

# Εισαγωγή στο SPSS

- Οι οριζόντιες γραμμές αντιστοιχούν στις N περιπτώσεις-πειραματικές μονάδες, ενώ οι κατακόρυφες στήλες στις  $p$  υπό μελέτη μεταβλητές.
- ΠΟΣΟΤΙΚΕΣ-ΠΟΙΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ
- ΠΑΡΑΔΕΙΓΜΑ                    ΓΙΑ                    ΕΙΣΑΓΩΓΗ  
ΔΕΔΟΜΕΝΩΝ

# ΟΝΟΜΑΣΙΑ ΜΕΤΑΒΛΗΤΩΝ

- Μπορεί να έχει το πολύ 64 χαρακτήρες στις συνήθεις γλώσσες και κάθε τέτοιο είναι μοναδικό,
- πρέπει να ξεκινά με γράμμα ή με @, #, ή \$,
- δεν πρέπει να περιέχει σημεία στίξης, αστεράκια καθώς και κενά,
- δεν μπορεί να περικλείονται οι λέξεις All, Ne, Eq, To, Le, Lt, By, Or, Gt, And, Not, Ge, With,
- μπορούν να γραφούν τόσο με μικρά όσο και με κεφαλαία γράμματα, τέλος
- καλό θα ήταν να αποφεύγουμε να έχουμε ως τελευταίο χαρακτήρα του ονόματός τους τόσο την τελεία όσο και την κάτω παύλα.

# Variable View

1. Name: εισάγουμε την κωδική ονομασία που επιθυμούμε
2. Type: το λογισμικό με βάση τις τιμές που πληκτρολογούμε καθορίζει αυτόματα τον τύπο της μεταβλητής
  - Numeric: αριθμητικές (numeric) με 2 δεκαδικά ψηφία (decimals) και συνολικό μήκος (δηλώνεται στο πλαίσιο Width) 8 θέσεων (για τον υπολογισμό λαμβάνονται υπόψη το πρόσημο, το ακέραιο μέρος, η δεκαδική τελεία καθώς και το δεκαδικό μέρος).
  - Comma. Αριθμητικές τιμές που έχουν το κόμμα «,» ανά τρεις θέσεις και την τελεία «.» ως υποδιαστολή.
  - Dot. Αριθμητικές τιμές που έχουν τελεία «.» ανά τρεις θέσεις και το κόμμα «,» ως υποδιαστολή.

# Variable View

- Scientific notation. Ποσοτική μεταβλητή της οποίας οι αριθμητικές τιμές γράφονται σε επιστημονική μορφή. Για παράδειγμα στη μορφή 123, 1.23E2, 1.23D2, 1.23E+2, και 1.23+2
- Date. Ημερομηνίες
- Dollar. Τιμές δολαρίου, με ή χωρίς το σύμβολο (\$), με την τελεία ως υποδιαστολή και το κόμμα ανά τρεις θέσεις.
- Custom currency. Αριθμητικές τιμές ο τρόπος εμφάνισής τους καθορίζεται από το χρήστη.
- String. Μεταβλητή που δεν είναι αριθμητική και επομένως δεν χρησιμοποιείται στους υπολογισμούς.

# Variable View

- **Labels:** δηλώνεται η πλήρης περιγραφή του ονόματος της μεταβλητής.
- **Values** μπορούμε να δηλώσουμε τα ονόματα των διάφορων περιπτώσεων των κατηγορικών μεταβλητών.
- **Missing Values** καθορίζουμε τις τιμές των ελλιπών τιμών μίας μεταβλητής.
- **Columns, Align** και **Measure** καθορίζουμε το μήκος κάθε στήλης, την στοίχιση και το είδος της μεταβλητής.

# ΕΙΣΑΓΩΓΙΚΑ

- ΑΠΟΘΗΚΕΥΣΗ ΔΕΔΟΜΕΝΩΝ
- ΑΝΑΓΝΩΣΗ ΑΡΧΕΙΩΝ EXCEL
- ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ ΔΕΔΟΜΕΝΩΝ  
**TRANSFORM COMPUTE VARIABLE**
- ΕΠΑΝΑΚΩΔΙΚΟΠΟΙΗΣΗ  
**TRANSFORM RECODE**
- ΕΠΙΛΟΓΗ ΠΕΙΡΑΜΑΤΙΚΩΝ ΜΟΝΑΔΩΝ  
**DATA SELECT CASES**

# ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

## ΠΟΙΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ

- **ΠΙΝΑΚΑΣ ΣΥΧΝΟΤΗΤΩΝ**

Προκύπτει από την απαρίθμηση και καταγραφή των δειγματικών τιμών στην αντίστοιχη κατηγορία. Περιλαμβάνει τη στήλη των Συχνοτήτων, Σχετικών Συχνοτήτων. Επιπλέον, μπορούν να συμπεριληφθούν στον πίνακα συχνοτήτων διατάξιμων μόνο ποιοτικών μεταβλητών η στήλη των Αθροιστικών Συχνοτήτων και η στήλη των Αθροιστικών σχετικών Συχνοτήτων

# ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

## ΠΟΙΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ

Συχνότητα παριστάνει τον αριθμό των φορών που μία κατηγορία της ποιοτικής μεταβλητής εμφανίζεται στο

Σχετική συχνότητα παριστάνει το ποσοστό επί τοις εκατό των φορών εμφάνισης μιας τιμής στο δείγμα

Αθροιστική συχνότητα παριστάνει το πλήθος των τιμών του δείγματος που είναι μικρότερες ή το πολύ ίσες από μία τιμή

Αθροιστική σχετική συχνότητα παριστάνει το ποσοστό επί τοις εκατό των τιμών του δείγματος που είναι μικρότερες ή ίσες από μία τιμή.



# ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

## ΠΟΙΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ

- **ΡΑΒΔΟΓΡΑΜΜΑ (Bar chart)**

Μία ειδική γραφική παράσταση που στον οριζόντιο άξονα σημειώνονται οι κατηγορίες στις οποίες τα μέλη του πληθυσμοί κατατάσσονται, ενώ στον κατακόρυφο άξονα οι αντίστοιχες συχνότητες (εναλλακτικά οι αντίστοιχες σχετικές συχνότητες).

# ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

## ΠΟΙΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ

- **ΚΥΚΛΙΚΟ ΔΙΑΓΡΑΜΜΑ (Pie chart)**

Ένας κυκλικός δίσκος χωρισμένος σε τομείς, όσες και οι κατηγορίες στις οποίες τα μέλη του πληθυσμού κατατάσσονται. Το εμβαδό κάθε τομέα απεικονίζει το ποσοστό των ατόμων που ανήκουν στην αντίστοιχη κατηγορία

**ΥΛΟΠΟΙΗΣΗ: ΔΙΑΔΙΚΑΣΙΑ FREQUENCIES**

# ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ ΠΟΣΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ

- ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ

- ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ ΘΕΣΗΣ- ΔΙΑΣΠΟΡΑΣ

- Μέση τιμή, Διάμεσος, Επικρατούσα τιμή, ποσοστιαία σημεία, Τυπική απόκλιση, εύρος, διακύμανση,

- Κατανομής: συντελεστές κύρτωσης και λοξότητας.

- ΙΣΤΟΓΡΑΜΜΑ

- ΦΥΛΛΟΓΡΑΦΗΜΑ

- ΘΗΚΟΓΡΑΜΜΑ

# ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ

- ΜΕΣΗ ΤΙΜΗ (mean)
- ΔΙΑΜΕΣΟΣ (median)
- ΕΠΙΚΡΑΤΟΥΣΑ ΤΙΜΗ
- ΕΥΡΟΣ
- ΔΙΑΚΥΜΑΝΣΗ
- ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ
- ΠΟΣΟΣΤΙΑΙΑ ΣΗΜΕΙΑ
- ΤΕΤΑΡΤΗΜΟΡΙΑ

# ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ

- ΣΥΝΤΕΛΕΣΗΣ ΚΥΡΤΩΣΗΣ

$$b_2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{n S^4}$$

Πλατύκυρτες: τιμές μικρότερες του 3

Λεπτόκυρτες: τιμές μεγαλύτερες του 3

Μεσόκυρτες: τιμές ίσες με 3.

- Το SPSS αφαιρεί την τιμή 3.

# ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ

- Συντελεστής λοξότητας

$$b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{n S^3}$$

Λοξή δεξιά: τιμή μεγαλύτερη του 0.

Λοξή αριστερά: τιμή μικρότερη του 0.

Συμμετρική: τιμή ίση με 0.

# ΙΣΤΟΓΡΑΜΜΑ

- Το ιστόγραμμα συχνοτήτων συνίσταται από ένα σύνολο από συγγενή ορθογώνια παραλληλόγραμμα των οποίων το ύψος είναι ανάλογο με τη συχνότητα κάθε ομάδας και το μήκος τους ανάλογο με το μήκος της ομάδας. Οι τιμές της μεταβλητής τοποθετούνται στον οριζόντιο άξονα, ενώ οι συχνότητες στον κατακόρυφο άξονα.

# ΦΥΛΛΟΓΡΑΦΗΜΑ

- Μία παραλλαγή του ιστογράμματος είναι το φυλλογράφημα (steam and leaf plot). Το φυλλογράφημα είναι το ιστόγραμμα με περιστροφή 90 μοιρών, δηλαδή επιδεικνύεται στη μία μεριά. Το μήκος κάθε γραμμής αντιστοιχεί στον αριθμό των παρατηρήσεων που ανήκουν στο διάστημα. Το φυλλογράφημα αναπαριστά κάθε τιμή με μία αληθινή τιμή.



# ΘΗΚΟΓΡΑΜΜΑ

- Στο θηκόγραμμα παριστάνονται η διάμεσος, το 25ο και 75ο ποσοστιαίο σημείο και οι ακραίες τιμές («αντιφατικές» τιμές σε σχέση με τις υπόλοιπες παρατηρούμενες τιμές του συνόλου δεδομένων).
- Το κάτω άκρο του κουτιού είναι το 25ο ποσοστιαίο σημείο και το πάνω άκρο το 75ο. Η διάμεσος παριστάνεται από μία οριζόντια γραμμή μέσα στο κέντρο του κουτιού.

# ΘΗΚΟΓΡΑΜΜΑ

- Η μικρότερη και μεγαλύτερη παρατηρούμενη τιμή σημειώνεται με μία οριζόντια γραμμή whiskers.
- Ο άνω και κάτω φράκτης  
1ο τεταρτημόριο – 1.5 \* ενδοτεταρτημοριακό εύρος  
και  
3ο τεταρτημόριο +1.5 \* ενδοτεταρτημοριακό εύρος

# ΥΛΟΠΟΙΗΣΗ ΓΙΑ ΤΗΝ ΑΠΟΚΤΗΣΗ ΠΕΡΙΓΡΑΦΙΚΩΝ ΜΕΤΡΩΝ ΠΟΣΟΤΙΚΗΣ ΜΕΤΑΒΛΗΤΗΣ

- DESCRIPTIVES
- FREQUENCIES
- EXPLORE

# ΕΞΕΤΑΣΗ ΤΗΣ ΣΧΕΣΗΣ ΔΥΟ ΜΕΤΑΒΛΗΤΩΝ ΔΥΟ ΠΟΙΟΤΙΚΕΣ

- ΠΙΝΑΚΕΣ ΣΥΝΑΦΕΙΑΣ
- ΕΞΕΤΑΣΗ ΤΗΣ ΥΠΑΡΞΗΣ Ή ΟΧΙ ΣΧΕΣΗΣ
- ΕΝΤΑΣΗ ΚΑΙ ΦΥΣΗ ΤΗΣ ΣΧΕΣΗΣ

# ΔΥΟ ΠΟΙΟΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΑ CROSSTABS

- **ΠΙΝΑΚΑΣ ΣΥΝΑΦΕΙΑΣ**

Ο πίνακας συνάφειας είναι ένας πίνακας διδιάστατος (στο επίπεδο) με  $r$  γραμμές, όσες οι κατηγορίες της μίας ποιοτικής μεταβλητής, και  $c$  στήλες όσες οι κατηγορίες της άλλης ποιοτικής μεταβλητής. Έτσι δημιουργούνται  $r \times c$  κελιά (κυψελίδες), κάθε ένα από τα οποία παριστάνει ένα συνδυασμό των τιμών των δύο μεταβλητών και στα οποία καταγράφονται οι παρατηρούμενες συχνότητες εμφάνισής τους.

# ΕΛΕΓΧΟΣ ΑΝΕΞΑΡΤΗΣΙΑΣ

- Επιτυγχάνεται με το στατιστικό

$$X^2 = \frac{\sum_{i=1}^r \sum_{j=1}^c (O_{ij} - E_{ij})^2}{E_{ij}}$$

όπου

$$E_{ij} = \frac{\sum_{i=1}^r O_{ij} \sum_{j=1}^c O_{ij}}{\sum_{i=1}^r \sum_{j=1}^c O_{ij}} = \frac{\sum_{i=1}^r O_{ij} \sum_{j=1}^c O_{ij}}{n}$$

με κρίσιμη περιοχή

$$X^2 \geq X_{a,(r-1)(c-1)}^2$$

# ΕΝΤΑΣΗ ΤΗΣ ΣΧΕΣΗΣ

- Ο **συντελεστής συνάφειας** (contingency coefficient).

$$C = \sqrt{\frac{X^2}{(X^2 + n)}}$$

- ο **συντελεστής Phi** (αναφέρεται και ως συντελεστής του Pearson)

$$\Phi = \sqrt{\frac{X^2}{n}}$$

- ο **συντελεστής V του Cramer**

$$V = \sqrt{\frac{X^2}{n \min(r-1, c-1)}}$$

# ΕΝΤΑΣΗ ΤΗΣ ΣΧΕΣΗΣ

- Ο συντελεστής **Lambda**, επίσης γνωστός και ως **Goodman-Kruskal lambda** και οι συντελεστές αβεβαιότητας (uncertainty coefficient) γνωστοί και ως Theil's U.
- Διατάξιμες: **Gamma** (ο zero-order για 2-way tables και ο conditional για 3-way έως 10-way tables), ο **Kendall's tau-b** (κατάλληλος για συμμετρικούς πίνακες), ο **Kendall's tau-c** (κατάλληλος για μη συμμετρικούς) και ο **Somers' d** (κατάλληλος για περιπτώσεις όπου η μία από τις δύο μεταβλητές μπορεί να θεωρηθεί εξαρτημένη, ενώ η άλλη ανεξάρτητη).



# ΑΛΛΑ ΚΡΙΤΗΡΙΑ ΚΑΙ ΜΕΤΡΑ ΣΥΝΑΦΕΙΑΣ

- Kappa του Kohen για πίνακες συνάφειας που έχουν τις ίδιες κατηγορίες στις στήλες και στις γραμμές. Παίρνει τιμές στο  $[-1, 1]$ .
- Το μέτρο Risk είναι διαθέσιμο για  $2 \times 2$  πίνακες, αλλά δε χρησιμοποιείται συχνά στην πράξη.
- McNemar χρησιμοποιείται για να ελέγξουμε τη μεταβολή στη συμπεριφορά των ατόμων πριν και μετά μία μέθοδο.
- McNemar-Bowker.
- Cochran's and Mantel-Haenszel statistics: ανεξαρτησία μεταξύ δύο δίτιμων μεταβλητών

# ΔΥΟ ΠΟΣΟΤΙΚΕΣ

- Η μελέτη της σχέσης δύο ποσοτικών μεταβλητών μπορεί να γίνει:
  - α) με σκοπό την τεκμηρίωση της σχέσης που έχουν
  - β) με σκοπό να καταλήξουμε σε μία μαθηματική σχέση που τις συνδέει (ΠΑΛΙΝΔΡΟΜΗΣΗ)
  - γ) με σκοπό τη σύγκριση των μέσων τιμών, όταν αυτές οι δύο μεταβλητές αποτελούν μετρήσεις πριν και μετά την εφαρμογή μίας μεθόδου (τεστ συγκρίσεως ζευγών).

# ΤΕΚΜΗΡΙΩΣΗ ΤΗΣ ΣΧΕΣΗΣ

- ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ

## Διαδικασία Scatterplot ή Chart Builder

Το διάγραμμα διασποράς δεν είναι τίποτε άλλο παρά το γράφημα των τιμών των δύο ποσοτικών μεταβλητών. Στον οριζόντιο άξονα τοποθετούνται οι τιμές εκείνης της μεταβλητής που ενδέχεται να έχει το ρόλο της εξαρτημένης, ενώ στον κατακόρυφο οι τιμές της εξαρτημένης. Η απεικόνιση αυτή μας βοηθά να έχουμε μία πρώτη υπόνοια για την ύπαρξη ή όχι κάποιας μαθηματικής σχέσης και ποιας μορφής (γραμμική, τετραγωνική, εκθετική κ.ο.κ.).

# ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

## Διαδικασία Correlate Bivariate

- Ο συντελεστής συσχέτισης του Pearson μας δίνει το βαθμό γραμμικής (και μόνο) εξάρτησης δύο ποσοτικών τυχαίων μεταβλητών

$$r(X, Y) = r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

- Η υπόθεση της μη ύπαρξης γραμμικής εξάρτησης ελέγχεται (υπό την προϋπόθεση ότι οι τ.μ.  $X$  και  $Y$  ακολουθούν κανονική κατανομή ή τουλάχιστον μία από αυτές) με το ΣΤΑΤΙΣΤΙΚΟ ΤΕΣΤ

$$t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}}$$

- Κρίσιμη περιοχή:

$$|t| \geq t_{\alpha/2, n-2}$$

# Μερικός συντελεστής συσχέτισης

## Διαδικασία Correlate Partial

- Με τον μερικό συντελεστή συσχέτισης αυτό που προσπαθούμε να κάνουμε είναι να διαπιστώσουμε αν η μεταβλητή ελέγχου είναι εκείνη που προκαλεί τη γραμμική σχέση μεταξύ των μεταβλητών μας.

# ΜΙΑ ΠΟΙΟΤΙΚΗ- ΜΙΑ ΠΟΣΟΤΙΚΗ

- Έστω ότι μας ενδιαφέρει να αναζητήσουμε τη σχέση μεταξύ μίας ποσοτικής μεταβλητής και μίας ποιοτικής μεταβλητής, με δύο ή περισσότερες κατηγορίες. Ουσιαστικά αυτό που θέλουμε να ελέγξουμε είναι αν οι πληθυσμιακές μέσες τιμές (της ποσοτικής μεταβλητής) δύο ή περισσότερων ομάδων (που καθορίζονται από την ποιοτική μεταβλητή) δε διαφέρουν στατιστικά σημαντικά. Στην περίπτωση που η ποιοτική μεταβλητή έχει δύο κατηγορίες τότε ο παραπάνω έλεγχος ανάγεται σε έλεγχο για τις μέσες τιμές, ενώ στην περίπτωση περισσότερων κατηγοριών ανάγεται σε πρόβλημα Ανάλυσης Διακύμανσης κατά ένα παράγοντα.