικές κατανομές. Αυτό σημαίνει ότι θα πρέπει να προηγηθεί έλεγχος κανονικότητας. Στην περίπτωση που οι κατανομές δεν ακολουθούν κανονική κατανομή προτείνεται η χρήση του Mann - Whitney test. Ανάλογα με το εάν οι διασπορές των δύο δειγμάτων είναι ίσες ή όχι επιλέγεται και ειδική εκδοχή του t-test. Ο έλεγχος ισότητας των διασπορών γίνεται με χρήση της F - κατανομής (F-test).

Όταν η μεταβλητή είναι ποσοτική και τα δείγματα είναι εξαρτημένα.

T - test ζευγαρωτών παρατηρήσεων

Προϋπόθεση:

(α) τα δύο δείγματα θα πρέπει να έχουν κανονικές κατανομές. Αυτό σημαίνει ότι θα πρέπει να προηγηθεί έλεγχος κανονικότητας. Στην περίπτωση που οι κατανομές δεν ακολουθούν κανονική κατανομή προτείνεται η χρήση του test Wilcoxon.

7.2.2 Σύγκριση της ίδιας μεταβλητής σε περισσότερα από δύο δείγματα.

Όταν η μεταβλητή είναι κατηγορική και τα δείγματα είναι ανεξάρτητα.

Πίνακας Συνάφειας. Ο πίνακας έχει n γραμμές (όπου n = ο αριθμός των δειγμάτων) και k στήλες (όπου k = οι τιμές της μεταβλητής). Μπορεί να ισχύει και το αντίστροφο.

χ^2 test ομοιογένειας. Προσοχή το χ^2 εφαρμόζεται στις συχνότητες και όχι στα ποσοστά.

Προϋπόθεση:

(α) σε όλα τα κελιά του πίνακα συνάφειας να έχουμε θεωρητικές συχνότητες >1 και για το 80% >5 .

Όταν η μεταβλητή είναι κατηγορική και τα δείγματα είναι εξαρτημένα.

Πίνακας Συνάφειας. Ο πίνακας έχει n γραμμές (όπου $n =$ ο αριθμός των περιπτώσεων) και k στήλες (όπου $k =$ ο αριθμός των δειγμάτων).

Προϋποθέσεις:

(α) οι μεταβλητές να είναι δίτιμες (dichotomous)

Test Cochran ($Q - test$).

Όταν η μεταβλητή είναι διατακτική και τα δείγματα είναι ανεξάρτητα.

Kruskal & Wallis test (μη παραμετρικό test). Βασική παραδοχή αποτελεί η υπόθεση ότι οι διάμεσοι των πληθυσμών είναι ίσες.

Όταν η μεταβλητή είναι διατακτική και τα δείγματα είναι εξαρτημένα.

Friedman test. Βασική παραδοχή αποτελεί η υπόθεση ότι οι διάμεσοι των πληθυσμών είναι ίσες.

Όταν η μεταβλητή είναι συνεχής και τα δείγματα είναι ανεξάρτητα.

Ανάλυση Διασποράς ANOVA.

Προϋποθέσεις:

(α) η κατανομή των παρατηρήσεων θα πρέπει να ακολουθεί την κανονική κατανομή,

(β) η μεταβλητή να παρουσιάζει την ίδια διασπορά σε όλους τους πληθυσμούς.

Προσοχή στον υπολογισμό και την εκτίμηση των υπολοίπων.

Όταν η μεταβλητή είναι συνεχής και τα δείγματα είναι εξαρτημένα.

Σε γενικές γραμμές υιοθετείται η Ανάλυση Διασποράς.

7.2.3 Σύγκριση δύο μεταβλητών στο ίδιο δείγμα

Όταν η εξαρτημένη μεταβλητή είναι συνεχής και η ανεξάρτητη επίσης συνεχής.

Γραμμική παλινδρόμηση και υπολογισμός του συντελεστή γραμμικής συσχέτισης Pearson.

Προϋπόθεση:

Η κατανομή των παρατηρήσεων θα πρέπει να ακολουθεί την κανονική κατανομή.

Όταν η εξαρτημένη μεταβλητή είναι συνεχής και η ανεξάρτητη είναι διατακτική.

Μη παραμετρική παλινδρόμηση και υπολογισμός του συντελεστή Spearman.

Όταν η εξαρτημένη μεταβλητή είναι συνεχής και η ανεξάρτητη είναι κατηγορική.

Ανάλυση Διασποράς με έναν παράγοντα και υπολογισμός του συντελεστή η^2 .

Παρατήρηση: ανάλογα με την κατανομή της εξαρτημένης μεταβλητής χρησιμοποιείται παραμετρικό ή μη-παραμετρικό test.

Όταν η εξαρτημένη μεταβλητή είναι διατακτική και η ανεξάρτητη είναι συνεχής.

Μη παραμετρική παλινδρόμηση και υπολογισμός του συντελεστή Spearman.

Όταν η εξαρτημένη μεταβλητή είναι διατακτική και η ανεξάρτητη είναι επίσης διατακτική.

Μη παραμετρική παλινδρόμηση και υπολογισμός του συντελεστή Spearman.

Όταν η εξαρτημένη μεταβλητή είναι διατακτική και η ανεξάρτητη είναι κατηγορική.

χ^2 test.

Όταν η εξαρτημένη μεταβλητή είναι κατηγορική και η ανεξάρτητη είναι συνεχής.

χ^2 test.

Όταν η εξαρτημένη μεταβλητή είναι κατηγορική και η ανεξάρτητη είναι διατακτική.

χ^2 test.

Όταν η εξαρτημένη μεταβλητή είναι κατηγορική και η ανεξάρτητη είναι επίσης κατηγορική.

χ^2 test.

8

ΒΙΒΛΙΟΓΡΑΦΙΑ

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Stevens, S. (1946) On the theory of scales and measurement. *Science*, 103, 667-680.
- Κατσιόλλης, Ι. (1998). *Οι μικροϋπολογιστές στις κοινωνικές επιστήμες: επιστημονική εμπειρική έρευνα και στατιστικές αναλύσεις*. Αθήνα: Gutenberg.
- Μιχαλοπούλου, Κ. (1992). *Κλίμακες Μέτρησης Στάσεων*. Αθήνα: Οδυσσέας
- Τσάντας, Ν., Μωυσιάδης, Χ., Μπαγιάτης, Ν., & Χατζηπαντελής, Θ. (1999). *Ανάλυση δεδομένων με τη βοήθεια στατιστικών πακέτων*. Θεσσαλονίκη: Ζήτη.
- Norusis, M. (2005). *Οδηγός Ανάλυσης Δεδομένων με το SPSS 12.0* (Κ.Περάκης, Trans.). Αθήνα: Κλειδάριθμος.
- Yerkes, R. M., & Dowdson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative and Neurological Psychology*, 18, 459-482.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Ρούσσο, Π., & Τσαούσης, Ι. (2003). *Στατιστική εφαρμοσμένη στις κοινωνικές επιστήμες*. Αθήνα: Ελληνικά Γράμματα
- Champion, D. J. (1981). *Basic statistics for social research* (2d ed.). New York: Macmillan.

Χρήσιμες διευθύνσεις στο Διαδίκτυο (εκτός αυτών που αναφέρονται στο κείμενο)

<http://www.ats.ucla.edu/stat/spss/> (ιδιαίτερα στις θέσεις SPSS Starter Kit, What statistical analysis should I use?, Annotated Output και την κατηγορία Links by Topic)

<http://www.ats.ucla.edu/stat/spss/notes2/analyze.htm>
(προσφέρει πολύ καλή ανάλυση για τις εντολές που πρέπει να ακολουθούνται για κάθε τύπο ανάλυσης)

Για τον **συντελεστή συσχέτισης** βλ. ενδεικτικά:
http://bcs.whfreeman.com/ips4e/cat_010/applets/CorrelationRegression.html

<http://www.stat.vt.edu/~sundar/java/applets/Correlation.html>

http://bcs.whfreeman.com/bps3e/content/cat_010/applets/twovarcalcbps.html

<http://bcs.whfreeman.com/bps3e/>

<http://davidmlane.com/hyperstat/index.html>

Για το **T-TEST** βλ. ενδεικτικά:

http://www.une.edu.au/WebStat/unit_materials/c6_common_statistical_tests/index.html

Για την **Ανάλυση Διακύμανσης** (ANOVA) βλ. ενδεικτικά:

<http://www.physics.csbsju.edu/stats/anova.html>

<http://web.umn.edu/~psyworld/virtualstat/anova/anovacalc.html>

<http://faculty.vassar.edu/lowry/ank3.html>

<http://faculty.vassar.edu/lowry/ank4.html>

http://www.une.edu.au/WebStat/unit_materials/c7_anova/index.html

<http://www.graphpad.com/quickcalcs/posttest1.cfm>

<http://home.ubalt.edu/ntsbarsh/Business-stat/otherapplets/ANOVA2Rep.htm>

<http://faculty.vassar.edu/lowry/corr3.html>

<http://home.ubalt.edu/ntsbarsh/Business-stat/otherapplets/ANOVAdep.htm>

<http://faculty.vassar.edu/lowry/anova2x2.html>

<http://faculty.vassar.edu/lowry/anova2x3.html>

<http://home.ubalt.edu/ntsbarsh/Business-stat/otherapplets/ANOVATwo.htm>

<http://faculty.vassar.edu/lowry/corr4>

http://www.une.edu.au/WebStat/unit_materials/c7_anova/twoway_anova.htm