

ΕΙΣΑΓΩΓΗ ΣΤΟ SPSS

Ε. Σαμόλη, Β. Σύψα

Σκοπός της πρακτικής

- Εισαγωγή στις στατιστικές εφαρμογές των πακέτων SPSS
- Κατανόηση της βασικής στατιστικής ανάλυσης των ιατρικών δεδομένων

Εισαγωγή στο SPSS

- Ευρύ φάσμα διαφόρων στατιστικών δοκιμασιών.
- Εύκολος χειρισμός δεδομένων και γραφικών με μεγάλη συμβατότητα με άλλα πακέτα
- Εύκολος χειρισμός εντολών μέσα από το menu bar
- Διαφορετικά παράθυρα με ανεξάρτητα menu bar για ευκολότερο χειρισμό (data, output, chart, command window)

1. Τα παράθυρα του SPSS

Ανοίγοντας το SPSS εμφανίζεται το data **editor window** στο οποίο βρίσκεται το menu bar με όλες τις προσφερόμενες διαδικασίες, στατιστικές ή μη.

Γενικά το SPSS αποτελείται από τα ακόλουθα παράθυρα:

- το **data editor window** έχει δυο φύλλα. Στο φύλλο με το όνομα DATA VIEW εμφανίζονται τα δεδομένα του χρησιμοποιούμενου αρχείου και το οποίο είναι αρχικά κενό. Αποτελείται από κελιά που σχηματίζονται από γραμμές και στήλες όπου οι γραμμές αντιπροσωπεύουν τις εγγραφές (cases) και οι στήλες τις μεταβλητές (variables) του αρχείου. Στο παράθυρο αυτό μπορούν να δημιουργηθούν νέα αρχεία (να καταχωρηθούν δηλαδή δεδομένα) ή να ανοιχθούν ήδη υπάρχοντα. Όλα αυτά τα αρχεία έχουν την κατάληξη “sav”.

- το **output window** όπου εμφανίζονται τα αποτελέσματα των επιλογών που κάνει ο χρήστης του προγράμματος από το menu (σχήματα, κατανομές κλπ). Ο χρήστης μπορεί να σβήσει ότι δεν χρειάζεται από το output και να σώσει τα αποτελέσματα/σχήματα σε αρχεία για περαιτέρω χρήση. Όλα τα output αρχεία έχουν την κατάληξη “spo”.

- το **syntax window** όπου μπορεί κανείς να γράψει τις διάφορες εντολές αντί να χρησιμοποιήσει το menu και στη συνέχεια να τις τρέξει με το RUN ή, αν έχει χρησιμοποιήσει το μενού, να τις κάνει paste από το output window και στη συνέχεια να τις σώσει σε αρχείο για περαιτέρω χρήση. Τα syntax αρχεία έχουν την κατάληξη “sps”.
- το **chart editor window** όπου γίνονται μεταβολές στα σχήματα του output.

2. Αρχεία δεδομένων

2.1. Εισαγωγή δεδομένων

Μπαίνοντας στο SPSS, το data editor window-data view είναι κενό. Στις στήλες του μπορούν να οριστούν οι μεταβλητές του νέου αρχείου. Κάνοντας double-click στον τίτλο **var** κάθε στήλης μεταφερόμαστε στο παράθυρο **variable view** όπου ορίζουμε το όνομα της μεταβλητής, τον τύπο της (type, πχ numeric, string κλπ) και ποιους αριθμοί υποδηλώνουν τα missing values. Όταν οριστούν οι μεταβλητές του αρχείου μπορεί κανείς να αρχίσει την καταχώρηση των δεδομένων ανά γραμμή. Κελιά που αφήνονται κενά σε κάποια εγγραφή θεωρούνται ελλείπουσες τιμές (missing values). Επειδή συχνά το κενό κελί δημιουργεί σύγχυση για το αν η πληροφορία ξεχάστηκε να καταχωρηθεί, είναι σε εκκρεμότητα ή είναι πράγματι missing value συνηθίζεται οι τελευταίες να καταχωρούνται με κάποια αριθμό (συνήθως 9, 99, 999 ανάλογα με την κατανομή μας). Η κίνηση μέσα στο παράθυρο γίνεται με τα βελάκια ενώ ισχύουν και τα παρακάτω :

- | | |
|--|-------------------|
| -μετακίνηση στην αρχή της γραμμής | : αριστερό βελάκι |
| -μετακίνηση στο τέλος της γραμμής | : δεξί βελάκι |
| -μετακίνηση στο πρώτο κελί του αρχείου | : Home |
| -μετακίνηση στο τελευταίο κελί του αρχείου | : End |

2.2 Χρήση προϋπαρχόντων αρχείων δεδομένων

Για να επεξεργαστεί κανείς τα δεδομένα ενός αρχείου, πρέπει αρχικά να το ανοίξει από το menu με το **File – Open – Data** και να το επιλέξει, οπότε αυτό θα εμφανιστεί στο data editor window.

Όπως αναφέρθηκε, τα αρχεία δεδομένων του SPSS έχουν την κατάληξη **sav** (ή **sys** αν είναι παλιότερη έκδοση του SPSS). Μπορεί επίσης να γίνει εισαγωγή αρχείων άλλης μορφής, όπως για πχ Excel (επιλέγοντας το αντίστοιχο format του αρχείου).

3. Διαχείριση δεδομένων (data management)

Πέρα από το υπο-μενου κάτω από την εντολή **Edit** με τις γνωστές επιλογές (αντιγραφή, διαγραφή κλπ), υπάρχουν τα υπο-μενου κάτω από την εντολή **Data** και κάτω από την εντολή **Transform**. Οι βασικές εντολές για τη διαχείριση δεδομένων βρίσκονται μέσα στις επιλογές **Data** ή **Transform** του μενού και είναι οι εξής :

3.1 Κάτω από την επιλογή Data:

- **SORT** : ταξινομεί τις εγγραφές του αρχείου με βάση την τιμή μίας ή περισσότερων μεταβλητών κατά αύξουσα (ascending) ή φθίνουσα (descending) σειρά.
- **SELECT CASES** : περιορίζει την ανάλυση σε συγκεκριμένες υποομάδες χρησιμοποιώντας κριτήρια που συνήθως βασίζονται στις τιμές κάποιων μεταβλητών (π.χ. ηλικία>40). Οι μη επιλεγμένες εγγραφές μπορούν είτε να διατηρηθούν στο αρχείο (unselected cases are filtered) ή να διαγραφούν εντελώς από αυτό (unselected cases are deleted). Αν στη συνέχεια θέλει κανείς να χρησιμοποιήσει όλες τις εγγραφές ή να επιλέξει κάποιες με βάση άλλο κριτήριο αρκεί να επαναλάβει τη διαδικασία select cases.
- **SPLIT FILE** : για την επεξεργασία των δεδομένων κατά ομάδες, δηλαδή χωρίζει κατά κάποιο τρόπο το αρχείο σε ομάδες.
- **MERGE FILES** : χρησιμοποιείται για να ενώσει δύο αρχεία με δύο διαφορετικούς τρόπους :
 - ένωση αρχείων με τις ίδιες μεταβλητές αλλά διαφορετικές εγγραφές (**Merge files - Add cases**) : προσθέτει νέες εγγραφές/παρατηρήσεις (cases) στο χρησιμοποιούμενο αρχείο από δεύτερο αρχείο.
 - ένωση αρχείων με τις ίδιες εγγραφές αλλά διαφορετικές μεταβλητές (**Merge files -Add Variables**) : προσθέτει σε ήδη υπάρχουσες εγγραφές επιπλέον μεταβλητές από δεύτερο αρχείο. Απαραίτητη προϋπόθεση είναι και τα

δύο αρχεία να είναι ταξινομημένα ως προς την τιμή κάποιας κοινής μεταβλητής (πχ αριθμός ατόμου) ώστε οι νέες μεταβλητές να αντιστοιχηθούν στη σωστή εγγραφή.

3.2 Κάτω από την επιλογή **EDIT** :

- **INSERT VARIABLE / INSERT CASE** : για την εισαγωγή σε συγκεκριμένη θέση καινούργιας μεταβλητής ή παρατήρησης.

3.3 Κάτω από την επιλογή **TRANSFORM** :

- **COMPUTE** : δημιουργεί μεταβλητές που είναι αριθμητικοί μετασχηματισμοί άλλων μεταβλητών (πχ μία νέα μεταβλητή που είναι ο λογάριθμος μίας ήδη υπάρχουσας).

- **RECODE** : επανακωδικοποιεί τις τιμές μίας ήδη υπάρχουσας μεταβλητής ή δημιουργεί νέα μεταβλητή με βάση τις επανακωδικοποιημένες τιμές της υπάρχουσας μεταβλητής (πχ μία μεταβλητή που υποδηλώνει την ηλικία και είναι κωδικοποιημένη ως 1 : <20 έτη, 2 : 20-30 έτη, 3 : >30 έτη μπορεί να γίνει recode ώστε η κατηγορία >30 έτη να ενοποιηθεί με την κατηγορία 20-30 θέτοντας σαν old value to 3 και σαν new value to 2).

4. Δημιουργία-μεταβολές σχημάτων

Η επιλογή **GRAPH** του μενού προσφέρει τη δυνατότητα για τη δημιουργία διαφόρων ειδών σχημάτων όπως ιστογράμματα (**HISTOGRAM**), πίτες (**PIE**), θηκογράμματα (**BOXPLOT**) κλπ. Γενικά τα γραφικά έχουν επιλογές για την χρήση της μεταβλητής (δηλαδή αν θα χρησιμοποιηθεί όπως είναι ή κάποιο αντίστοιχο μέτρο της όπως η μέση τιμή) και για πολλαπλά γραφικά στο ίδιο διάγραμμα (πχ η γραφική παράσταση μιας μεταβλητής σε υπο-στρώματα που καθορίζονται από μία άλλη μεταβλητή). Με την επιλογή του είδους σχήματος, εμφανίζεται ένα παράθυρο που ζητάει τη μεταβλητή/μεταβλητές που απαιτούνται για το σχεδιασμό του. Για να κάνει ο χρήστης αλλαγές στο σχήμα που προκύπτει αρκεί να το κάνει double-click οπότε μεταφέρεται σε νέο παράθυρο, το chart editor, όπου προσφέρονται επιλογές όπως διαμόρφωση αξόνων (scale, range, tick marks), προσθήκη τίτλων, μεταβολή χρωμάτων κλπ. Τελειώνοντας τις αλλαγές, ο χρήστης φεύγει από το chart editor

κάνοντας File-Close και επιστρέφει έτσι στο output window όπου το σχήμα εμφανίζεται πια ανανεωμένο.

5. Στατιστική Ανάλυση στο SPSS

Οι εντολές που χρησιμοποιούνται στην στατιστική ανάλυση βρίσκονται κάτω από την επιλογή **ANALYSE** του menu bar. Οι βασικότερες εντολές που χρησιμοποιούνται ανάλογα με το είδος της ανάλυσης παρουσιάζονται παρακάτω :

Περιγραφική Στατιστική → Descriptive Statistics

- **Frequencies:** Δίνει κατανομές συχνοτήτων, εκατοστημόρια, χαρακτηριστικές τιμές θέσης και διασποράς. Υπάρχει η επιλογή δημιουργίας γραφικών μέσα από την εντολή για τον έλεγχο της κατανομής (ιστόγραμμα, bar chart). (Παράγονται για τις μεταβλητές που μεταφέρονται από τον κατάλογο των μεταβλητών από τα αριστερά προς τα δεξιά)
- Οι χαρακτηριστικές τιμές θέσης και διασποράς παρέχονται και μέσα από το **Descriptives**
- Για ανάλυση κατά ομάδες υπάρχει το **Explore**
- Για την παραγωγή πινάκων υπάρχει το **Crosstabs** (όπου καθορίζονται οι μεταβλητές που θα χρησιμοποιηθούν για τις σειρές και για τις στήλες). Υπάρχει η δυνατότητα δημιουργίας πινάκων κατά υπο-ομάδες (*strata*) και η παραγωγή διαφόρων δοκιμασιών όπως το χ^2 (chi-square). Περισσότερες επιλογές για την δημιουργία πινάκων βρίσκονται στο **Tables – Custom Tables**

Σύγκριση μέσων τιμών δύο ομάδων (παραμετρικές μέθοδοι) →

Compare Means

- **Independent Samples T-test :** Εισάγουμε τις μεταβλητές που θέλουμε να συγκρίνουμε στο κουτί *Test variables* και το όνομα της μεταβλητής που καθορίζει τις δύο συγκρινόμενες ομάδες στο κουτί *Grouping variable*, όπου πρέπει να καθοριστούν και οι τιμές που διαφοροποιούν τις δύο ομάδες μέσα από το *Define groups*.

- **Paired Sample t-test**, όπου καθορίζουμε τα ζεύγη των μεταβλητών υπό εξέταση

Διερεύνηση συσχέτισης μεταξύ δύο ποιοτικών μεταβλητών

- χ^2 για τη διερεύνηση της σχέσης μεταξύ δύο ποιοτικών μεταβλητών, που ορίζεται μέσα από την εντολή **Crosstabs**

Διερεύνηση συσχέτισης μεταξύ δύο ποσοτικών μεταβλητών

- **Pearson** correlation coefficient για τη συσχέτιση μεταξύ δύο ποσοτικών μεταβλητών, που ορίζεται μέσα από την εντολή **Correlate - Bivariate**, όπου δηλώνουμε τα ονόματα των μεταβλητών.

Στατιστική Εξάρτηση - Παλινδρόμηση

→ Regression

- **Linear** (Γραμμική Παλινδρόμηση), όπου καθορίζεται η εξαρτημένη και οι ανεξάρτητες μεταβλητές στις αντίστοιχες θέσεις. Υπάρχει η επιλογή εμφάνισης στα αποτελέσματα των διαστημάτων εμπιστοσύνης, του πίνακα διακύμανσης και των προγνωστικών μέτρων του μοντέλου, όπως επίσης και διάφορες επιλογές για γραφικές παραστάσεις και δημιουργία νέων μεταβλητών που θα περιέχουν τα υπόλοιπα ή τις εκτιμώμενες τιμές.
- **Binary Logistic** (Λογαριθμιστική παλινδρόμηση), με αντίστοιχες επιλογές με linear regression. Επιπλέον επιτρέπει τον ορισμό ποιοτικών μεταβλητών με κατηγορία αναφοράς την πρώτη ή την τελευταία κατηγορία της μεταβλητής. Εδώ προσοχή ορίζουμε 1=cases, 0=controls.

Σύγκριση μεταβλητών με μη-παραμετρικές μεθόδους → Nonparametric tests

- **Wilcoxon** και **Sign**: για την σύγκριση δύο σχετιζόμενων δειγμάτων (αντίστοιχα του t-test), όπου ορίζονται μέσα από την εντολή **Nonparametric tests - 2 Related Samples**, τσεκάροντας τα αντίστοιχα κουτάκια.

- **Spearman**: για την μη-παραμετρική συσχέτιση δύο ποσοτικών μεταβλητών, το οποίο ορίζεται μέσα από την εντολή **Correlate-Bivariate** τσεκάροντας το κουτί για **Spearman**.

6. Αποθήκευση - εκτύπωση - έξοδος

Αποθήκευση : Οι εντολές **SAVE** και **SAVE AS** αποθηκεύουν σε αρχεία ότι υπάρχει στο τρέχων παράθυρο.

Εκτύπωση : Η εντολή **PRINT** χρησιμοποιείται για την εκτύπωση δεδομένων, αποτελεσμάτων και σχημάτων. Ειδικά για τις εκτυπώσεις από το output window, μπορεί κανείς να τυπώσει επιλεκτικά κάνοντας ένα click για να επιλέξει πχ ένα σχήμα και διαλέγοντας το **SELECTION** από το **PRINT RANGE**.

Έξοδος από το SPSS : με **FILE - CLOSE**.