Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών
Τμήμα Φιλοσοφίας, Παιδαγωγικής και Ψυχολογίας
Τομέας Ψυχολογίας

Εισαγωγή
στη Στατιστική Επεξεργασία Δεδομένων
με το SPSS for Windows

Επιμέλεια:
Λέκτορας Βασίλης Γ. Παυλόπουλος

Αθήνα, 2006

Περιεχόμενα

1. ΕΙΣΑΓΩΓΗ

Το SPSS (Statistical Package for Social Sciences) είναι ένα ισχυρό πρόγραμμα στατιστικής επεξεργασίας δεδομένων, το οποίο έχει φτάσει ήδη αισίως στη 14η έκδοση. Η στατιστική ανάλυση με το SPSS ακολουθεί συνήθως την παρακάτω διαδικασία:

1. Καταχωρίζουμε τα στοιχεία στο παράθυρο επεξεργασίας δεδομένων (Data editor).
2. Ελέγχουμε την επάρκεια των δεδομένων για στατιστική ανάλυση (βλ. Παράρτημα Β) και, ενδεχομένως, κάνουμε τις απαραίτητες διορθωτικές κινήσεις.
3. Επιλέγουμε τη στατιστική ανάλυση από το μενού της εντολής «Analyze».
4. Συμπληρώνουμε το πλαίσιο διαλόγου της στατιστικής ανάλυσης με τις απαιτούμενες πληροφορίες και εκτελούμε την εντολή.
5. Το αποτέλεσμα, σε μορφή πινάκων ή/και σχημάτων, εμφανίζεται στο παράθυρο προβολής αποτελεσμάτων (Viewer). Από εδώ μπορούμε να το επεξεργαστούμε, να το αποθηκεύσουμε ή να το εκτυπώσουμε.

Προσοχή! Σε όλα τα παραπάνω βήματα πρέπει να γνωρίζουμε εκ των προτέρων ποιο ερευνητικό ερώτημα θα επιχειρήσουμε να απαντήσουμε και ποια ανάλυση ενδείκνυται για τη φύση των δεδομένων και τους σκοπούς της έρευνας (βλ. και Παράρτημα Α). Επίσης, πρέπει να είμαστε σε θέση να «διαβάσουμε» τα αποτελέσματα για να τα αξιοποιήσουμε. Το SPSS δεν κάνει τίποτα από αυτά για μας – ευτυχώς!...

Άνοιγμα αρχείου. Μπορούμε να ανοίξουμε ένα υπάρχον αρχείο επιλέγοντας από τη γραμμή εντολών «File» → «Open» και, στη συνέχεια, τον τύπο του αρχείου (βλ. παρακάτω). Όταν το κάνουμε, εμφανίζεται ένα πλαίσιο διαλόγου όπου μπορούμε να βρούμε το αρχείο από τον αλφαβητικό κατάλογο με τα διαθέσιμα αρχεία του φακέλου.

Αποθήκευση της εργασίας μας. Για να αποθηκεύσουμε τα περιεχόμενα ενός παραθύρου του SPSS, επιλέγουμε από τη γραμμή εντολών «File» → «Save». Αν είναι η πρώτη φορά που αποθηκεύουμε το αρχείο, θα εμφανιστεί το πλαίσιο διαλόγου «Save As…», όπου θα πληκτρολογήσουμε το όνομα που επιθυμούμε. Ανάλογα με τον τύπο του αρχείου, το πρόγραμμα δίνει αυτόματα την αντίστοιχη προέκταση (βλ. παρακάτω). Ας σημειωθεί ότι το SPSS δεν αποθηκεύει αυτόματα τα περιεχόμενα όλων των παραθύρων που είναι ανοιχτά, αλλά μόνο τα περιεχόμενα του ενεργού παράθυρου εργασίας.

Έξοδος από το SPSS. Από τη γραμμή εντολών, επιλέγουμε «File» → «Exit». Αν ξεχάσουμε να αποθηκεύσουμε την εργασία μας πριν κλείσουμε το SPSS, το ίδιο το πρόγραμμα θα μας το υπενθυμίσει κατά την έξοδο.

2. ΤΑ ΠΑΡΑΘΥΡΑ ΤΟΥ SPSS

Ανάλογα με το είδος της εργασίας που εκτελείται, το SPSS εμφανίζει στην επιφάνεια εργασίας διάφορα παράθυρα ταυτοχρόνως:

- **Το παράθυρο επεξεργασίας δεδομένων (Data editor).** Εδώ μπορούμε να δημιουργήσουμε ένα αρχείο δεδομένων. Από την έκδοση 10 του SPSS και εξής, υπάρχουν διαθέσιμοι δύο τρόποι προβολής του παραθύρου αυτού: η **προβολή δεδομένων (Data view)** και η **προβολή μεταβλητών (Variable view)**. Το παράθυρο επεξεργασίας δεδομένων ανοίγει αυτόματα με την εκκίνηση του SPSS και στη γραμμή τίτλου αναγράφεται το όνομα «Untitled». Όταν αποθηκεύσουμε τα δεδομένα, στη γραμμή τίτλου εμφανίζεται το όνομα που δώσαμε.
- **Το παράθυρο προβολής αποτελεσμάτων (Viewer).** Εδώ εμφανίζονται τα αποτελέσματα των στατιστικών αναλύσεων, με τη μορφή πινάκων ή/και σχημάτων. Το παράθυρο ανοίγει αυτόματα όταν εκτελέσουμε μία στατιστική ανάλυση και στη γραμμή τίτλου αναγράφεται το όνομα «Output1». Εάν αποθηκεύσουμε τα αποτελέσματα, στη γραμμή τίτλου αναγράφεται το όνομα που δώσαμε.

- **Το παράθυρο επεξεργασίας εντολών (Syntax editor).** Εδώ μπορούμε να πληκτρολογήσουμε εντολές στατιστικών αναλύσεων, αντί να τις επιλέξουμε από τη γραμμή εντολών. Το παράθυρο αυτό δεν ανοίγει αυτόματα και η λειτουργία του προϋποθέτει γνώση σύνταξης της γλώσσας εντολών του SPSS.
- **Το παράθυρο επεξεργασίας γραφημάτων (Chart editor)**. Εδώ μπορούμε να επεξεργαστούμε ένα γράφημα που έχουμε ήδη δημιουργήσει με το SPSS.

3. ΤΥΠΟΙ ΑΡΧΕΙΩΝ ΤΟΥ SPSS

Ανάλογα με το περιεχόμενό τους, τα αρχεία του SPSS μπορεί να είναι:

- **Αρχεία δεδομένων (Data files).** Παίρνουν αυτόματα την προέκταση *.SAV. Περιέχουν τα δεδομένα που καταχωρίζουμε στο παράθυρο επεξεργασίας δεδομένων (Data editor).
- **Αρχεία αποτελεσμάτων (Output files).** Παίρνουν αυτόματα την προέκταση *.SPO. Περιέχουν τα αποτελέσματα των στατιστικών αναλύσεων. Μπορούμε να τα επεξεργαστούμε μέσα από το παράθυρο προβολής αποτελεσμάτων (Viewer). Από εκεί μπορούμε επίσης, αν θέλουμε, να τα αντιγράψουμε και να τα επικολλήσουμε απευθείας στο κείμενο των ευρημάτων της έρευνας.
- **Αρχεία εντολών (Syntax files).** Παίρνουν αυτόματα την προέκταση *.SPS. Περιέχουν εντολές του SPSS.
- **Αρχεία γραφημάτων (Chart files).** Παίρνουν αυτόματα την προέκταση *.CHT. Περιέχουν γραφήματα που δημιουργήθηκαν με εντολές του SPSS.

4. ΔΗΜΙΟΥΡΓΙΑ ΑΡΧΕΙΟΥ ΔΕΔΟΜΕΝΩΝ

4.1. Καταχώριση των δεδομένων

Η καταχώριση των δεδομένων γίνεται στην **προβολή δεδομένων (Data view)** του **παράθυρου επεξεργασίας δεδομένων (Data editor)**, η οποία εμφανίζεται αυτόματα με την εκκίνηση του SPSS. Το παράθυρο αυτό έχει τη μορφή ενός τεράστιου φύλλου εργασίας, δηλ. είναι ουσιαστικά ένας πίνακας με σειρές και στήλες.

Οι **στήλες** του αρχείου δεδομένων αντιπροσωπεύουν **μεταβλητές** (π.χ. φύλο, ηλικία, τις απαντήσεις στις ερωτήσεις ενός ερωτηματολογίου, κλπ.), ενώ στις **σειρές** καταχωρίζονται τα στοιχεία για κάθε **άτομο** που συμμετείχε στην έρευνα. Αν, για παράδειγμα, σε κάθε συμμετέχοντα χορηγήθηκαν δύο ερωτηματολόγια, τα δεδομένα και των δύο αυτών ερωτηματολογίων θα καταχωριστούν στην ίδια σειρά, εφόσον συμπληρώθηκαν από το ίδιο άτομο. Επομένως, αφού ολοκληρώσουμε την καταχώριση των δεδομένων, το αρχείο θα περιέχει τόσες σειρές όσα τα άτομα του δείγματος, και τόσες στήλες όσες οι μεταβλητές της έρευνας.

Μόλις καταχωρίσουμε ένα στοιχείο σε φατνίο (cell) του αρχείου δεδομένων, αυτόματα το SPSS θεωρεί ότι προσθέτουμε ένα άτομο στο δείγμα μας, ακόμα και αν όλα τα υπόλοιπα φατνία της σειράς αυτής είναι κενά. Το SPSS εκλαμβάνει τα κενά φατνία ως ελλιπή στοιχεία (missing data). Σε περίπτωση που επιθυμούμε να διαγράψουμε για οποιοδήποτε λόγο μια σειρά του αρχείου δεδομένων (δηλαδή θέλουμε να αφαιρέσουμε ένα άτομο από το δείγμα), επιλέγουμε το γκρίζο φατνίο που βρίσκεται στο αριστερό περιθώριο της σειράς (και που αναγράφει τον αύξοντα αριθμό της) και πατάμε το πλήκτρο «Delete» ή επιλέγουμε «Edit» → «Cut» από τη γραμμή εντολών.

Αν και το SPSS παρέχει τη δυνατότητα χειρισμού πληροφοριών που εκφράζονται με σύμβολα ή γλωσσικούς χαρακτήρες, στην πλειοψηφία των ερευνών το αρχείο δεδομένων περιέχει αποκλειστικά αριθμούς. Αυτό σημαίνει ότι οι πληροφορίες που προέρχονται από κατηγορικές μεταβλητές ή ποιοτικές διαβαθμίσεις πρέπει να **κωδικο-**

ποιηθούν σε αριθμητική μορφή. Για παράδειγμα, τα κατηγορικά δεδομένα της μεταβλητής «φύλο» μπορούν να κωδικοποιηθούν ως 1 = «άνδρας» και 2 = «γυναίκα». Φυσικά, η κωδικοποίηση θα μπορούσε να γίνει και αντίστροφα, αφού οι αριθμοί εκφράζουν κατηγορίες και όχι ποσότητες. Παρομοίως, τα ποιοτικά δεδομένα της μεταβλητής «επίπεδο εκπαίδευσης» μπορούν να κωδικοποιηθούν ως 1 = «Α/θμια», 2 = «Β/θμια» και 3 = «Γ/θμια». Εδώ όμως η σειρά της κωδικοποίησης έχει μια συγκεκριμένη λογική (είναι αύξουσα) εφόσον η διαφορά ανάμεσα στις τρεις τιμές-επίπεδα της μεταβλητής «επίπεδο εκπαίδευσης» είναι ποσοτική, αν και όχι ίσων διαστημάτων: όσο μεγαλύτερη η τιμή, τόσο ανώτερη η βαθμίδα εκπαίδευσης.

4.2. Παραμετροποίηση των μεταβλητών

Για να ορίσουμε διάφορες παραμέτρους των μεταβλητών του αρχείου δεδομένων, χρησιμοποιούμε την **προβολή μεταβλητών (Variable view)** του **παράθυρου επεξεργασίας δεδομένων (Data editor)**[1] επιλέγοντας το αντίστοιχο καρτελάκι στο κάτω αριστερό μέρος του παραθύρου Data editor. Σε αυτή την κατάσταση προβολής κάθε σειρά περιέχει τα στοιχεία μίας μεταβλητής, ως εξής:

- **Name.** Τα ονόματα των μεταβλητών στο SPSS είναι ενιαίες λέξεις (χωρίς κενά). Εκτός από αλφαβητικούς χαρακτήρες, μπορούν επιπλέον να χρησιμοποιηθούν τα σύμβολα . @ # _ $. Ωστόσο, ένα όνομα πρέπει υποχρεωτικά να αρχίζει μόνο με αλφαβητικό χαρακτήρα, ενώ δεν επιτρέπεται η χρήση των συμβόλων ! ? ' * . Αν και στις νεότερες εκδόσεις του προγράμματος είναι αποδεκτοί οι χαρακτήρες του ελληνικού αλφάβητου, όπως και τα μεγάλα ονόματα (άνω των 8 χαρακτήρων), πάντως συνιστάται η χρήση ονομάτων από 8, το πολύ, λατινικούς χαρακτήρες. Έτσι διασφαλίζεται η συμβατότητα του αρχείου δεδομένων μεταξύ εκδόσεων του SPSS και μεταξύ υπολογιστών.
- **Type.** Εδώ ορίζουμε τον τύπο της μεταβλητής (π.χ. αριθμητική, αλφαριθμητική, ημερομηνία, κλπ.), τον αριθμό των ακέραιων και των δεκαδικών ψηφίων των αριθμητικών μεταβλητών.
- **Labels.** Προαιρετικά, μπορούμε να δώσουμε εκτενή περιγραφή της μεταβλητής, περισσότερο κατανοητή από ό,τι το όνομα, η οποία εμφανίζεται στα αποτελέσματα των στατιστικών αναλύσεων για να διευκολύνει την ανάγνωσή τους. Για παράδειγμα, στη μεταβλητή «educfa» μπορούμε να χρησιμοποιήσουμε ως label τη φράση «education of father».
- **Values.** Προαιρετικά, επίσης, μπορούμε να δώσουμε μια σύντομη περιγραφή για τις τιμές-επίπεδα των κατηγορικών μεταβλητών. Αν, π.χ., έχουμε κωδικοποιήσει τις βαθμίδες εκπαίδευσης ως 1, 2 και 3, μπορούμε να χρησιμοποιήσουμε τα value labels 1 = «Α/θμια», 2 = «Β/θμια» και 3 = «Γ/θμια». Οι περιγραφές αυτές θα εμφανίζονται αντί για τις τιμές 1, 2 και 3, ώστε να διευκολύνουν την ανάγνωση των ευρημάτων. Σε κάθε περίπτωση, πάντως, συνιστάται να διατηρούμε ένα αντίγραφο του πρωτοκόλλου κωδικοποίησης των δεδομένων.
- **Missing values.** Μπορούμε να ορίσουμε κάποια τιμή που θα καταχωρίζουμε όταν δεν υπάρχουν διαθέσιμα στοιχεία για ένα άτομο. Αν, για παράδειγμα, δεν γνωρίζουμε το μορφωτικό επίπεδο ενός ατόμου, μπορούμε να ορίσουμε την τιμή 9 ως missing value. Πάντως, εάν αφήσουμε κενά τα σχετικά φατνία, το SPSS καταλαβαίνει αυτόματα ότι πρόκειται για ελλιπή στοιχεία και ονομάζει τις πληροφορίες αυτές ως system missing.
- **Columns.** Μπορούμε να αλλάξουμε την προκαθορισμένη ρύθμιση για το πλάτος της στήλης στην προβολή δεδομένων. Αυτό δεν έχει κάποια επίπτωση στις στατιστικές αναλύσεις.

[1] Μέχρι και την έκδοση 9 του SPSS, η παραμετροποίηση των μεταβλητών γινόταν μέσω του πλαισίου διαλόγου της εντολής «Data» → «Define variable».

- **Measure.** Επιλέγουμε το είδος των δεδομένων της μεταβλητής: αριθμητικά (scale), ποιοτικά-τακτικές τιμές (ordinal) ή κατηγορικά (nominal).

5. ΤΡΟΠΟΠΟΙΗΣΗ ΚΑΙ ΕΠΙΛΟΓΗ ΔΕΔΟΜΕΝΩΝ

Το SPSS επιτρέπει να τροποποιήσουμε τις μεταβλητές και τα δεδομένα, αφού τα έχουμε αποθηκεύσει στον υπολογιστή. Αυτή η δυνατότητα είναι ιδιαίτερα χρήσιμη επειδή πολλές φορές οι στατιστικές αναλύσεις δεν αφορούν στις αρχικές μεταβλητές αλλά σε παράγωγες μεταβλητές, δηλ. αλγεβρικούς συνδυασμούς των αρχικών στοιχείων (για παράδειγμα, το άθροισμα των απαντήσεων ενός ερωτηματολογίου). Οι διαθέσιμες επιλογές βρίσκονται στην εντολή «Transform» της γραμμής εντολών.

COMPUTE. Πραγματοποιεί αλγεβρικές πράξεις με τα δεδομένα μίας ή περισσότερων μεταβλητών και αποθηκεύει τα αποτελέσματα σε νέα μεταβλητή.

Έστω ότι έχουμε ένα ερωτηματολόγιο που περιλαμβάνει 10 κλειστές ερωτήσεις (ας τις ονομάσουμε «q1», «q2», ..., «q10») με πιθανές απαντήσεις 1 = «Ναι» και 0 = «Όχι». Ενδέχεται να θέλουμε να υπολογίσουμε μία συνολική τιμή για κάθε άτομο, προσθέτοντας τις απαντήσεις που έδωσε σε όλες τις ερωτήσεις. Επιλέγουμε τις εντολές: «TRANSFORM» → «COMPUTE…». Στο πλαίσιο διαλόγου που εμφανίζεται, στη θέση «Target variable» γράφουμε το όνομα της νέας μεταβλητής που θα δημιουργήσουμε, (π.χ. «score») ενώ στη θέση «Numeric expression» γράφουμε τον αλγεβρικό χειρισμό, π.χ. «q1 + q2 + q3 + q4 + q5 + q6 + q7 + q8 + q9 + q10». Η εκτέλεση της εντολής δημιουργεί μια νέα μεταβλητή με το όνομα «score», η οποία αποτελείται από το άθροισμα των 10 ερωτήσεων του ερωτηματολογίου.

Όταν κάνετε πολύπλοκους υπολογισμούς με την εντολή «Compute» θυμηθείτε να βάλετε σε παρένθεση τις πράξεις που πρέπει να προηγηθούν (π.χ. τους πολλαπλασιασμούς ή τις διαιρέσεις), ώστε να μην προκύψουν παραπλανητικά αποτελέσματα.

RECODE. Αλλάζει την αρχική κωδικοποίηση των τιμών μιας μεταβλητής.

Ας πούμε ότι στη μεταβλητή «educ» («επίπεδο εκπαίδευσης») θέλουμε να ενώσουμε τους «απόφοιτους Α/θμιας» (τιμή 1) μαζί με τους «απόφοιτους Β/θμιας» (τιμή 2) σε μία ομάδα, διατηρώντας παράλληλα τους «απόφοιτους Γ/θμιας» (τιμή 3) ως ξεχωριστή ομάδα. Θα δώσουμε την εντολή: «TRANSFORM» → «RECODE» → «INTO DIFFERENT VARIABLES». Στο παράθυρο που ανοίγει, επιλέγουμε τη μεταβλητή που σκοπεύουμε να τροποποιήσουμε (π.χ. την «educ») και τη μεταφέρουμε στο πλαίσιο «Input variable». Στη θέση «Output variable» γράφουμε ένα όνομα για τη νέα μεταβλητή που θα δημιουργηθεί (ας πούμε, «educ2»). Στη συνέχεια, από το κουμπί «Old and new values» μεταφερόμαστε σε νέο πλαίσιο διαλόγου, όπου γράφουμε την αρχική τιμή («Old value»), τη νέα τιμή («New value»), επιλέγουμε το κουμπί «Add» και επαναλαμβάνουμε μέχρι να αντιστοιχήσουμε όλες τις αρχικές τιμές με τις νέες (π.χ. 1 → 1, 2 → 1, 3 → 2). Με τα κουμπιά «Continue» και «OK» ενεργοποιούμε την εντολή. Ως αποτέλεσμα, δημιουργείται μια νέα μεταβλητή, η «educ2», η οποία περιέχει δύο (αντί για τρεις) κωδικοποιημένες τιμές για το επίπεδο εκπαίδευσης, 1 = «Α/θμια+Β/θμια» και 2 = «Γ/θμια».

Οι πιο παρατηρητικοί ίσως προσέξουν ότι το SPSS μας δίνει τη δυνατότητα να ανακωδικοποιήσουμε τις τιμές μιας μεταβλητής χωρίς να δημιουργήσουμε νέα (μέσα από την εντολή «TRANSFORM» → «RECODE» → «INTO SAME VARIABLES»). Σε αυτή την περίπτωση όμως δεν θα μπορέσουμε να επιστρέψουμε στις αρχικές τιμές αν για κάποιο λόγο το θελήσουμε. Επομένως, η χρήση της επιλογής αυτής απαιτεί περίσκεψη γιατί μπορεί να προκαλέσει απώλεια δεδομένων.

Η εντολή «Recode» είναι ιδιαίτερα χρήσιμη σε έρευνες με ερωτηματολόγια που περιέχουν ερωτήσεις αντίστροφης βαθμολόγησης. Για παράδειγμα, ένα ερωτηματολόγιο ενδοπροσωπικού-εξωπροσωπικού ελέγχου (locus of control) περιλαμβάνει μερι-

κές ερωτήσεις ενδοπροσωπικού ελέγχου και μερικές ερωτήσεις εξωπροσωπικού ελέγχου. Πριν υπολογίσουμε το συνολικό βαθμό, ας πούμε, ενδοπροσωπικού ελέγχου, πρέπει πρώτα να εφαρμόσουμε την εντολή «Recode» στις ερωτήσεις εξωπροσωπικού ελέγχου (αν οι πιθανές απαντήσεις ήταν 1 = «Ναι» και 0 = «Όχι», τότε θα αντιστρέψουμε την κωδικοποίηση, δηλαδή: 0 → 1 και 1 → 0). Με τον τρόπο αυτό εξασφαλίζουμε ότι όλες οι θετικές απαντήσεις των συμμετεχόντων (τιμή 1) δηλώνουν το ίδιο κέντρο ελέγχου (στο παράδειγμά μας, ενδοπροσωπικό).

Ας σημειωθεί ότι οι εντολές για τροποποίηση των δεδομένων και δημιουργία νέων μεταβλητών πρέπει να εκτελούνται μετά την ολοκλήρωση της καταχώρισης των στοιχείων της έρευνας. Π.χ., αν καταχωρίσουμε νέα στοιχεία στο αρχείο δεδομένων αφού έχουμε ήδη δημιουργήσει μερικές παράγωγες μεταβλητές με την εντολή «Compute», οι τιμές των μεταβλητών αυτών δεν θα συμπληρωθούν αυτόματα για τους νέους συμμετέχοντες. Δηλ. πρέπει να επαναλάβουμε την εντολή «Compute» ώστε να δημιουργήσουμε ξανά τις παράγωγες μεταβλητές.

***

Μπορούμε επίσης, αν θέλουμε, να επιλέξουμε ένα μόνο μέρος των δεδομένων για να εφαρμόσουμε μία στατιστική ανάλυση. Αυτό γίνεται από την εντολή «Select cases». Παρακάτω παρουσιάζεται μία από τις πιο συνηθισμένες περιπτώσεις χρήσης της εντολής αυτής:

SELECT CASES. Επιλέγει ένα υποσύνολο των δεδομένων για τις επόμενες στατιστικές αναλύσεις, χρησιμοποιώντας ως κριτήριο κάποια μεταβλητή. Προσοχή: Η εντολή παραμένει ενεργή μέχρι να την ακυρώσουμε!

Από τη γραμμή εντολών επιλέγουμε διαδοχικά «DATA» → «SELECT CASES…». Στο πλαίσιο διαλόγου που εμφανίζεται, από το κουμπί «If condition is satisfied» επιλέγουμε τη μεταβλητή που θα χρησιμοποιήσουμε ως κριτήριο και τη μεταφέρουμε στο κουτί όπου μπορούμε να πραγματοποιήσουμε αλγεβρικές πράξεις. Για παράδειγμα, πληκτρολογώντας «gender = 1» επιλέγουμε μόνο τους άνδρες για τις επόμενες στατιστικές αναλύσεις (εάν η κωδικοποίηση της μεταβλητής «gender» ήταν 1 = «άνδρες» και 2 = «γυναίκες»). Μπορούμε επίσης να χρησιμοποιήσουμε τα σύμβολα < και > για να ορίσουμε μία συνθήκη επιλογής των δεδομένων. Ας πούμε ότι θέλουμε να χρησιμοποιήσουμε σε μια ανάλυση μόνον όσα άτομα έχουν ηλικία («age») άνω των 15 ετών. Πληκτρολογούμε «age > 15» και ενεργοποιούμε την εντολή επιλέγοντας «Continue» και «OK».

Όταν θελήσουμε να συμπεριλάβουμε ξανά το συνολικό δείγμα στις στατιστικές αναλύσεις, θα επαναλάβουμε τα βήματα «DATA» → «SELECT CASES…» και στο πλαίσιο διαλόγου θα επιλέξουμε «All cases».

6. ΣΤΑΤΙΣΤΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

Μεγάλο μέρος της ισχύος του SPSS βρίσκεται κρυμμένο στην εντολή «Analyze»[2]. Από εδώ μπορούμε να ενεργοποιήσουμε μια σειρά από στατιστικές αναλύσεις, άλλες στοιχειωδώς απλές και άλλες τερατωδώς πολύπλοκες. Η επιλογή μιας στατιστικής ανάλυσης από το μενού της εντολής «Analyze» ανοίγει ένα πλαίσιο διαλόγου στο οποίο πρέπει να συμπληρώσουμε ορισμένα στοιχεία για την εκτέλεση της εντολής.

Το βασικότερο στοιχείο στο πλαίσιο διαλόγου μιας στατιστικής ανάλυσης είναι να επιλέξουμε τις μεταβλητές που θέλουμε να χρησιμοποιήσουμε για τη συγκεκριμένη ανάλυση και να τις μεταφέρουμε στο αντίστοιχο κουτί. Η διαδικασία αυτή γίνεται ως εξής: Στα αριστερά του πλαισίου διαλόγου υπάρχει ένας κατάλογος με όλες τις διαθέσιμες μεταβλητές της έρευνας. Επιλέγοντας μία μεταβλητή και πατώντας το κουμπί

[2] Σε παλαιότερες εκδόσεις του SPSS, η εντολή «Analyze» εμφανιζόταν ως «Statistics».

➔ μεταφέρουμε τη μεταβλητή σε διπλανό κουτί, για την ανάλυση. Σε μερικές περιπτώσεις μπορεί να χρειαστεί να γνωρίζουμε εκ των προτέρων ποιες μεταβλητές θα χρησιμοποιήσουμε ως ανεξάρτητες και ποιες ως εξαρτημένες, για να τις τοποθετήσουμε στην αντίστοιχη θέση.

Στο πλαίσιο διαλόγου μιας στατιστικής ανάλυσης μπορούμε προαιρετικά να επιλέξουμε και ορισμένες ρυθμίσεις που βρίσκονται κρυμμένες πίσω από κουμπιά, όπως «Options», «Cells», «Statistics», κ.ά. Είναι λογικό ότι αυτές οι σημειώσεις δεν μπορούν να καλύψουν όλες τις πιθανές λειτουργίες των εντολών. Θα παρουσιαστούν μόνο εκείνες που χρησιμοποιούνται στις περισσότερες περιπτώσεις. Ο χρήστης παροτρύνεται να καταφεύγει στη βοήθεια του προγράμματος (εντολή «Help» από τη γραμμή εντολών) και, κυρίως, στην ενδεικτική βιβλιογραφία για διευκρινήσεις σε θέματα που δεν παρουσιάζονται εδώ αναλυτικά.

Όταν ολοκληρώσουμε τη συμπλήρωση του πλαισίου διαλόγου επιλέγουμε το κουμπί «OK» για την εκτέλεση της εντολής. Η κίνηση αυτή ενεργοποιεί τον επεξεργαστή του SPSS και μας μεταφέρει στο παράθυρο προβολής των αποτελεσμάτων (Viewer) όπου μετά από μερικά δευτερόλεπτα εμφανίζονται τα πολυπόθητα αποτελέσματα. Το παράθυρο των αποτελεσμάτων χωρίζεται σε δύο μέρη: στο αριστερό μέρος υπάρχουν τα περιεχόμενα των αναλύσεων συνοπτικά, ενώ στο δεξί μέρος υπάρχουν οι πίνακες των αναλύσεων σε πλήρη ανάπτυξη. Για γρήγορη μετακίνηση μέσα στο αρχείο των αποτελεσμάτων, μπορούμε να επιλέξουμε κατευθείαν την ανάλυση που θέλουμε να δούμε από το αριστερό τμήμα του παραθύρου. Εκεί μπορούμε, επίσης, να επιλέξουμε με ένα «κλικ» όλους τους πίνακες μιας ανάλυσης και να τους αντιγράψουμε, μετακινήσουμε ή διαγράψουμε, ανοίγοντας το μενού «Edit» της γραμμής εντολών.

Παρακάτω εξετάζονται ορισμένες από τις πιο βασικές στατιστικές αναλύσεις που μπορεί να εκτελέσει το SPSS, οι οποίες συνοδεύονται από εφαρμοσμένο παράδειγμα και υποδείξεις για τον τρόπο μεταφοράς των αποτελεσμάτων στο κείμενο ενός ερευνητικού άρθρου και την κατασκευή πινάκων και σχημάτων.

6.1. Περιγραφικοί στατιστικοί δείκτες μονομεταβλητών

FREQUENCIES. Υπολογίζει την κατανομή συχνότητας των τιμών μίας μεταβλητής. Η εντολή του SPSS είναι «ANALYZE» → «DESCRIPTIVE STATISTICS» → «FREQUENCIES…».

Στο παρακάτω παράδειγμα παρουσιάζεται η κατανομή συχνότητας για τις μεταβλητές «gender» («φύλο») και «educmo» («εκπαίδευση μητέρας»):

Frequencies

Statistics

		GENDER φύλο	EDUCMO εκπαίδευση μητέρας
N	Valid	69	64
	Missing	0	5

Frequency Table

GENDER φύλο

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 αγόρι	42	60,9	60,9	60,9
	2 κορίτσι	27	39,1	39,1	100,0
	Total	69	100,0	100,0	

EDUCMO εκπαίδευση μητέρας

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Δημοτικό	1	1,4	1,6	1,6
	2 Γυμνάσιο	4	5,8	6,3	7,8
	3 Λύκειο	26	37,7	40,6	48,4
	4 ΑΕΙ/ΤΕΙ	33	47,8	51,6	100,0
	Total	64	92,8	100,0	
Missing	System	5	7,2		
Total		69	100,0		

Ο πίνακας «Statistics» περιέχει τον αριθμό των έγκυρων («valid») και των ελλιπών («missing») στοιχείων για κάθε μεταβλητή. Στους πίνακες της ενότητας «Frequency Table», η στήλη «Frequency» περιέχει τις απόλυτες επιμέρους συχνότητες, η στήλη «Percent» περιέχει τις σχετικές επιμέρους συχνότητες (ποσοστά %) και η στήλη «Cum Percent» περιέχει τις αθροιστικές σχετικές συχνότητες. Η στήλη «Valid Percent» περιέχει τις σχετικές συχνότητες, όπως προκύπτουν αφού αποκλειστούν τα ελλιπή δεδομένα (missing data). Συνήθως οι πληροφορίες αυτές καταγράφονται σε ρέοντα λόγο, χωρίς να απαιτείται η προσθήκη πίνακα, περίπου ως εξής:

Ως προς το φύλο, το δείγμα αποτελείται από 42 (60,9%) αγόρια και 27 (39,1%) κορίτσια. Ως προς το επίπεδο εκπαίδευσης της μητέρας, 1 (1,4%) άτομο έχει μητέρα απόφοιτη Δημοτικού, 4 (5,8%) έχουν μητέρα απόφοιτη Γυμνασίου, 26 (37,7%) έχουν μητέρα απόφοιτη Λυκείου και 33 (47,8%) άτομα έχουν μητέρα απόφοιτη ΑΕΙ/ΤΕΙ. Δεν υπάρχουν στοιχεία για το επίπεδο εκπαίδευσης της μητέρας 5 (7,2%) ατόμων.

6.2. Σύγκριση ποσοστιαίων αναλογιών κατηγορικής διμεταβλητής

CROSSTABS. Υπολογίζει την κατανομή συχνότητας μιας διμεταβλητής. Χρησιμοποιείται με κατηγορικά ή ποιοτικά δεδομένα (κατηγορικές διμεταβλητές). Η εντολή στο SPSS είναι: «ANALYZE» → « DESCRIPTIVE STATISTICS» → «CROSSTABS…».

Στο παρακάτω παράδειγμα ελέγχουμε αν διαφέρει η συχνότητα εμφάνισης της συνήθειας του καπνίσματος (μεταβλητή «smoke») μεταξύ μαθητών Α' και Β' Λυκείου (μεταβλητή «grade»). Πέραν των προεπιλογών της εντολής, ζητήσαμε επιπλέον τα ποσοστά για τις σειρές και τις στήλες (κουμπί «Cells…»), καθώς και τον υπολογισμό του στατιστικού κριτηρίου χ^2 (κουμπί «Statistics»). Το αποτέλεσμα έχει ως εξής:

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
SMOKE κάπνισμα * GRADE τάξη φοίτησης	69	100,0%	0	,0%	69	100,0%

SMOKE κάπνισμα * GRADE τάξη φοίτησης Crosstabulation

			GRADE τάξη φοίτησης		
			1 Α Λυκ.	2 Β Λυκ.	Total
SMOKE κάπνισμα	0 όχι	Count	17	14	31
		% within SMOKE κάπνισμα	54,8%	45,2%	100,0%
		% within GRADE τάξη φοίτησης	63,0%	33,3%	44,9%
	1 ναι	Count	10	28	38
		% within SMOKE κάπνισμα	26,3%	73,7%	100,0%
		% within GRADE τάξη φοίτησης	37,0%	66,7%	55,1%
Total		Count	27	42	69
		% within SMOKE κάπνισμα	39,1%	60,9%	100,0%
		% within GRADE τάξη φοίτησης	100,0%	100,0%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	5,831[b]	1	,016		
Continuity Correction[a]	4,695	1	,030		
Likelihood Ratio	5,882	1	,015		
Fisher's Exact Test				,025	,015
Linear-by-Linear Association	5,747	1	,017		
N of Valid Cases	69				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 12,13.

Ο πρώτος πίνακας, «Case processing summary», δείχνει τα έγκυρα και τα ελλιπή στοιχεία. Ο δεύτερος πίνακας δείχνει την κατανομή συχνότητας της διμεταβλητής. Σε κάθε φατνίο του πίνακα αυτού, η πρώτη σειρά δηλώνει την απόλυτη συνδυαστική συχνότητα, η δεύτερη σειρά δηλώνει τη σχετική συνδυαστική συχνότητα (%) για τις σειρές και η τρίτη σειρά τη σχετική συνδυαστική συχνότητα (%) για τις στήλες. Σε ξεχωριστό πίνακα παρουσιάζονται τα αποτελέσματα του στατιστικού ελέγχου με το κριτήριο χ^2. Στον πίνακα αυτό, η τιμή χ^2 εμφανίζεται στη δεύτερη στήλη («Value») της πρώτης σειράς («Pearson chi-square»). Στην ίδια σειρά, σε διπλανές στήλες υπάρχουν οι βαθμοί ελευθερίας («d.f.») και το επίπεδο στατιστικής σημαντικότητας («Asymp. Sig. (2-sided)»). Υπενθυμίζεται ότι τιμές ίσες ή μικρότερες από ,05 θεωρούνται στατιστικώς σημαντικές ($p < 0,05$). Αυτό σημαίνει ότι οι διαφορές που παρατηρούνται στις συχνότητες των φατνίων του πίνακα είναι συστηματικές και δεν οφείλονται στη διακύμανση των τυχαίων δειγμάτων. Θα καταγράφαμε τα ευρήματα αυτά ως εξής:

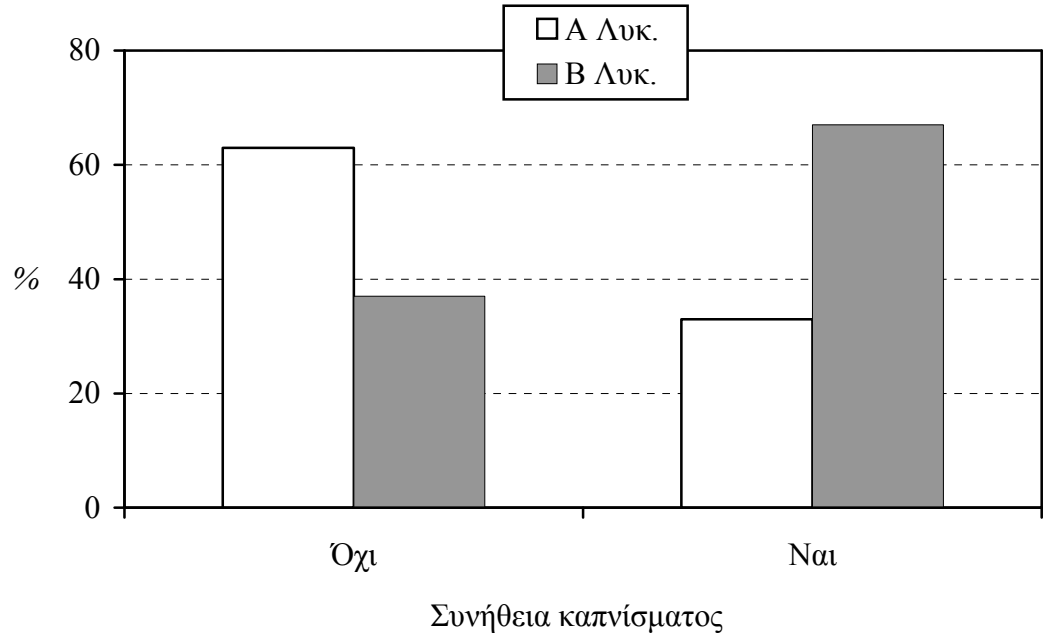
Όπως έδειξε το κριτήριο χ^2, οι απαντήσεις των μαθητών στο ερώτημα σχετικά με τη συνήθεια του καπνίσματος διαφέρουν ανάλογα με την τάξη φοίτησης, $\chi^2(1, n = 69) = 5,83$, $p = 0,016$. Ειδικότερα, μόνο το 37% (10 μαθητές) της Α' Λυκείου δήλωσαν ότι καπνίζουν, ενώ το αντίστοιχο ποσοστό για τη Β' Λυκείου είναι 66,7% (28 μαθητές) (βλ. Πίν. 1 και Σχ. 1).

Πίνακας 1.

Κατανομή (απόλυτες και σχετικές συχνότητες) της συνήθειας καπνίσματος ως προς την τάξη φοίτησης

Κάπνισμα	Τάξη φοίτησης					
	Α Λυκείου		Β Λυκείου		Σύνολο	
	f	%	*f*	%	*f*	%
Όχι	17	63,0	14	33,3	31	44,9
Ναι	10	37,0	28	66,7	38	55,1
Σύνολο	27	39,1	42	60,9	69	100

Σημείωση. Οι σχετικές συνδυαστικές συχνότητες έχουν υπολογιστεί για τις στήλες (τάξη φοίτησης).



Σχήμα 1. Κατανομή (%) της συνήθειας καπνίσματος ως προς την τάξη φοίτησης.

Για να είναι έγκυρος ο έλεγχος με το κριτήριο χ^2, πρέπει τουλάχιστον το 75% των φατνίων του πίνακα να έχουν συνδυαστική συχνότητα μεγαλύτερη από 5. Αν δεν συμβαίνει αυτό, τότε το SPSS υπολογίζει μεν την τιμή χ^2 αλλά συγχρόνως προειδοποιεί για το ποσοστό των φατνίων με συνδυαστική συχνότητα < 5. Στην περίπτωση αυτή, τα αποτελέσματα του ελέγχου χ^2 πρέπει να ερμηνευθούν με επιφύλαξη.

***

T-TEST. Χρησιμοποιείται για να συγκρίνουμε τη στατιστική σημαντικότητα των διαφορών των μέσων όρων: (α) της ίδιας μεταβλητής σε δύο διαφορετικές ομάδες ατόμων (ανεξάρτητα δείγματα), και (β) δύο διαφορετικών μεταβλητών σε μια, την ίδια ομάδα ατόμων (εξαρτημένα δείγματα). Ανάλογα, λοιπόν, με το είδος της σύγκρισης, υπάρχουν δύο είδη *t*-test:

6.3. Σύγκριση δύο μέσων όρων (ανεξάρτητα δείγματα)

INDEPENDENT SAMPLES T-TEST. Συγκρίνει τους μέσους όρους μίας αριθμητικής μεταβλητής μεταξύ **δύο** ομάδων (πρόκειται δηλ. για μικτή διμεταβλητή). Η εντολή στο SPSS είναι: «ANALYZE» → «COMPARE MEANS» → «INDEPENDENT-SAMPLES T TEST…»

Το αποτέλεσμα της ανάλυσης καταλήγει σε δύο πίνακες. Στον πρώτο, «Group statistics», εμφανίζονται περιγραφικοί στατιστικοί δείκτες: το μέγεθος («N») κάθε ομάδας, ο μέσος όρος («Mean»), η τυπική απόκλιση («Std. Deviation») και το τυπικό σφάλμα του μέσου όρου («Std. Error Mean»). Στο δεύτερο πίνακα παρουσιάζεται το *t*-test, ως εξής: υπολογίζονται δύο τιμές, μία *t*-τιμή για την περίπτωση που τα δείγματα είναι ομοιογενή («Equal variances assumed») και μία *t*-τιμή για την περίπτωση που τα δείγματα είναι ανομοιογενή («Equal variances not assumed»).

Ο έλεγχος της ομοιογένειας της διασποράς μεταξύ των συγκρινόμενων δειγμάτων γίνεται με το Levene test. Αν η *F*-τιμή του Levene test είναι στατιστικώς σημαντική («Sig.» < ,050), αυτό σημαίνει ότι τα δείγματα είναι ανομοιογενή.

Στη συνέχεια, εκτός από την *t*-τιμή, δίνονται οι βαθμοί ελευθερίας («df»), το επίπεδο στατιστικής σημαντικότητας («Sig.»), η διαφορά μεταξύ των μέσων όρων («Mean Difference»), το τυπικό σφάλμα της διαφοράς των μέσων όρων («Std. Error Difference») και τα όρια του διαστήματος εμπιστοσύνης της διαφοράς των μέσων όρων («95% Confidence Interval of the Difference»).

T-Test

Group Statistics

	GRADE τάξη φοίτησης	N	Mean	Std. Deviation	Std. Error Mean
RISK αριθμός ριψοκίνδυνων συμπεριφορών	1 Α Λυκ.	27	3,22	1,28	,25
	2 Β Λυκ.	42	4,55	1,86	,29

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
RISK αριθμός ριψοκίνδυνων συμπεριφορών	Equal variances assumed	8,282	,005	-3,233	67	,002	-1,33	,41	-2,14	-,51
	Equal variances not assumed			-3,499	66,645	,001	-1,33	,38	-2,08	-,57

Στο παράδειγμά μας υπολογίσαμε τους μέσους όρους του αριθμού ριψοκίνδυνων συμπεριφορών («risk») ως προς την τάξη φοίτησης («grade») εφήβων μαθητών. Θα συνοψίζαμε τα αποτελέσματα αυτά σε κείμενο ως εξής:

Ο έλεγχος των μέσων όρων με το κριτήριο *t* για ανεξάρτητα δείγματα έδειξε ότι ο αριθμός ριψοκίνδυνων συμπεριφορών συνδέεται σημαντικά με την τάξη φοίτησης. Ειδικότερα, οι μαθητές της Β' Λυκείου ($M = 4,55$, $SD = 1,86$) αναφέρουν μεγαλύτερο αριθμό ριψοκίνδυνων συμπεριφορών από ό,τι οι μαθητές της Α' Λυκείου ($M = 3,22$, $SD = 1,28$), $t(66,64) = 3,50$, $p = 0,001$.

6.4. Σύγκριση δύο μέσων όρων (εξαρτημένα δείγματα)

PAIRED SAMPLES T-TEST. Χρησιμοποιείται για να συγκρίνουμε αν ο μέσος όρος μιας αριθμητικής μεταβλητής διαφέρει από τον μέσο όρο μιας άλλης αριθμητικής μεταβλητής (η οποία έχει μετρηθεί με την ίδια μονάδα μέτρησης) σε μία, την ίδια ομάδα ατόμων. Πρόκειται δηλ. για αριθμητική διμεταβλητή σε εξαρτημένα δείγματα. Η σχετική εντολή στο SPSS είναι: «ANALYZE» → «COMPARE MEANS» → «PAIRED-SAMPLES T TEST…»

Το αποτέλεσμα δίνει τρεις πίνακες. Στον πρώτο πίνακα δίνονται περιγραφικοί στατιστικοί δείκτες των δύο μεταβλητών: μέσοι όροι («Mean»), μέγεθος δείγματος («N»), τυπικές αποκλίσεις («Std. Deviation»), τυπικό σφάλμα των μέσων όρων («Std. Error Mean»). Στο δεύτερο πίνακα δίνεται ο δείκτης συνάφειας Pearson *r* («Correlation») μεταξύ των δύο μεταβλητών και το επίπεδο στατιστικής σημαντικότητάς του («Sig.»). Στον τρίτο πίνακα δίνονται τα στοιχεία από τη σύγκριση των μέσων όρων («Paired differences»), η τιμή *t*, οι βαθμοί ελευθερίας («df») και το επίπεδο στατιστικής σημαντικότητας («Sig.»).

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	RISK2 αριθμός βλαπτικών συμπεριφορών (μετά)	4,03	69	1,77	,21
	RISK1 αριθμός βλαπτικών συμπεριφορών (πριν)	4,90	69	1,49	,18

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	RISK2 αριθμός βλαπτικών συμπεριφορών (μετά) & RISK1 αριθμός βλαπτικών συμπεριφορών (πριν)	69	,492	,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	RISK2 - RISK1	-,87	1,66	,20	-1,27	-,47	-4,346	68	,000

Στο παραπάνω υποθετικό παράδειγμα συγκρίναμε τους μέσους όρους βλαπτικών συμπεριφορών που εκδήλωσε μια ομάδα 69 εφήβων πριν (μεταβλητή «risk1») και μετά (μεταβλητή «risk2») από ένα πρόγραμμα παρέμβασης. Τα αποτελέσματα συνοψίζονται ως εξής:

Όπως έδειξε το κριτήριο *t* για εξαρτημένα δείγματα, ο μέσος όρος αριθμού βλαπτικών συμπεριφορών ήταν μικρότερος μετά την παρέμβαση (M = 4,03, SD = 1,77), συγκριτικά με την αντίστοιχη τιμή πριν την παρέμβαση (M = 4,90, SD = 1,49): $t(68) = -4{,}35$, $p < 0{,}001$ (βλ. Πίν. 2).

Πίνακας 2.

Μέσοι όροι και τυπικές αποκλίσεις του αριθμού βλαπτικών συμπεριφορών των εφήβων πριν και μετά την εφαρμογή του προγράμματος παρέμβασης

	Πριν την παρέμβαση		Μετά την παρέμβαση		*t*
	M	*SD*	*M*	*SD*	(*df* = 68)
Αριθμός βλαπτικών συμπεριφορών	4,90	1,49	4,03	1,77	-4,35***

Σημείωση. *** $p < 0{,}001$.

6.5. Σύγκριση περισσότερων των δύο μέσων όρων

ANOVA. Η ανάλυση διακύμανσης (Analysis of Variance, ANOVA) χρησιμοποιείται για τον έλεγχο της στατιστικής σημαντικότητας των διαφορών των μέσων όρων περισσότερων από δύο ομάδων. Επιπλέον, με την ANOVA μπορούμε να ελέγξουμε την αλληλεπίδραση δύο ή περισσότερων ανεξάρτητων (κατηγορικών) μεταβλητών πάνω στην εξαρτημένη (αριθμητική) μεταβλητή. Δηλαδή, μπορεί να έχουμε μία μικτή διμεταβλητή, τριμεταβλητή ή ακόμα και πολυμεταβλητή.

Το SPSS παρέχει πολλές δυνατότητες σχεδιασμού και παραμετροποίησης μοντέλων ανάλυσης διακύμανσης, κυρίως μέσω της ανάλυσης General Linear Model, στο μενού εντολών: «ANALYZE» → «GENERAL LINEAR MODEL» → «UNIVARIATE…». Επειδή όμως οι προεπιλεγμένες ρυθμίσεις δίνουν στοιχειώδη εικόνα (στην ουσία, μόνο τον πίνακα της ανάλυσης διακύμανσης), συνήθως απαιτείται να ζητήσουμε την εμφάνιση επιπλέον πληροφοριών μέσα από το κουμπί «Options…», ως εξής:

(α) Στην ενότητα «Estimated Marginal Means», μεταφέρουμε τα περιεχόμενα του καταλόγου «Factor(s) and Factor Interactions» στον διπλανό κατάλογο «Display Means for:». Αν κάποια ανεξάρτητη μεταβλητή έχει περισσότερες από δύο τιμές-επίπεδα, τσεκάρουμε την επιλογή «Compare main effects» και από το μενού «Confidence interval adjustment» επιλέγουμε το post hoc κριτήριο του Bonferroni.

(β) Στην ενότητα «Display», τσεκάρουμε τις επιλογές «Estimates of effect size» και «Homogeneity tests».

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
ethngrp εθνική ομάδα	1	Αλβανοί	144
	2	Πόντιοι	96
adv αντιξοότητα	1	χαμηλή	98
	2	μέτρια	95
	3	υψηλή	47

Levene's Test of Equality of Error Variances [a]

Dependent Variable: prcdscr αντιλαμβανόμενη διάκριση

F	df1	df2	Sig.
1.412	5	234	.221

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+ethngrp+adv+ethngrp * adv

Tests of Between-Subjects Effects

Dependent Variable: prcdscr αντιλαμβανόμενη διάκριση

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	33.963[a]	5	6.793	10.057	.000	.177
Intercept	1405.695	1	1405.695	2081.154	.000	.899
ethngrp	10.395	1	10.395	15.391	.000	.062
adv	16.690	2	8.345	12.355	.000	.096
ethngrp * adv	4.900	2	2.450	3.627	.028	.030
Error	158.053	234	.675			
Total	1797.345	240				
Corrected Total	192.017	239				

a. R Squared = .177 (Adjusted R Squared = .159)

Estimated Marginal Means

1. Grand Mean

Dependent Variable: prcdscr αντιλαμβανόμενη διάκριση

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
2.605	.057	2.493	2.718

2. εθνική ομάδα

Estimates

Dependent Variable: prcdscr αντιλαμβανόμενη διάκριση

εθνική ομάδα	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 Αλβανοί	2.829	.076	2.680	2.978
2 Πόντιοι	2.381	.085	2.213	2.549

Pairwise Comparisons

Dependent Variable: prcdscr αντιλαμβανόμενη διάκριση

(I) εθνική ομάδα	(J) εθνική ομάδα	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1 Αλβανοί	2 Πόντιοι	.448*	.114	.000	.223	.673
2 Πόντιοι	1 Αλβανοί	-.448*	.114	.000	-.673	-.223

Based on estimated marginal means

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

a. Adjustment for multiple comparisons: Bonferroni.

Univariate Tests

Dependent Variable: prcdscr αντιλαμβανόμενη διάκριση

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Contrast	10.395	1	10.395	15.391	.000	.062
Error	158.053	234	.675			

The F tests the effect of εθνική ομάδα. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

3. αντιξοότητα

Estimates

Dependent Variable: prcdscr αντιλαμβανόμενη διάκριση

αντιξοότητα	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 χαμηλή	2.458	.086	2.289	2.627
2 μέτρια	2.317	.087	2.145	2.489
3 υψηλή	3.040	.120	2.804	3.276

Pairwise Comparisons

Dependent Variable: prcdscr αντιλαμβανόμενη διάκριση

(I) αντιξοότητα	(J) αντιξοότητα	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1 χαμηλή	2 μέτρια	.141	.122	.754	-.154	.436
	3 υψηλή	-.582*	.147	.000	-.937	-.227
2 μέτρια	1 χαμηλή	-.141	.122	.754	-.436	.154
	3 υψηλή	-.723*	.148	.000	-1.081	-.365
3 υψηλή	1 χαμηλή	.582*	.147	.000	.227	.937
	2 μέτρια	.723*	.148	.000	.365	1.081

Based on estimated marginal means

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

a. Adjustment for multiple comparisons: Bonferroni.

Univariate Tests

Dependent Variable: prcdscr αντιλαμβανόμενη διάκριση

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Contrast	16.690	2	8.345	12.355	.000	.096
Error	158.053	234	.675			

The F tests the effect of αντιξοότητα. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

4. εθνική ομάδα * αντιξοότητα

Dependent Variable: prcdscr αντιλαμβανόμενη διάκριση

				95% Confidence Interval	
εθνική ομάδα	αντιξοότητα	Mean	Std. Error	Lower Bound	Upper Bound
1 Αλβανοί	1 χαμηλή	2.762	.105	2.555	2.969
	2 μέτρια	2.697	.106	2.488	2.906
	3 υψηλή	3.028	.171	2.691	3.366
2 Πόντιοι	1 χαμηλή	2.154	.135	1.888	2.420
	2 μέτρια	1.938	.139	1.664	2.211
	3 υψηλή	3.052	.168	2.721	3.382

Στο παραπάνω παράδειγμα, ελέγξαμε τη στατιστική σημαντικότητα των διαφορών των μέσων όρων αντιλαμβανόμενης διάκρισης μεταναστών και παλιννοστούντων στην Ελλάδα (μεταβλητή: «prcdscr») ως προς τις μεταβλητές: εθνική ομάδα («ethngrp») και επίπεδο αντιξοότητας («adv»).

Ο πρώτος πίνακας, «Between-Subjects Factors», δείχνει τις τιμές-επίπεδα των ανεξάρτητων μεταβλητών.

Ο δεύτερος πίνακας («Levene's Test of Equality of Error Variances») εμφανίζει το αποτέλεσμα από τον έλεγχο ομοιογένειας της διασποράς με το test του Levene (το οποίο, παρεμπιπτόντως, πρέπει να είναι στατιστικώς *ασήμαντο* για να διασφαλιστεί η εγκυρότητα της ανάλυσης διακύμανσης).

Ο επόμενος πίνακας, «Tests of Between-Subjects Effects», συνοψίζει την ανάλυση διακύμανσης. Στις στήλες του πίνακα αυτού υπάρχουν τα αθροίσματα τετραγώνων των αποκλίσεων («Type III Sum of Squares»), οι βαθμοί ελευθερίας («df»), τα μέσα τετράγωνα των αποκλίσεων («Mean Square»), η *F*-τιμή, το επίπεδο στατιστικής σημαντικότητας («Sig.») και ο δείκτης η^2 («Partial Eta Squared»). Στις σειρές του πίνακα υπάρχουν οι κύριες επιδράσεις των ανεξάρτητων μεταβλητών («ethngrp», «adv»), η αλληλεπίδραση των ανεξάρτητων μεταβλητών («ethngrp * adv») και πληροφορίες για τη διασπορά εντός των ομάδων («Error»). Αγνοείστε τις σειρές «Corrected Model», «Intercept», «Total» και «Corrected Total». Υπενθυμίζεται ότι για να θεωρηθεί η επίδραση κάποιας πηγής διασποράς ως στατιστικώς σημαντική, θα πρέπει η τιμή στη στήλη «Sig.» να είναι μικρότερη από .050 (δηλ. $p < 0,05$).

Στην ενότητα «Estimated Marginal Means» υπάρχει ο γενικός μέσος όρος της εξαρτημένης μεταβλητής για το συνολικό δείγμα («1. Grand Mean»), οι μέσοι όροι των κύριων επιδράσεων («2. εθνική ομάδα», «3. αντιξοότητα») και της αλληλεπίδρασης («4. εθνική ομάδα * αντιξοότητα»). Για κάθε πηγή διασποράς, εκτός από το μέσο όρο («Mean») και το τυπικό σφάλμα του μέσου όρου («Std. Error»), δίνονται επιπλέον οι πολλαπλές συγκρίσεις με το κριτήριο του Bonferroni (βλ. πίνακες «Pairwise Comparisons»), τις οποίες ζητήσαμε από το κουμπί «Options». Ενδιαφέρον εδώ παρουσιάζει μόνο το αποτέλεσμα για την αντιξοότητα, όπου από την επισκόπηση της στήλης «Sig.» διαπιστώνουμε ότι στατιστικώς σημαντικές είναι οι διαφορές των μέσων όρων αντιλαμβανόμενης διάκρισης μεταξύ των ομάδων χαμηλής και υψηλής αντιξοό-

τητας (p < 0,001), μέτριας και υψηλής αντιξοότητας (p < 0,001), υψηλής και χαμηλής αντιξοότητας (p < 0,001, η σύγκριση αυτή επαναλαμβάνεται) και υψηλής και μέτριας αντιξοότητας (p < 0,001). Αγνοείστε τον πίνακα «Univariate Tests» που ακολουθεί.

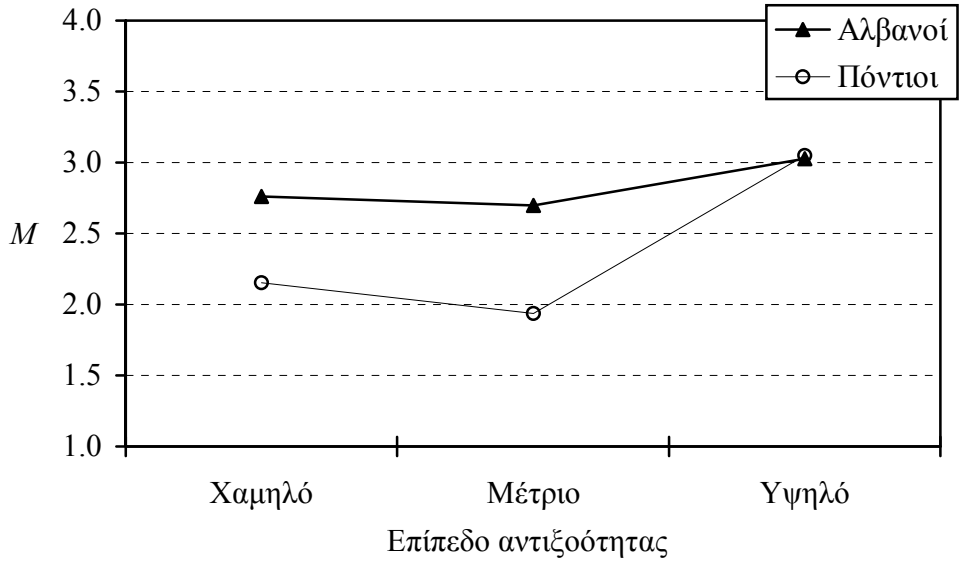
Θα συνοψίζαμε τα αποτελέσματα αυτά στην αντίστοιχη ενότητα ενός ερευνητικού άρθρου με κείμενο, πίνακα και σχήμα, ως εξής:

Πίνακας 3.

Μέσοι όροι αντιλαμβανόμενης διάκρισης των μεταναστών ως προς το επίπεδο αντιξοότητας

	Επίπεδο αντιξοότητας			
	Χαμηλό	Μέτριο	Υψηλό	F
	M	M	M	(df = 2, 234)
Αντιλαμβανόμενη διάκριση	$2{,}46_{\alpha}$	$2{,}32_{\alpha}$	$3{,}04_{\beta}$	3,63*

Σημείωση. * p < 0,05. Μέσοι όροι που μοιράζονται κοινό δείκτη *δεν* διαφέρουν σημαντικά μεταξύ τους σύμφωνα με το post hoc κριτήριο του Scheffé για p < 0,05. Η κλίμακα βαθμολόγησης κυμαίνεται από 1 = «καθόλου» μέχρι 5 = «πάρα πολύ».



Σχήμα 2. Αλληλεπίδραση του επιπέδου αντιξοότητας και της εθνικής ομάδας πάνω στους μέσους όρους αντιλαμβανόμενης διάκρισης.

Σχεδιάστηκε ένα μοντέλο διπαραγοντικής ανάλυσης διακύμανσης 2 (εθνική ομάδα: Αλβανοί, Πόντιοι) Χ 3 (επίπεδο αντιξοότητας: χαμηλό, μέτριο, υψηλό) με εξαρτημένη μεταβλητή το βαθμό αντιλαμβανόμενης διάκρισης.

Ο έλεγχος της ομοιογένειας της διασποράς με το κριτήριο του Levene κατέληξε σε στατιστικώς ασήμαντο αποτέλεσμα, $F(5, 234) = 1,41$, $p = 0,221$. Οι κύριες επιδράσεις των ανεξάρτητων παραγόντων ήταν στατιστικώς σημαντικές. Όσον αφορά την εθνική ομάδα, οι Αλβανοί ($M = 2,83$, $SE = 0,08$) σημείωσαν υψηλότερο μέσο όρο αντιλαμβανόμενης διάκρισης από ό,τι οι Πόντιοι ($M = 2,38$, $SE = 0,08$), $F(1, 234) = 15,39$, $p < 0,001$. Επιπλέον, το επίπεδο αντιξοότητας διαφοροποίησε τους μέσους όρους αντιλαμβανόμενης διάκρισης, $F(2, 234) = 12,36$, $p < 0,001$. Από τις πολλαπλές συγκρίσεις με το post hoc κριτήριο του Bonferroni προέκυψε ότι οι μετανάστες που αντιμετωπίζουν υψηλή αντιξοότητα ($M = 3,04$, $SE = 0,12$) είχαν υψηλότερο μέσο όρο από ό,τι οι μετανάστες με μέτρια ($M = 2,32$, $SE = 0,09$) και χαμηλή ($M = 2,46$, $SE = 0,09$) αντιξοότητα, οι οποίοι δεν διέφεραν μεταξύ τους (βλ. Πίν. 3).

Στατιστικώς σημαντική όμως ήταν και η αλληλεπίδραση των ανεξάρτητων μεταβλητών, $F(2, 234) = 3,63$, $p = 0,028$. Όπως φαίνεται στο Σχ. 2, οι Αλβανοί με χαμηλό και μέτριο επίπεδο αντιξοότητας ανέφεραν υψηλότερο μέσο όρο αντιλαμβανόμενης διάκρισης από ό,τι οι Πόντιοι αντίστοιχου επιπέδου αντιξοότητας. Ενώ όταν η αντιξοότητα είναι υψηλή, Αλβανοί και Πόντιοι δεν διαφέρουν μεταξύ τους και, μάλιστα, σημειώνουν τον υψηλότερο μέσο όρο αντιλαμβανόμενης διάκρισης από όλες τις συνδυαστικές ομάδες.

6.6. Συνάφεια αριθμητικών διμεταβλητών

CORRELATION. Η εντολή αυτή υπολογίζει τη συνάφεια μεταξύ αριθμητικών διμεταβλητών χρησιμοποιώντας τον τύπο του δείκτη συνάφειας Pearson *r* ως προεπιλογή. Δίνεται, επίσης, το επίπεδο στατιστικής σημαντικότητας του δείκτη και το μέγεθος του δείγματος. Υπενθυμίζεται ότι ο δείκτης Pearson *r* είναι έγκυρος μόνο σε περίπτωση που τηρούνται συγκεκριμένες προϋποθέσεις για την εφαρμογή του: αριθμητικά δεδομένα, ευθύγραμμη συνάφεια, ισοδιαστημική κλίμακα μέτρησης. Σε περίπτωση που

δεν πληρούνται οι παραπάνω προϋποθέσεις, μπορεί κανείς να επιλέξει τον απαραμετρικό δείκτη συνάφειας Spearman *rho*. Η εντολή στο SPSS είναι: «ANALYZE» → «CORRELATE» → «BIVARIATE…».

Στο παρακάτω παράδειγμα εξετάζουμε τη συνάφεια ανάμεσα σε τέσσερις κλίμακες αυτοαντίληψης της ικανότητας (μεταβλητές «h_sch», «h_soc», «h_athl», «h_self»). Η πρώτη σειρά σε κάθε φατνίο του πίνακα δείχνει το δείκτη Pearson r. Στη δεύτερη σειρά δίνεται το επίπεδο στατιστικής σημαντικότητας του δείκτη («Sig.») και στην τρίτη σειρά το μέγεθος του δείγματος («N»). Παρατηρείστε ότι στη διαγώνιο που ξεκινά από το άνω αριστερό τμήμα του πίνακα και καταλήγει στο κάτω δεξί τμήμα του, δίνεται η συνάφεια κάθε μεταβλητής με τον εαυτό της (αυτοσυνάφεια), η οποία, φυσικά, είναι πάντα 1,00. Επίσης, ο πίνακας είναι συμμετρικός ως προς τη διαγώνιο.

Ως προς την ερμηνεία, υπενθυμίζεται ότι ο δείκτης Pearson r δείχνει μόνο τον τρόπο που συμμεταβάλλονται οι μεταβλητές, χωρίς να δίνει πληροφορίες για τη φύση της συνάφειας. Επομένως, κάθε στατιστικώς σημαντικός δείκτης δεν συνεπάγεται την ύπαρξη αιτιώδους σχέσης μεταξύ των δύο μεταβλητών. Αυτό εξαρτάται από τον μεθοδολογικό σχεδιασμό της έρευνας.

Correlations

		H_SCH	H_SOC	H_ATHL	H_SELF
H_SCH σχολική ικανότητα	Pearson Correlation	1,000	,247*	,342**	,567**
	Sig. (2-tailed)	,	,041	,004	,000
	N	69	69	69	69
H_SOC σχέσεις με συνομηλίκους	Pearson Correlation	,247*	1,000	,045	,378**
	Sig. (2-tailed)	,041	,	,712	,001
	N	69	69	69	69
H_ATHL αθλητική ικανότητα	Pearson Correlation	,342**	,045	1,000	,294*
	Sig. (2-tailed)	,004	,712	,	,014
	N	69	69	69	69
H_SELF σφαιρική αυτοαξία	Pearson Correlation	,567**	,378**	,294*	1,000
	Sig. (2-tailed)	,000	,001	,014	,
	N	69	69	69	69

*. Correlation is significant at the 0.05 level (2-tailed).

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

Υπολογίστηκαν οι δείκτες συνάφειας *r* του Pearson ανάμεσα στις κλίμακες αυτοαντίληψης της ικανότητας. Όλοι οι δείκτες ήταν θετικής κατεύθυνσης και στατιστικώς σημαντικοί (με την εξαίρεση της συνάφειας ανάμεσα στις «σχέσεις με συνομηλίκους» και την «αθλητική ικανότητα», η οποία ήταν στατιστικώς ασήμαντη). Το μέγεθος των δεικτών κυμάνθηκε από 0,25 («σχέσεις με συνομηλίκους» Χ «σχολική ικανότητα») μέχρι 0,57 («σφαιρική αυτοαξία» Χ «σχολική ικανότητα»). Δηλαδή, όσο πιο θετικά αξιολογείται από τον έφηβο ένας τομέας αυτοαντίληψης της ικανότητας, τόσο πιο θετικά τείνουν να αξιολογούνται και άλλοι τομείς της αυτοαντίληψης, αν και το μέγεθος των συσχετίσεων αυτών κυμαίνεται από χαμηλό έως μέτριο (βλ. Πίν. 4).

Πίνακας 4.

Συνάφεια (Pearson r) μεταξύ των τομέων αυτοεκτίμησης

Τομείς αυτοεκτίμησης	1.	2.	3.	4.
1. Σχολική ικανότητα	1,00			
2. Σχέσεις με συνομηλίκους	0,25*	1,00		
3. Αθλητική ικανότητα	0,34**	0,04	1,00	
4. Σφαιρική αυτοαξία	0,57***	0,38***	0,29*	1,00

Σημείωση. * $p < 0,05$. ** $p < 0,01$. *** $p < 0,001$.

6.7. Στατιστική πρόβλεψη

REGRESSION. Η ανάλυση πολλαπλής παλινδρόμησης (multiple regression) χρησιμοποιείται για να προβλέψουμε στατιστικώς τις τιμές μιας εξαρτημένης μεταβλητής-κριτηρίου από μία ή περισσότερες ανεξάρτητες μεταβλητές πρόβλεψης. Αποτελεί προέκταση του απλού δείκτη συνάφειας διμεταβλητών εφόσον υπολογίζει τον δείκτη πολλαπλής συνάφειας, δηλ. το βαθμό συσχέτισης ανάμεσα στην εξαρτημένη μεταβλητή και πολλές ανεξάρτητες μεταβλητές συγχρόνως. Σε αντίθεση με τον υπολογισμό της συνάφειας όμως, η χρήση της ανάλυσης παλινδρόμησης υπονοεί την ύπαρξη αιτιώδους σχέσης ανάμεσα στην εξαρτημένη μεταβλητή και τις ανεξάρτητες. Όσο μεγαλύτερη η συνάφεια κάθε ανεξάρτητης μεταβλητής με την εξαρτημένη και όσο μικρότερη η συνάφεια των ανεξάρτητων μεταβλητών μεταξύ τους, τόσο αυξάνεται η ακρίβεια πρόβλεψης της εξαρτημένης μεταβλητής.

Ασφαλώς η ανάλυση παλινδρόμησης είναι περισσότερο πολύπλοκη τεχνική από όσο παρουσιάζεται εδώ και γι' αυτό απαιτείται προσεκτικός έλεγχος των προϋποθέσεων για την εφαρμογή της. Υπενθυμίζεται ότι, αφού ουσιαστικά πρόκειται για προέκταση του δείκτη συνάφειας Pearson r, πρέπει τόσο η εξαρτημένη όσο και οι ανεξάρτητες μεταβλητές να είναι αριθμητικές, η συνάφεια να είναι ευθύγραμμη και η κλίμακα μέτρησης ίσων διαστημάτων.

Η εντολή στο SPSS είναι: «ANALYZE» → «REGRESSION» → «LINEAR...». Στο πλαίσιο διαλόγου, αφού συμπληρώσουμε την εξαρτημένη («Dependent»)και τις ανεξάρτητες («Independent») μεταβλητές, μπορούμε επίσης να επιλέξουμε τη μέθοδο («Method»). Προεπιλεγμένη μέθοδος είναι η «Enter», η οποία υπολογίζει την επίδραση όλων των ανεξάρτητων μεταβλητών ταυτόχρονα πάνω στην εξαρτημένη. Ιδιαίτερα χρήσιμη όμως είναι και η βηματική μέθοδος «Stepwise», η οποία ιεραρχεί τις ανεξάρτητες μεταβλητές ανάλογα με τη βελτίωση που επιφέρουν στην ικανότητα πρόβλεψης της εξαρτημένης.

Στο παρακάτω παράδειγμα, επιχειρούμε να προβλέψουμε τον αριθμό ριψοκίνδυνων συμπεριφορών («risk») από τις ανεξάρτητες μεταβλητές: επίγνωση του κινδύνου («perdang»), επίγνωση των επιπτώσεων («percons»), αντιλαμβανόμενη ικανότητα ελέγχου («control») και επιρροή των συνομηλίκων («peerinfl»). Χρησιμοποιήσαμε τη μέθοδο «enter» για λόγους απλότητας του παραδείγματος. Τα αποτελέσματα δίνονται σε τέσσερις πίνακες. Ο πρώτος πίνακας, «Variables Entered/Removed», δείχνει τις ανεξάρτητες μεταβλητές και τη μέθοδο που επιλέχθηκε. Ο δεύτερος πίνακας, «Model Summary» δίνει το δείκτη πολλαπλής συνάφειας («R»), τον συντελεστή προσδιορισμού R^2 για το δείγμα («R Square», δηλ. το ποσοστό της διασποράς των

τιμών της εξαρτημένης μεταβλητής που εξηγείται από την επίδραση των ανεξάρτητων), την εκτίμηση του συντελεστή R^2 για τον πληθυσμό («Adjusted R Square») και το τυπικό σφάλμα της εκτίμησης αυτής («Std. Error of the Estimate»). Ο τρίτος πίνακας («ANOVA») δείχνει το αποτέλεσμα της ανάλυσης διακύμανσης, το οποίο δηλώνει εάν η κλίση της γραμμής παλινδρόμησης είναι σημαντικά διαφορετική του μηδενός. Τέλος, ο πίνακας «Coefficients» παρουσιάζει τους μη προσαρμοσμένους («B») και τους προσαρμοσμένους («Beta») συντελεστές παλινδρόμησης για καθεμιά από τις ανεξάρτητες μεταβλητές ξεχωριστά. Από την επισκόπηση της στήλης «Sig.» μπορούμε να εντοπίσουμε ποιοι συντελεστές είναι στατιστικώς σημαντικοί, δηλαδή ποιες ανεξάρτητες μεταβλητές συμβάλλουν σημαντικά στην πρόβλεψη της εξαρτημένης.

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	PEERINFL επιρροή συνομηλίκων, PERDANG επίγνωση κινδύνου, CONTROL ικανότητα ελέγχου, PERCONS επίγνωση επιπτώσεων[a]	,	Enter

a. All requested variables entered.

b. Dependent Variable: RISK αριθμός ριψοκίνδυνων συμπεριφορών

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,639[a]	,409	,372	1,41

a. Predictors: (Constant), PEERINFL επιρροή συνομηλίκων, PERDANG επίγνωση κινδύνου, CONTROL ικανότητα ελέγχου, PERCONS επίγνωση επιπτώσεων

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	87,422	4	21,856	11,056	,000[a]
	Residual	126,520	64	1,977		
	Total	213,942	68			

a. Predictors: (Constant), PEERINFL επιρροή συνομηλίκων, PERDANG επίγνωση κινδύνου, CONTROL ικανότητα ελέγχου, PERCONS επίγνωση επιπτώσεων

b. Dependent Variable: RISK αριθμός ριψοκίνδυνων συμπεριφορών

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	5,673	1,407		4,032	,000
	PERDANG επίγνωση κινδύνου	-,842	,178	-,534	-4,731	,000
	PERCONS επίγνωση επιπτώσεων	,014	,272	,006	,051	,960
	CONTROL ικανότητα ελέγχου	,097	,207	,046	,470	,640
	PEERINFL επιρροή συνομηλίκων	,667	,174	,373	3,838	,000

a. Dependent Variable: RISK αριθμός ριψοκίνδυνων συμπεριφορών

Χρησιμοποιήθηκε η ανάλυση πολλαπλής παλινδρόμησης (μέθοδος enter) για να ελεγχθεί η δυνατότητα πρόβλεψης του αριθμού ριψοκίνδυνων συμπεριφορών από ψυχολογικούς και κοινωνικούς παράγοντες. Ως μεταβλητές πρόβλεψης χρησιμοποιήθηκαν: η επίγνωση του κινδύνου, η επίγνωση των επιπτώσεων, η αντιλαμβανόμενη ικανότητα ελέγχου και η επιρροή των συνομηλίκων. Ο δείκτης πολλαπλής συνάφειας είναι ίσος με 0,64 και ο προσαρμοσμένος συντελεστής προσδιορισμού R^2 είναι ίσος με 0,37. Δηλαδή, 37% της διασποράς του αριθμού ριψοκίνδυνων συμπεριφορών μπορεί να ερμηνευθεί από την επίδραση των ανεξάρτητων μεταβλητών. Η κλίση της γραμμής παλινδρόμησης είναι σημαντικά διαφορετική του μηδενός, $F(4, 64) = 11,06$, $p < 0,001$. Από την επισκόπηση των συντελεστών παλινδρόμησης διαπιστώνουμε ότι δύο (από τις τέσσερις) ανεξάρτητες μεταβλητές συμβάλλουν σημαντικά στην πρόβλεψη της εξαρτημένης: η «επίγνωση του κινδύνου» ($\beta = -0,53$, $t = -4,73$, $p < 0,001$) και η «επιρροή των συνομηλίκων» ($\beta = 0,37$, $t = 3,84$, $p < 0,001$). Δηλαδή, όσο μικρότερη η επίγνωση του κινδύνου και όσο μεγαλύτερη η επιρροή των συνομηλίκων, τόσο μεγαλύτερος ο αριθμός ριψοκίνδυνων συμπεριφορών (βλ. Πίν. 4).

Πίνακας 4.

Ανάλυση παλινδρόμησης για τη στατιστική πρόβλεψη του αριθμού ριψοκίνδυνων συμπεριφορών από ψυχολογικούς και κοινωνικούς παράγοντες (N = 68)

Μεταβλητές πρόβλεψης	*B*	SE *B*	beta
Επίγνωση του κινδύνου	-0,84	0,18	-0,53***
Επίγνωση των επιπτώσεων	0,01	0,27	0,01
Ικανότητα ελέγχου	0,10	0,21	0,05
Επιρροή των συνομηλίκων	0,67	0,17	0,37***

Σημείωση. *** $p < 0,001$. Εξαρτημένη μεταβλητή: Αριθμός ριψοκίνδυνων συμπεριφορών (μέθοδος enter). $R^2 = 0,41$, $F(4, 64) = 11,06$, $p < 0,001$.

7. Ενδεικτική Βιβλιογραφία

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

Ανδριώτης, Κ. (2003). *Ποσοτική έρευνα και ανάλυση δεδομένων με τη χρήση του SPSS 11.5*. Αθήνα: Κλειδάριθμος.

Δαφέρμος, Β. (2005). *Κοινωνική Στατιστική με το SPSS*. Θεσσαλονίκη: Ζήτη.

Howitt, D., & Cramer, D. (2003). *Στατιστική με το SPSS 11 για Windows* (μτφρ. Κ. Καρανικολός). Αθήνα: Κλειδάριθμος.

Levesque, R. (2001). *Free General and Specialized Online SPSS tutorials*. Retrieved December 18, 2005, from http://pages.infinit.net/rlevesqu/spss.htm

Ρετινιώτης, Σ. (2003). *Στατιστική από τη θεωρία στην πράξη με το SPSS 11.0*. Αθήνα: Εκδόσεις Νέων Τεχνολογιών.

SPSS, Inc. (2005). *SPSS 13 for Windows student version (CD Rom).* Englewood Cliffs, NJ: Prentice-Hall.

Texas A&M University, Department of Statistics (2005). *SPSS tutorials*. Retrieved December 18, 2005, from http://www.stat.tamu.edu/spss.php

ΠΑΡΑΡΤΗΜΑ Α. Διάγραμμα απόφασης για στατιστική ανάλυση

➢ **Κατανομή συχνότητας κατηγορικής διμεταβλητής:**	
ΠΑΡΑΜΕΤΡΙΚΟ ΚΡΙΤΗΡΙΟ	ΑΠΑΡΑΜΕΤΡΙΚΟ ΚΡΙΤΗΡΙΟ
	κριτήριο χ^2

➢ **Συνάφεια αριθμητικής διμεταβλητής:**	
ΠΑΡΑΜΕΤΡΙΚΟΣ ΔΕΙΚΤΗΣ	ΑΠΑΡΑΜΕΤΡΙΚΟΣ ΔΕΙΚΤΗΣ
Pearson r	Spearman rho

➢ **Στατιστική πρόβλεψη (αριθμητικές μεταβλητές):**	
ΠΑΡΑΜΕΤΡΙΚΟ ΚΡΙΤΗΡΙΟ	ΑΠΑΡΑΜΕΤΡΙΚΟ ΚΡΙΤΗΡΙΟ
Ανάλυση παλινδρόμησης (Regression)	

➢ **Σύγκριση δύο μέσων όρων:**	
➢ **Ανεξάρτητα ή εξαρτημένα δείγματα;**	
ΠΑΡΑΜΕΤΡΙΚΑ ΚΡΙΤΗΡΙΑ	ΑΠΑΡΑΜΕΤΡΙΚΑ ΚΡΙΤΗΡΙΑ
t-test για δύο *ανεξάρτητα* δείγματα	Mann Whitney U
t-test για δύο *εξαρτημένα* δείγματα	Wilcoxon

➢ **Σύγκριση περισσότερων από δύο μέσων όρων:**	
➢ **Ανεξάρτητα ή εξαρτημένα δείγματα;**	
➢ **Μονοπαραγοντικά ή διπαραγοντικά δείγματα;**	
ΠΑΡΑΜΕΤΡΙΚΑ ΚΡΙΤΗΡΙΑ	ΑΠΑΡΑΜΕΤΡΙΚΑ ΚΡΙΤΗΡΙΑ
Μονοπαραγοντική ANOVA (F-κριτήριο) για *ανεξάρτητα* δείγματα	Kruskall-Wallis
Διπαραγοντική ANOVA (F-κριτήριο) για *ανεξάρτητα* δείγματα	
Μονοπαραγοντική ANOVA (F-κριτήριο) για *εξαρτημένα* δείγματα	Friedman

Σημείωση. Τα **απαραμετρικά κριτήρια** χρησιμοποιούνται όταν δεν πληρούνται οι προϋποθέσεις εφαρμογής των αντίστοιχων παραμετρικών κριτηρίων όπως, π.χ., πολύ μικρά δείγματα (N < 30), ιδιαίτερα ασύμμετρη κατανομή, ιδιαίτερα ανομοιογενείς ομάδες, κλπ. Στο SPSS ο υπολογισμός των απαραμετρικών στατιστικών δεικτών γίνεται μέσα από το μενού εντολών «ANALYZE» → «NONPARAMETRIC STATISTICS…».

ΠΑΡΑΡΤΗΜΑ Β. Προετοιμασία δεδομένων για στατιστική ανάλυση

Η προετοιμασία και ο σχολαστικός έλεγχος ενός αρχείου δεδομένων αποτελούν πολύ σημαντικά προκαταρκτικά βήματα πριν από τη διεξαγωγή οποιασδήποτε στατιστικής ανάλυσης και μπορεί να προστατέψουν τον ερευνητή από εντελώς λανθασμένα ή παραπλανητικά συμπεράσματα. Παρακάτω καταγράφονται μερικά από τα συνηθέστερα προβλήματα που ενδέχεται να εμφανιστούν στη φάση αυτή, καθώς και οι προτεινόμενοι τρόποι για την αντιμετώπισή τους. Η μελέτη τους συνίσταται ιδιαιτέρως στους πρωτόπειρους ερευνητές, ακόμα και αν (ή, καλύτερα, ακριβώς επειδή) δεν διαθέτουν σε βάθος γνώσεις Στατιστικής. Στην επιστήμη αυτή και στην έρευνα, γενικότερα, η επίγνωση του μέρους εκείνου της γνώσης που δεν κατέχουμε, αποτελεί την πιο αποτελεσματική δικλείδα ασφαλείας από τη λανθασμένη εφαρμογή των ημιτελών γνώσεών μας…

- **Εγκυρότητα.** Τα δεδομένα πρέπει να έχουν περαστεί στον υπολογιστή προσεκτικά, χωρίς λάθη παράλειψης ή αντιμετάθεσης. Προσοχή χρειάζεται στις κατηγορικές μεταβλητές που έχουν κωδικοποιηθεί με αριθμούς.

✓ Αντιμετώπιση: Ελέγχουμε την κατανομή συχνότητας, το εύρος των τιμών, το μέσο όρο και την τυπική απόκλιση κάθε μεταβλητής. Τις πληροφορίες αυτές μπορούμε να ζητήσουμε στο SPSS μέσω της εντολής «Analyze» → «Descriptive Statistics» → «Descriptives».

- **Ελλιπή δεδομένα (Missing data).** Ελλιπή δεδομένα έχουμε όταν τα υποκείμενα της έρευνας δεν παρέχουν πληροφορίες για όλες τις μεταβλητές. Τα ελλιπή στοιχεία εμφανίζονται στο αρχείο δεδομένων είτε ως συγκεκριμένες τιμές (π.χ. 99) είτε ως κενά. Ελέγχουμε τις κατανομές συχνότητας (στο SPSS: εντολή «Analyze» → «Descriptive Statistics» → «Frequencies») για να δούμε πόσα είναι τα ελλιπή δεδομένα. Ιδιαίτερα σημαντικό είναι να διαπιστώσουμε εάν η κατανομή των ελλιπών δεδομένων παρουσιάζει κανονικότητα, π.χ. αν εμφανίζονται συστηματικά σε μία συγκεκριμένη μεταβλητή ή αν συνδέονται με κάποια χαρακτηριστικά του δείγματος (π.χ. αν προέρχονται από άτομα χαμηλής μόρφωσης).

✓ Αντιμετώπιση: (α) Αν υπάρχει μεγάλος αριθμός ελλιπών δεδομένων σε συγκεκριμένα υποκείμενα ή μεταβλητές, διαγράφουμε τα υποκείμενα ή τις μεταβλητές. (β) Χωρίζουμε το δείγμα σε δύο ομάδες (με ελλιπή και χωρίς ελλιπή δεδομένα) και ελέγχουμε αν οι δύο αυτές ομάδες διαφοροποιούνται ως προς τις εξαρτημένες μεταβλητές. (γ) Συμπληρώνουμε τα ελλιπή δεδομένα υπολογίζοντας τις τιμές τους, είτε με βάση άλλες πηγές πληροφόρησης, είτε με το μέσο όρο του δείγματος ή της ομάδας στην οποία ανήκουν, είτε με εκτίμηση μέσω παλινδρομικής ανάλυσης. (δ) Επαναλαμβάνουμε τις στατιστικές αναλύσεις με και χωρίς τα ελλιπή δεδομένα και συγκρίνουμε τα αποτελέσματα. (ε) Πρόληψη: φροντίζουμε να μην έχουμε ελλιπή δεδομένα με τον κατάλληλο ερευνητικό σχεδιασμό.

- **Ακραίες τιμές (Outliers).** Οι ιδιαίτερα απομακρυσμένες από τον μέσο όρο τιμές επηρεάζουν έντονα τους στατιστικούς δείκτες και γι' αυτό πρέπει να απομονώνονται. Οι ακραίες τιμές μπορεί να προέρχονται από λανθασμένη εισαγωγή δεδομένων ή από υποκείμενα που προέρχονται από άλλο πληθυσμό, συγκριτικά με το υπόλοιπο δείγμα. Στις κατηγορικές μεταβλητές ακραία θεωρείται μια τιμή με πολύ χαμηλή συχνότητα εμφάνισης, π.χ. μικρότερη από 10%.

✓ Αντιμετώπιση: Ελέγχουμε μήπως έγινε λάθος στην εισαγωγή των δεδομένων. Μήπως κάποια μεταβλητή ή κάποιο υποκείμενο ευθύνονται για τις ακραίες τιμές; Σε αυτή την περίπτωση, διαγράφουμε τη συγκεκριμένη μεταβλητή ή το συγκεκριμένο υποκείμενο. Αν οι ακραίες τιμές είναι λίγες και εμφανίζονται τυχαία, διαγράφουμε τις συγκεκριμένες ακραίες τιμές. Μπορούμε, ακόμα, να χρησιμοποιήσουμε

μια αλγοριθμική μετατροπή της μεταβλητής, ώστε η κατανομή να αποκτήσει πιο ομαλή κατανομή.

- **Πραγματική συνάφεια (Honest correlation).** Η συνάφεια μεταξύ δύο μεταβλητών μπορεί να είναι τεχνητά αυξημένη (inflated) ή τεχνητά μειωμένη (deflated).

✓ Η συνάφεια είναι **τεχνητά αυξημένη** αν δύο παράγωγες μεταβλητές έχουν υπολογιστεί με βάση έναν αριθμό κοινών items (αντιμετώπιση: κρατάμε μόνο τη μία μεταβλητή στην ανάλυση και αφαιρούμε τη δεύτερη).

✓ Η συνάφεια είναι **τεχνητά μειωμένη** (α) αν μία μεταβλητή δεν έχει μεγάλο εύρος (αντιμετώπιση: υπάρχει μαθηματικός τύπος αναπροσαρμογής του δείκτη συνάφειας), (β) αν το δείγμα είναι ιδιαίτερα άνισα κατανεμημένο, π.χ. 90%-10%, στις τιμές μιας διχότομης μεταβλητής (αντιμετώπιση: διασπάμε μια μεγάλη κατηγορία σε μικρότερες όπου αυτό είναι δυνατό), ή (γ) αν ο μέσος όρος *M* μίας μεταβλητής είναι πολύ μεγάλος και η τυπική απόκλιση *s* πολύ μικρή, δηλ. αν ο λόγος μεταβλητότητας $\frac{s}{M}$ < 0,0001 (αντιμετώπιση: αφαιρούμε μια σταθερή τιμή από όλες τις τιμές της μεταβλητής).

- **Κανονικότητα (Normality).** Η κανονικότητα, δηλ. η κανονική κατανομή συχνότητας των τιμών μιας μεταβλητής αποτελεί προϋπόθεση για πολλές στατιστικές αναλύσεις. Έλεγχος της κανονικότητας της κατανομής μπορεί να γίνει με την επισκόπηση του ιστογράμματος συχνότητας. Επίσης, μπορεί να ελεγχθούν οι δείκτες συμμετρίας (skewness) και κύρτωσης (kurtosis) με το z-κριτήριο ($z_s = \frac{s}{s_s}$ και $z_k = \frac{k}{s_k}$).

✓ Αντιμετώπιση: Εάν εντοπίσουμε μια έντονα ασύμμετρη κατανομή χρειάζεται να τη μετατρέψουμε ώστε να γίνει πιο ομαλή. Ο τύπος μετατροπής εξαρτάται από το μέγεθος, την κατεύθυνση και τη μορφή της ασυμμετρίας.

- **Γραμμικότητα (Linearity).** Η γραμμικότητα αναφέρεται στη ευθύγραμμη συσχέτιση μεταξύ δύο μεταβλητών. Έλεγχος της γραμμικότητας μπορεί να γίνει με την επισκόπηση του διαγράμματος σκεδασμού.

✓ Αντιμετώπιση: Αν εντοπίσουμε διμεταβλητή με μη γραμμική (καμπύλη) συσχέτιση, π.χ. αριθμός συμπτωμάτων και δόση φαρμάκου, μπορούμε να μετατρέψουμε τη μία μεταβλητή σε διχότομη (dummy variable, δηλ. του τύπου 0-1) και να χρησιμοποιήσουμε αυτή για περαιτέρω αναλύσεις.

- **Ομοιοσκεδασμός (Homoscedasticity).** Σημαίνει ότι η διασπορά των τιμών μίας μεταβλητής είναι περίπου ίδια σε όλες τις ομάδες που ορίζονται από τις τιμές-επίπεδα μίας άλλης. Ο ετεροσκεδασμός προκαλείται συνήθως όταν η κατανομή μίας από τις δύο μεταβλητές δεν είναι ομαλή. Μπορεί να ελεγχθεί με επισκόπηση του διαγράμματος σκεδασμού.

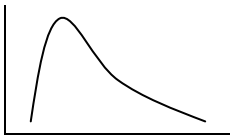
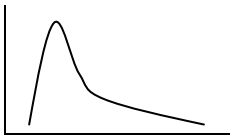
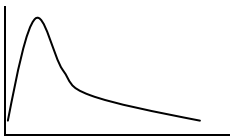
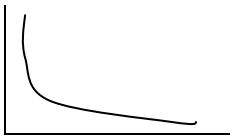
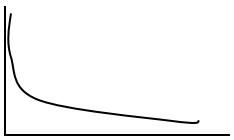
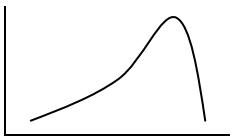
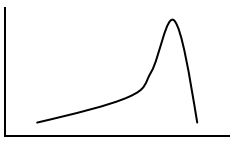
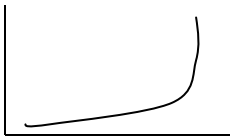
✓ Αντιμετώπιση: Ίσως χρειαστεί να μετατρέψουμε τη μεταβλητή που δεν έχει κανονική κατανομή. Πάντως, ο ετεροσκεδασμός δεν είναι τόσο κρίσιμος για τη στατιστική ανάλυση, όσο η κανονικότητα και η γραμμικότητα.

- **Πολυσυγγραμμικότητα (Multicollinearity).** Είναι πρόβλημα που συναντάμε σε ένα πίνακα συναφειών, όταν υπάρχουν ζεύγη μεταβλητών με πολύ υψηλή συνάφεια (r > 0,90). Συμβαίνει όταν διαφορετικές μεταβλητές μετρούν περίπου την ίδια ψυχολογική ιδιότητα, π.χ. δύο κλίμακες νοημοσύνης. Η πολυσυγγραμμικότητα δυσχεραίνει σημαντικά την ερμηνεία των ευρημάτων.

✓ Αντιμετώπιση: Αποκλείουμε από την ανάλυση μία από τις μεταβλητές που εμφανίζουν πολύ υψηλή συνάφεια μεταξύ τους.

Προτεινόμενες μετατροπές αριθμητικών μεταβλητών προκειμένου να αντιμετωπιστούν προβλήματα ασυμμετρίας/κύρτωσης

Πηγή: Tabachnick, B., & Fidell, L. (2006). *Using multivariate statistics* **(5th ed.). Boston: Allyn & Bacon.**

	Είδος ασυμμετρίας/κύρτωσης	*Μετατροπή στο SPSS*
	Μέτρια θετική ασυμμετρία	newX = SQRT(X)
	Μεγάλη θετική ασυμμετρία	newX = LOG10(X)
	Θετική ασυμμετρία (με τιμή 0)	newX = LOG10(X + C)
	Μεγάλη θετική ασυμμετρία σχήματος L	newX = 1 / X
	Μεγάλη θετική ασυμμετρία σχήματος L (με τιμή 0)	newX = 1 / (X + C)
	Μέτρια αρνητική ασυμμετρία	newX = SQRT(K – X)
	Μεγάλη αρνητική ασυμμετρία	newX = LOG10(K – X)
	Μεγάλη αρνητική ασυμμετρία σχήματος J	newX = 1 / (K – X)

C = μια σταθερά που προστίθεται σε κάθε τιμή της μεταβλητής έτσι ώστε η μικρότερη τιμή να είναι 1.
K = μια σταθερά που αφαιρείται από κάθε τιμή της μεταβλητής έτσι ώστε η μικρότερη τιμή να είναι 1 (συνήθως είναι ίση με τη μέγιστη τιμή + 1)