

Περιγραφικά Μέτρα– Μέτρα Θέσης

Δ. Κουλούμπού

Περιγραφικά Μέτρα

- Τα **στατιστικά περιγραφικά μέτρα** είναι αντιπροσωπευτικές τιμές οι οποίες περιγράφουν συγκεκριμένα χαρακτηριστικά σχετικά με την κατανομή μιας μεταβλητής.
- Λειτουργούν συμπληρωματικά με τους πίνακες και τα διαγράμματα για την περιγραφή αριθμητικών δεδομένων.

Περιγραφικά Μέτρα

Τα μέτρα αυτά διακρίνονται σε:

- **Μέτρα θέσης ή μέτρα κεντρικής τάσης**
- **Μέτρα διασποράς**
- **Μέτρα ασυμμετρίας και κύρτωσης**

Περιγραφικά Μέτρα

- **Μέτρα Θέσης** μιας κατανομής λέμε τα μέτρα (αριθμητικά μεγέθη) που μας δίνουν τη θέση του «κέντρου» των παρατηρήσεων στον οριζόντιο άξονα.
- Τα πιο συνηθισμένα μέτρα που χρησιμοποιούνται για να περιγράψουμε τη θέση ενός συνόλου δεδομένων πάνω στον οριζόντιο άξονα *Οχ* εκφράζοντας την «κατά μέσο όρο» απόστασή τους από την αρχή των αξόνων, είναι

Μέτρα Θέσης

Βασικά Μέτρα Θέσης:

- **Αριθμητικός Μέσος**
- **Διάμεσος**
- **Επικρατούσα Τιμή**

Αριθμητικός Μέσος ή Μέση Τιμή

- Το πιο γνωστό και διαδεδομένο μέτρο θέσης ή κεντρικής τάσης.
- Ο αριθμητικός μέσος για ένα σύνολο παρατηρήσεων x_1, x_2, \dots, x_n ορίζεται ως η ποσότητα

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Αριθμητικός Μέσος ή Μέση Τιμή

Παράδειγμα:

Παρακάτω δίνεται ο αριθμός των emails που έστειλε ένας γραμματέας κατά την διάρκεια της εργασίας του ανά ημέρα από ένα τυχαίο δείγμα 15 ημερών.

5,10,2,8,7,10,9,23,6,5,10,7,4,8,16

Η μέση τιμή των αριθμών emails είναι

$$\bar{x} = \frac{5 + 10 + 2 + 8 + 7 + 10 + 9 + 23 + 6 + 5 + 10 + 7 + 4 + 8 + 16}{15} = \frac{130}{15}$$
$$= 8,67$$

Πλεονεκτήματα και μειονεκτήματα του Αριθμητικού Μέσου

Παρακάτω παρουσιάζοντα συνοπτικά τα πλεονεκτήματα και τα μειονεκτήματα του Αριθμητικού Μέσου :

Πλεονεκτήματα:

- Είναι εύκολος στον υπολογισμό του.
- Αντικατοπτρίζει πιο πιστά την κεντρική τιμή της κατανομής σε σχέση με τους άλλους δείκτες.
- Μπορεί να χρησιμοποιηθεί για τον υπολογισμό των παραμέτρων του πληθυσμού (παραμετρικά τεστ).

Πλεονεκτήματα και μειονεκτήματα του Αριθμητικού Μέσου

Μειονεκτήματα:

- Είναι ευαίσθητος στις τιμές των δεδομένων της κατανομής. Για παράδειγμα, μια αλλαγή σε οποιαδήποτε τιμή προκαλεί διαφοροποίηση του μέσου όρου.
- Επειδή υπολογίζεται αλγεβρικά, η τιμή του είναι πιθανό να μην ανήκει στις τιμές της κατανομής.
- Είναι πολύ ευαίσθητος στις ακραίες τιμές.

Διάμεσος

Η διάμεσος (median) δενός δείγματος είναι η τιμή που χωρίζει το δείγμα ή τον πληθυσμό σε δύο ίσα μέρη, δηλαδή:

- Ο αριθμός των παρατηρήσεων που είναι μικρότερες ή ίσες από το δ να είναι ίσος με τον αριθμό των παρατηρήσεων που είναι μεγαλύτερες ή ίσες από το δ.
- Με άλλα λόγια, η διάμεσος (δ) χωρίζει την κατανομή του πληθυσμού ακριβώς στη μέση.

Υπολογισμός Διαμέσου

Η διάμεσος ενός συνόλου στοιχείων είναι η τιμή στο μέσον αυτού του συνόλου αφού τα δεδομένα ταξινομηθούν με αύξουσα σειρά.

- Αν ο αριθμός των παρατηρήσεων είναι περιττός, η διάμεσος είναι η τιμή της παρατήρησης που βρίσκεται ακριβώς στη μέση.
- Αν ο αριθμός των παρατηρήσεων είναι άρτιος, η διάμεσος είναι ο μέσος όρος των δύο μεσαίων παρατηρήσεων.

Υπολογισμός Διαμέσου

Δηλαδή:

Αν x_1, x_2, \dots, x_n οι τιμές των παρατηρήσεων μιας μεταβλητής ταξινομημένες σε αύξουσα σειρά, τότε:

- $\delta = x_{\frac{n+1}{2}}$, αν n περιττός
- $\delta = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$, αν n άρτιος

Υπολογισμός Διαμέσου

Παράδειγμα 1:

Η βαθμολογία στα εννέα μαθήματα προόδου ενός πρωτοετούς φοιτητή, έχει ως εξής: 8, 10, 7, 5, 6, 6, 7, 8, 9. Για να βρούμε την διάμεσο των βαθμών του φοιτητή θα πρέπει αρχικά να τις **διατάξουμε σε αύξουσα σειρά**.

Επομένως διατάσσουμε: 5, 6, 6, 7, 7, 8, 8, 9, 10.

Επειδή και εδώ, το πλήθος των παρατηρήσεων-βαθμών είναι περιττός αριθμός, αφού $n = 9$ η διάμεσος θα είναι ίση με τη μεσαία παρατήρηση, εφόσον προηγουμένως τις έχουμε διατάξει σε αύξουσα σειρά. Δηλαδή ισχύει ότι

$$\delta = \underline{x_{\frac{n+1}{2}}} = x_5 = 7$$

(Η πέμπτη παρατήρηση , δηλαδή το 7 είναι η διάμεσος του δείγματος).

Υπολογισμός Διαμέσου

Παράδειγμα 2:

Η βαθμολογία του πρώτου τριμήνου ενός μαθητή στα δέκα γραπτώς εξεταζόμενα μαθήματα, έχει ως εξής: 12, 15, 11, 14, 16, 15, 13, 15, 13, 16.

Για να βρούμε την διάμεσο των βαθμών του φοιτητή θα πρέπει αρχικά να τις **διατάξουμε σε αύξουσα σειρά**.

Τις διατάσσουμε και έχουμε: 11, 12, 13, 13, 14, 15, 15, 15, 16, 16.

Επειδή και εδώ, το πλήθος των παρατηρήσεων-βαθμών είναι άρτιος αριθμός, αφού $n = 10$ η διάμεσος θα είναι ίση με το ημιάθροισμα των δύο μεσαίων παρατηρήσεων, εφόσον προηγουμένως τις έχουμε διατάξει σε αύξουσα σειρά. Δηλαδή ισχύει ότι:

$$\delta = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_5 + x_6}{2} = \frac{14 + 15}{2} = 14,5$$

Πλεονεκτήματα και μειονεκτήματα της Διαμέσου

Παρακάτω παρουσιάζοντα συνοπτικά τα πλεονεκτήματα και τα μειονεκτήματα της διαμέσου:

Πλεονεκτήματα:

- Είναι εύκολο να υπολογιστεί
- Δεν επηρεάζεται από τις ακραίες τιμές, οπότε είναι καλύτερος δείκτης κεντρικής τάσης όταν έχουμε ασύμμετρη κατανομή. (Σε συμμετρικές κατανομές η διάμεσος συμπίπτει με την μέση τιμή)
- Μπορεί να υπολογιστεί ακόμη και όταν δεν γνωρίζουμε τις ακραίες τιμές.

Πλεονεκτήματα και μειονεκτήματα της Διάμεσου

Μειονεκτήματα:

- Δεν λαμβάνει υπόψη την ακριβή τιμή του κάθε στοιχείου.
- Αν οι τιμές της κατανομής είναι λίγες, τότε η διάμεσος δεν μπορεί να τις αντιπροσωπεύσει με ακρίβεια. Για παράδειγμα, αν έχουμε τις τιμές 2, 5, 8, 67 και 110, η διάμεσος είναι το 8.

Επικρατούσα Τιμή

- Η επικρατούσα τιμή (T_0) (mode) είναι η τιμή που συναντάται τις περισσότερες φορές σε μια κατανομή, δηλαδή η τιμή με τη μεγαλύτερη συχνότητα.
- Είναι το απλούστερο στατιστικό μέτρο κεντρικής τάσης και συνήθως χρησιμοποιείται για την περιγραφή κατηγορικών (ποιοτικών) μεταβλητών.
- Από όλα τα είδη στατιστικών μέτρων κεντρικής τάσης, η επικρατούσα τιμή είναι η ευκολότερα προσδιοριζόμενη, αφού για τον υπολογισμό της δεν χρειάζονται αλγεβρικοί υπολογισμοί αλλά μια απλή καταμέτρηση των δεδομένων.
- Στη γραφική αναπαράσταση μιας κατανομής (π.χ., ιστόγραμμα συχνοτήτων) η επικρατούσα τιμή, επειδή αντιπροσωπεύει την πιο συχνή τιμή, αντιστοιχεί στην κορυφή της κατανομής.

Επικρατούσα Τιμή

Η Επικρατούσα Τιμή (T_0) ενός δείγματος είναι η τιμή που εμφανίζεται με την μεγαλύτερη συχνότητα από όλες τις άλλες.

- Π.χ. Στις διπλανές παρατηρήσεις η τιμή 450 εμφανίζεται πιο συχνά (7 φορές).

Άρα Επικρατούσα Τιμή : $T_0 = 450$

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 425 | 430 | 430 | 435 | 435 | 435 | 435 | 435 | 440 | 440 |
| 440 | 440 | 440 | 445 | 445 | 445 | 445 | 445 | 450 | 450 |
| 450 | 450 | 450 | 450 | 450 | 460 | 460 | 460 | 465 | 465 |
| 465 | 470 | 470 | 472 | 475 | 475 | 475 | 480 | 480 | 480 |
| 480 | 485 | 490 | 490 | 490 | 500 | 500 | 500 | 500 | 510 |
| 510 | 515 | 525 | 525 | 525 | 535 | 549 | 550 | 570 | 570 |
| 575 | 575 | 580 | 590 | 600 | 600 | 600 | 600 | 615 | 615 |

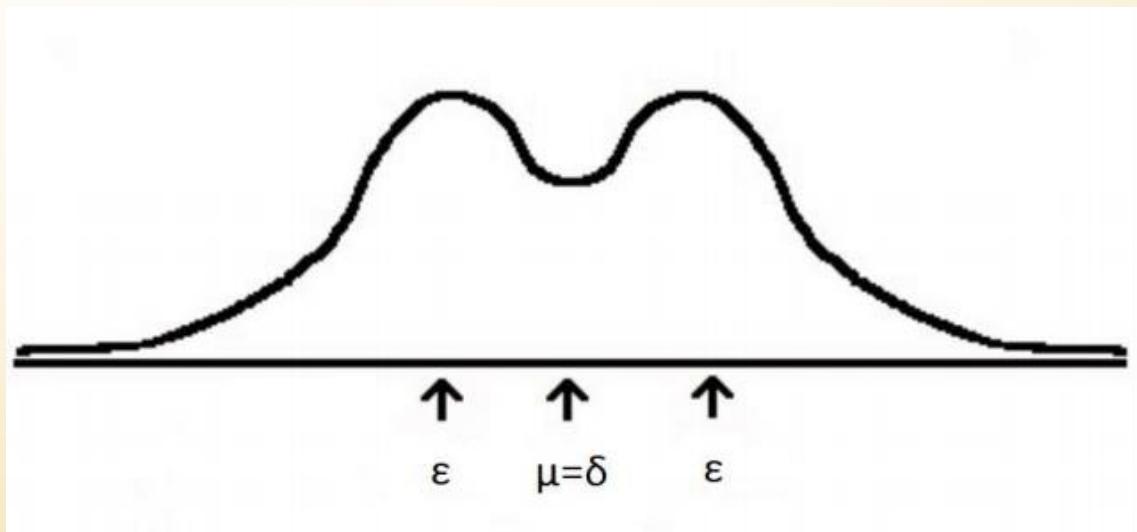
Επικρατούσα Τιμή

Στην περίπτωση που υπάρχουν δύο διαφορετικές τιμές με την ίδια μέγιστη συχνότητα, τότε επικρατούσες τιμές είναι και οι δύο, και η κατανομή αυτή ονομάζεται **δικόρυφη** (bimodal), ενώ όταν υπάρχουν περισσότερες από δύο, τότε η κατανομή λέγεται **πολυκόρυφη** (multimodal). Για παράδειγμα,

- αν είχαμε μια κατανομή όπου οι τιμές 5 και 9 είχαν τη μεγαλύτερη συχνότητα (εμφανίζονταν από 7 φορές η κάθε μία), τότε η εν λόγω κατανομή θα είχε δύο επικρατούσες τιμές.
- Τέλος, ενδέχεται να μην υπάρχει καμία επικρατούσα τιμή. Για να συμβεί αυτό αρκεί η κατανομή να είναι **ισοϋψής**.

Επικρατούσα Τιμή

Παράδειγμα δικόρυφης κατανομής:



Πλεονεκτήματα και μειονεκτήματα της Επικρατούσας Τιμής

Παρακάτω παρουσιάζοντα συνοπτικά τα πλεονεκτήματα και τα μειονεκτήματα της Επικρατούσας Τιμής:

Πλεονεκτήματα:

- Δείχνει την πιο συχνή τιμή της κατανομής.
- Μένει ανεπηρέαστη από ακραίες τιμές.
- Μπορεί να υπολογιστεί όταν οι ακραίες τιμές είναι άγνωστες.

Πλεονεκτήματα και μειονεκτήματα της Επικρατούσας Τιμής

Μειονεκτήματα:

- Δεν λαμβάνει υπόψη την ακριβή τιμή του κάθε στοιχείου.
- Δεν είναι πολύ χρήσιμη για μικρό αριθμό
- δεδομένων, στα οποία αρκετές τιμές εμφανίζονται το ίδιο συχνά (1, 1, 2, 3, 3, 4).

Υπολογισμός των στατιστικών μέτρων θέσης στο SPSS

Για τον υπολογισμό των τριών δεικτών κεντρικής τάσης στο SPSS θα χρησιμοποιήσουμε τα δεδομένα τα οποία περιέχονται στο αρχείο lesson3_datafile.sav

Τα δεδομένα αφορούν τις βαθμολογίες 50 φοιτητών στα μαθήματα Στατιστική, Λογιστική και πληροφορική. Περιέχεται επίσης η πληροφορία του φύλου του κάθε φοιτητή καθώς και το έτος σπουδών του.

Υπολογισμός των στατιστικών μέτρων θέσης στο SPSS

Για να υπολογίσουμε τα μέτρα θέσης για την βαθμολογία του κάθε μαθήματος η διαδικασία είναι η εξής:

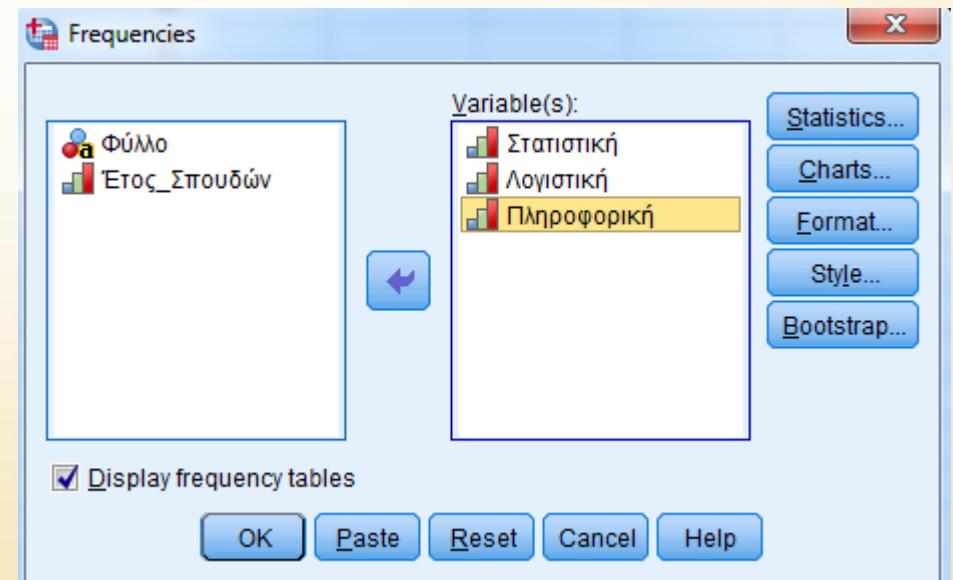
Από το αρχικό μενού που βρίσκεται στην κορυφή της οθόνης επιλέξτε **[Analyze> Descriptive> Statistics Frequencies...]**. Μετακινήστε τις μεταβλητές «Στατιστική» , «Λογιστική» και «Πληροφορική» στο πλαίσιο **[Variable(s):]**

(βλ. Εικόνα 1).

Υπολογισμός των στατιστικών μέτρων θέσης στο SPSS

Πλαίσιο διαλόγου Frequencies

- Εικόνα 1:



Υπολογισμός των στατιστικών μέτρων θέσης στο SPSS

Στη συνέχεια, πατήστε το κουμπί **[Statistics...]** και επιλέξτε :

- **[Mean]** για μέση τιμή
- **[Median]** για διάμεσο και
- **[Mode]** για επικρατούσα τιμή.
- **[Quartiles > Cut points for > 4]** για τα τεταρτημόρια.

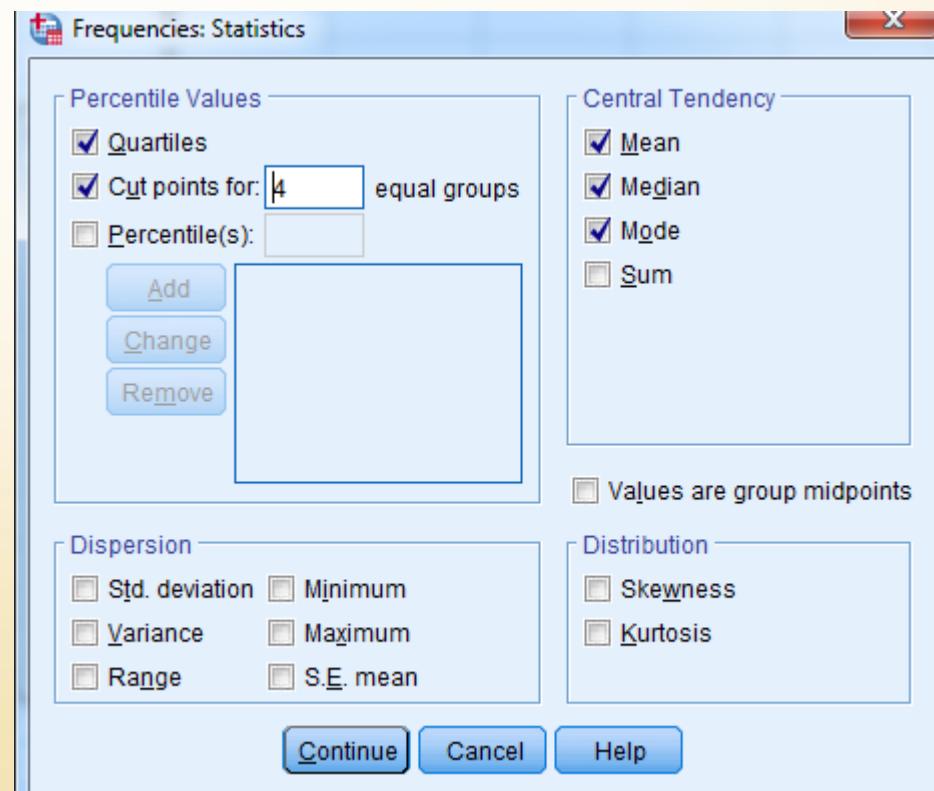
Τέλος, πατήστε **[Continue]** και μετά **[OK]**.

(βλ. Εικόνα 2)

Υπολογισμός των στατιστικών μέτρων θέσης στο SPSS

Πλαίσιο διαλόγου *Frequencies:*
Statistics

- *Εικόνα 2:*



Υπολογισμός των στατιστικών μέτρων θέσης στο SPSS

Στην Εικόνα 3 παρουσιάζονται τα αποτελέσματα του υπολογισμού, όπως εμφανίζονται στο SPSS Viewer. Ο πίνακας περιέχει τους τρεις δείκτες κεντρικής τάσης:

- **Mean** (ο αριθμητικό μέσος),
- **Median** (η διάμεσος)
- **Mode** (η επικρατούσα τιμή) καθώς και τα
- **Percentiles** (Q_1 , Q_2 , Q_3) τρία τεταρτημόρια για κάθε μία από τις τρεις μεταβλητές που επιλέξαμε.

Υπολογισμός των στατιστικών μέτρων θέσης στο SPSS

- Εικόνα 3:

| Statistics | | | | |
|-------------|---------|------------|-----------|-------------|
| | | Στατιστική | Λογιστική | Πληροφορική |
| N | Valid | 50 | 50 | 50 |
| | Missing | 0 | 0 | 0 |
| Mean | | 6,64 | 5,84 | 6,10 |
| Median | | 7,00 | 6,00 | 6,00 |
| Mode | | 6 | 5 | 6 |
| Percentiles | 25 | 5,75 | 5,00 | 4,75 |
| | 50 | 7,00 | 6,00 | 6,00 |
| | 75 | 8,00 | 7,00 | 8,00 |

Διαδικασία Explore (SPSS)

Η διαδικασία Explore μπορεί να χρησιμοποιηθεί για την απόκτηση πλήθους στατιστικών μέτρων καθώς και γραφικών παραστάσεων τόσο για το σύνολο των δεδομένων όσο και ξεχωριστά για κατηγορίες αυτών.

Ας υποθέσουμε για παράδειγμα ότι θέλουμε να μελετήσουμε την βαθμολογία των μαθημάτων «Στατιστική» και «Πληροφορική» ως προς το φύλο των φοιτητών.

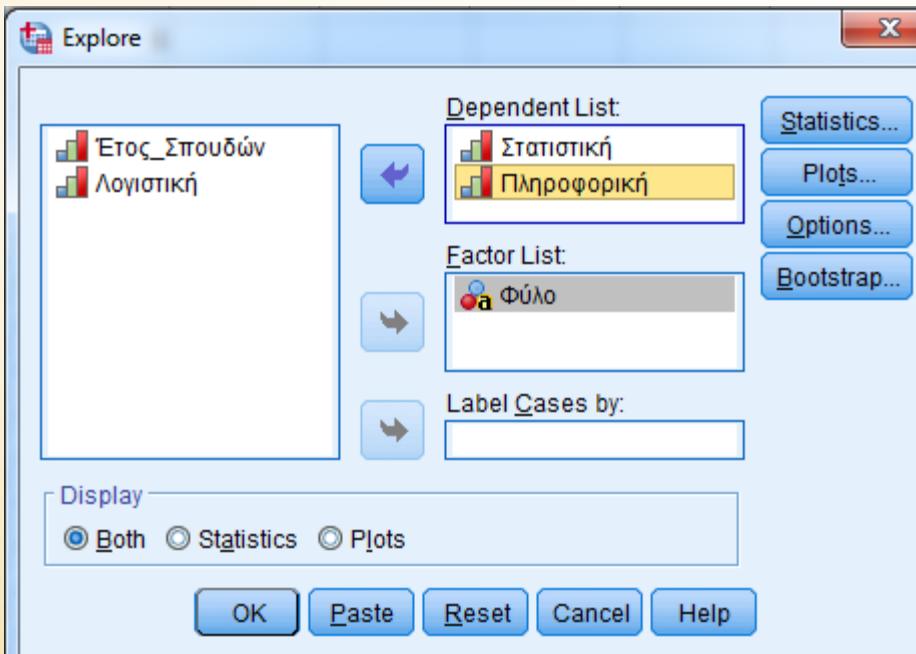
Διαδικασία Explore (SPSS)

- Από το κεντρικό παράθυρο διαλόγου επιλέγουμε: [Analyze >Descriptive Statistics>Explore].
- Στο παράθυρο διαλόγου που προκύπτει τοποθετούμε στο πλαίσιο **Dependent** τις ποσοτικές (υποχρεωτικά και μόνο) μεταβλητές που θέλουμε να αναλύσουμε. (Στο παράδειγμα μας «Στατιστική» και «Πληροφορική»)
- Στο πλαίσιο **Factor list** τοποθετούμε τις πιθανές ποιοτικές-κατηγορικές μεταβλητές (και μόνο) ως προς τις κατηγορίες των οποίων θέλουμε να προχωρήσουμε την ανάλυση μας, π.χ. έτος σπουδών, φύλο κ.ο.κ. (Στο παράδειγμα μας «Φύλο»)

(βλ. Εικόνα 4)

Διαδικασία Explore (SPSS)

Εικόνα 4:



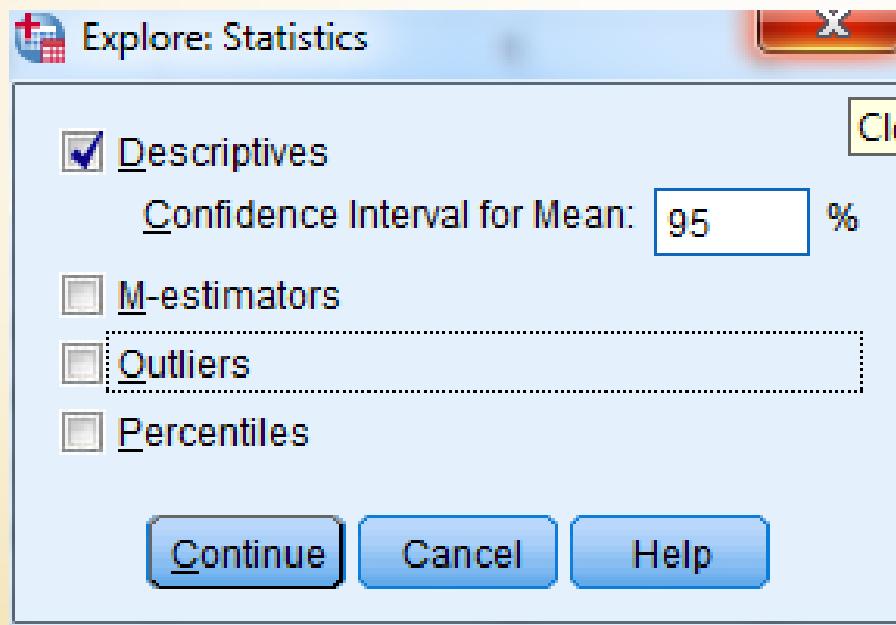
Διαδικασία Explore (SPSS)

Επιπρόσθετα διατηρώντας την προεπιλογή Display Both (στο κάτω αριστερό άκρο του παραθύρου) έχουμε τη δυνατότητα απόκτησης τόσο στατιστικών μέτρων όσο και γραφημάτων.

- Από την επιλογή Statistics επιλέγουμε **Descriptives**, για απόκτηση των κυριότερων περιγραφικών μέτρων, όπως η διάμεσος, η μέση τιμή, η τυπική απόκλιση κ.ά. (εικόνα 5) και στην συνέχεια πατάμε **Continue**.

Διαδικασία Explore (SPSS)

Εικόνα 5:

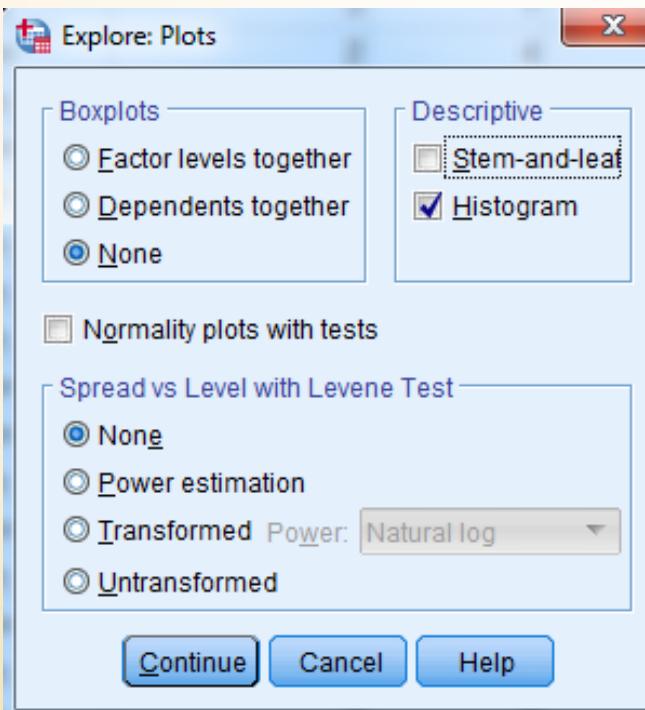


Διαδικασία Explore (SPSS)

Από την επιλογή Plots για την ώρα στο Boxplots επιλέγουμε none, στις επιλογές του Spread vs Level with Levene Test επιλέγουμε none, και στην επιλογή descriptive επιλέγουμε Histogram (βλ. Εικόνα 6) και στην συνέχεια πατάμε **Continue** και μετά **ok**

Διαδικασία Explore (SPSS)

Εικόνα 6:



Ερμηνεία Αποτελεσμάτων

Ο παρακάτω πίνακας που εμφανίζεται στο SPSS Viewer μας πληροφορεί ότι από τους 50 συμμετέχοντες 30 ήταν κορίτσια (φύλο 1) και 20 αγόρια (φύλο 2) σε κάθε μάθημα χωρίς να υπάρχουν ελλιπείς τιμές.

| Φύλο | Case Processing Summary | | | | | | |
|-------------|-------------------------|---------|---------|---------|-------|---------|--------|
| | Cases | | Missing | | Total | | |
| | N | Percent | N | Percent | N | Percent | |
| Στατιστική | 1 | 30 | 100,0% | 0 | 0,0% | 30 | 100,0% |
| | 2 | 20 | 100,0% | 0 | 0,0% | 20 | 100,0% |
| Πληροφορική | 1 | 30 | 100,0% | 0 | 0,0% | 30 | 100,0% |
| | 2 | 20 | 100,0% | 0 | 0,0% | 20 | 100,0% |

Ερμηνεία Αποτελεσμάτων

Στον πίνακα **Descriptives** μας δίνονται διάφορα περιγραφικά μέτρα (και όχι μόνο) για τη μεταβλητές «Στατιστική» και «Πληροφορική» που περιγράφει την βαθμολογία σε αυτά τα μαθήματα των φοιτητών ανά φύλο.

Ερμηνεία Αποτελεσμάτων

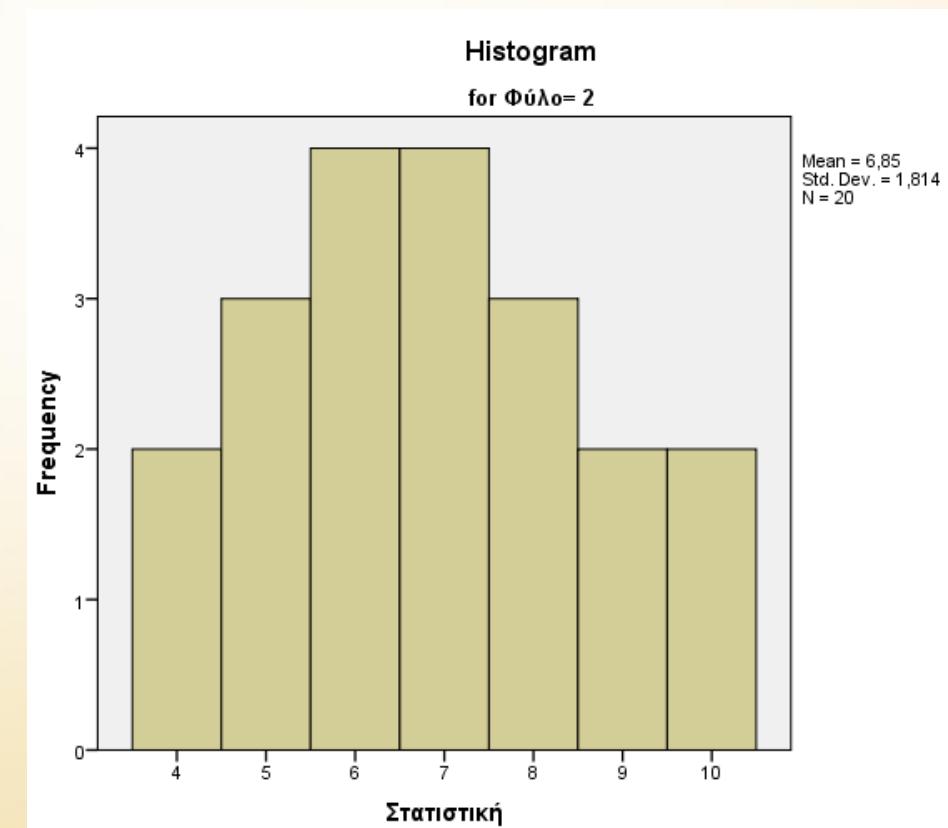
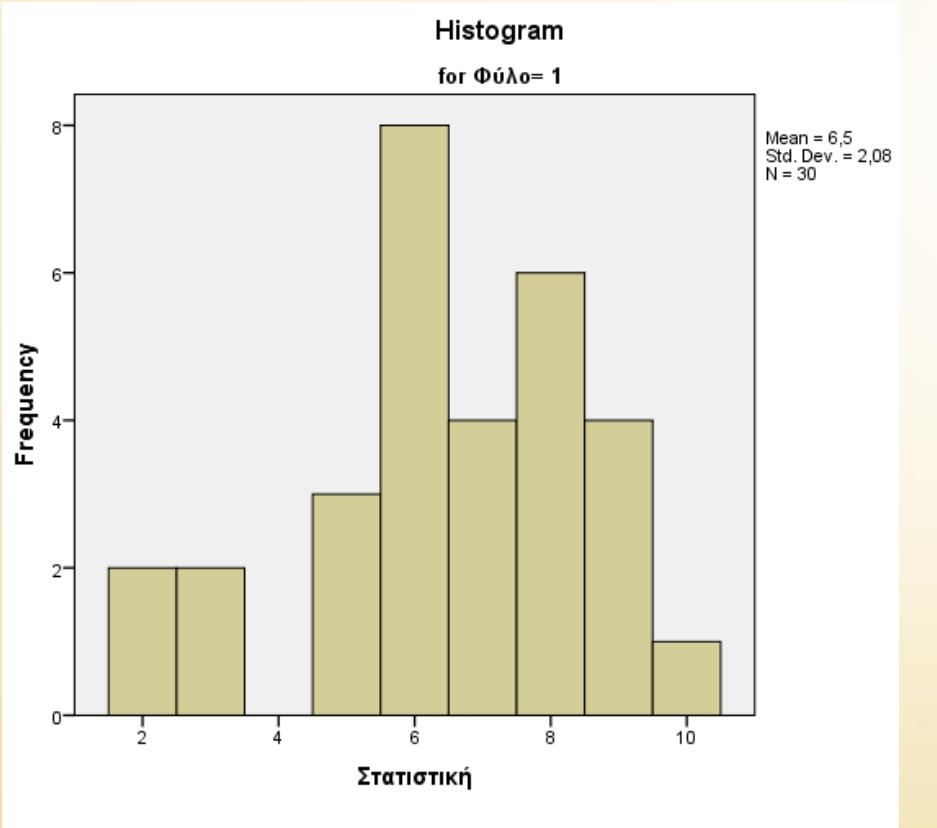
| Descriptives | | | | |
|--|------|----------------------------------|----------------------------|--------------|
| | Φύλο | Statistic | Std. Error | |
| Στατιστική | 1 | Mean | 6,50 | ,380 |
| | | 95% Confidence Interval for Mean | Lower Bound Upper Bound | 5,72 7,28 |
| | | 5% Trimmed Mean | | 6,57 |
| | | Median | | 6,50 |
| | | Variance | | 4,328 |
| | | Std. Deviation | | 2,080 |
| | | Minimum | | 2 |
| | | Maximum | | 10 |
| | | Range | | 8 |
| | | Interquartile Range | | 2 |
| | | Skewness | | -,628 ,427 |
| | | Kurtosis | | -,001 ,833 |
| | | Mean | | 6,85 ,406 |
| | | 95% Confidence Interval for Mean | Lower Bound Upper Bound | 6,00 7,70 |
| ΠΑΔΑ - ΣΔΟΚΕ ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ ΚΑΙ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ | 2 | 5% Trimmed Mean | | 6,83 |
| | | Median | | 7,00 |
| | | Variance | | 3,292 |
| | | Std. Deviation | | 1,814 |
| | | Minimum | | 4 |
| | | Maximum | | 10 |
| | | Range | | 6 |
| | | Interquartile Range | | 3 |
| | | Skewness | | -,188 ,512 |
| | | Kurtosis | | -,779 ,992 |

| | | | | |
|-------------|---|----------------------------------|----------------------------|--------------|
| Πληροφορική | 1 | Mean | 6,37 | ,400 |
| | | 95% Confidence Interval for Mean | Lower Bound Upper Bound | 5,55 7,18 |
| | | 5% Trimmed Mean | | 6,43 |
| | | Median | | 6,50 |
| | | Variance | | 4,792 |
| | | Std. Deviation | | 2,189 |
| | | Minimum | | 2 |
| | | Maximum | | 10 |
| | | Range | | 8 |
| | | Interqua Skewne | Double-click to activate | 3 |
| | | Kurtosis | | -,564 ,833 |
| | 2 | Mean | 5,70 | ,465 |
| | | 95% Confidence Interval for Mean | Lower Bound Upper Bound | 4,73 6,67 |
| | | 5% Trimmed Mean | | 5,72 |
| | | Median | | 6,00 |
| | | Variance | | 4,326 |
| | | Std. Deviation | | 2,080 |
| | | Minimum | | 1 |
| | | Maximum | | 10 |
| | | Range | | 9 |
| | | Interquartile Range | | 3 |
| | | Skewness | | -,140 ,512 |
| | | Kurtosis | | 358 ,992 |

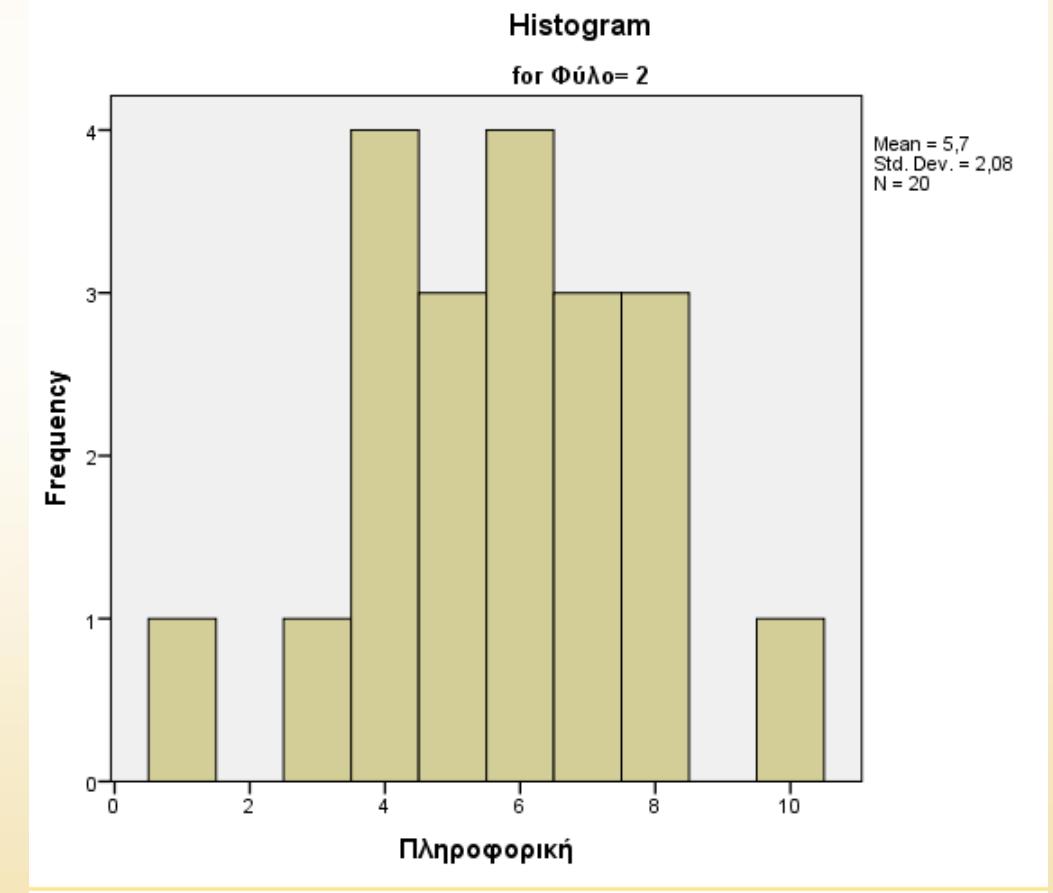
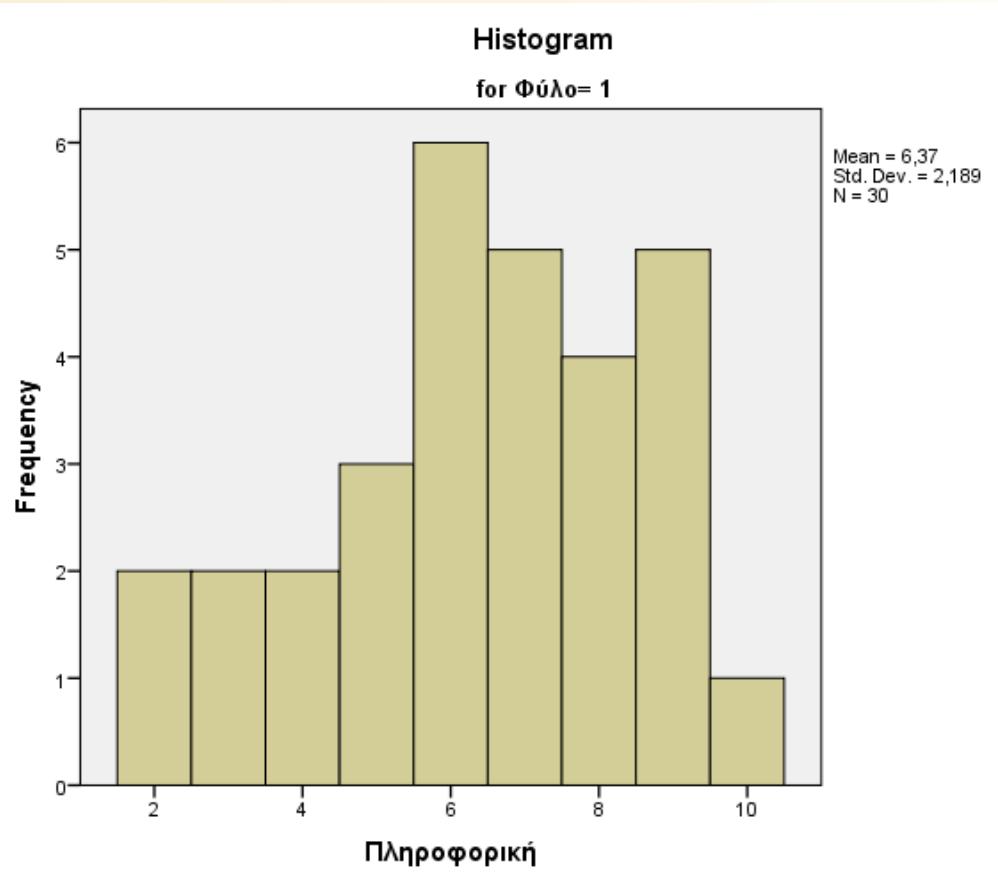
Γραφήματα

Επιπλέον έχουμε τα ιστόγραμμα των μεταβλητών «Στατιστική» και «Πληροφορική» ως προς το φύλο.

Γραφήματα



Γραφήματα



Βιβλιογραφία

**Εισαγωγική Διαχείριση
του IBM SPSS, Χαλικιάς
Μιλτιάδης, Λάλου
Παναγιώτα, Μανωλέσου
Αλεξάνδρα. Εκδόσεις
Κάλλιπος**

**Data Analysis in
Management with SPSS
Software , J.P. Verma ,
Springer , 2012**