

ΕΘΝΙΚΟ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ - ΙΑΤΡΙΚΗ ΣΧΟΛΗ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

«ΕΦΑΡΜΟΣΜΕΝΗ ΝΕΥΡΟΑΝΑΤΟΜΙΑ»

«Βιοστατιστική, Μεθοδολογία και Συγγραφή Επιστημονικής Μελέτης»

Ενότητα 1: Πληθυσμός και δείγμα – Είδη Μεταβλητών -
Περιγραφική στατιστική

Δρ.Ευσταθία Παπαγεωργίου, Αναπληρώτρια Καθηγήτρια
Τμήμα Ιατρικών Εργαστηρίων

Συλλογή δεδομένων

- Μια κύρια πρωτογενής πηγή συλλογής στοιχείων είναι η διενέργεια πληθυσμιακών ερευνών
 - *Π.χ. απογραφικών ή δειγματοληπτικών.*

Μέθοδοι συλλογής δεδομένων 1/2

- Οι κυριότεροι τρόποι συλλογής των πληροφοριών είναι οι ακόλουθοι:
 - **Ταχυδρομικά**, η μέθοδος αυτή έχει μικρό κόστος, αλλά απαντά συνήθως πολύ μικρό ποσοστό ερωτώμενων (20-40%).
 - **Τηλεφωνικά**, έχει και αυτή η μέθοδος μικρό κόστος και εύκολο τρόπο επιλογής τυχαίου δείγματος, π.χ. τυχαίοι αριθμοί από τηλεφωνικό κατάλογο, όμως πρέπει να γίνεται προσεκτικά και σε ώρες που να μην είναι ενοχλητικές για τους ερωτώμενους. Συνήθως ανταποκρίνεται το 60-80% των ερωτώμενων, αλλά πολλές φορές οι απαντήσεις είναι μικρής εγκυρότητας.

Μέθοδοι συλλογής δεδομένων 2/2

- **Συνέντευξη**, ίσως ο πιο ασφαλής τρόπος αλλά με μεγάλο κόστος. Οι ερευνητές επίσης πρέπει να είναι κατάλληλα καταρτισμένοι για την επιτυχή διεξαγωγή της έρευνας. Συνήθως ανταποκρίνεται το 70-80% των ερωτώμενων.
- **Συνδυασμός των παραπάνω μεθόδων**, η μέθοδος αυτή είναι δαπανηρή αλλά δίνει τα καλύτερα αποτελέσματα και με την μικρότερη πιθανότητα σφάλματος.

Η απογραφή

- Η μέθοδος της απογραφής στηρίζεται στη συγκέντρωση στοιχείων από όλες τις μονάδες ενός πληθυσμού που ενδιαφέρει στη συγκεκριμένη έρευνα.
 - Επειδή το σύνολο των ερωτώμενων είναι μεγάλο, παράλληλα με το μεγάλο πλήθος των μεταβλητών που εξετάζονται, η μέθοδος αυτή απαιτεί μεγάλο χρόνο και πολύ μεγάλο κόστος για την συλλογή των στοιχείων και την εξαγωγή συμπερασμάτων.
 - ✓ Έτσι συχνά οι ερευνητές καταφεύγουν στη διαδικασία της δειγματοληψίας, που κρύβει και μεγαλύτερο σφάλμα.

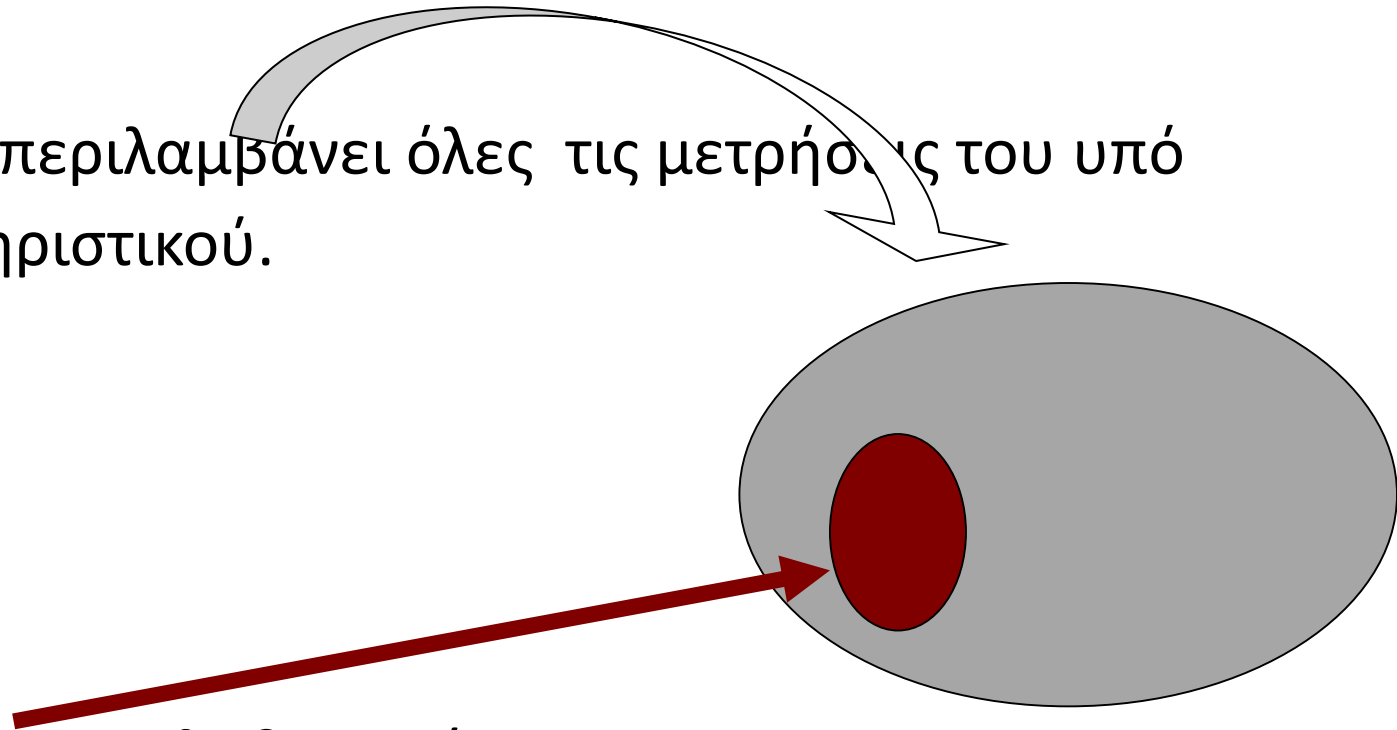
Σημαντικοί στατιστικοί όροι 1/2

Πληθυσμός:

Το σύνολο που περιλαμβάνει όλες τις μετρήσεις του υπό μελέτη χαρακτηριστικού.

Δείγμα:

Ένα υποσύνολο του πληθυσμού.



Σημαντικοί στατιστικοί όροι 2/2

- **Πληθυσμός Στόχος:**

Ο υπό μελέτη πληθυσμός/ στον οποίο ο ερευνητής θέλει να γενικεύσει τα αποτελέσματά του.

- **Απλό στοιχείο:**

Κάθε στοιχείο του πληθυσμού.

- **Δειγματοληπτική μονάδα:**

Η μικρότερη μονάδα από την οποία μπορεί να επιλεγεί το δείγμα.

- **Δειγματοληπτικό πλαίσιο**

Το σύνολο των δειγματοληπτικών μονάδων.

- **Τρόπος Δειγματοληψίας**

Μέθοδος επιλογής δείγματος από το δειγματοληπτικό πλαίσιο.

Για ποιο λόγο παίρνουμε δείγμα;

- Όταν θέλουμε να εξάγουμε πληροφορία για μεγάλους πληθυσμούς
 - ↓ Μείωση κόστους
 - ↓ Ελαχιστοποίηση χρόνου
 - ↘ Όταν είναι αδύνατο να μελετήσουμε ολόκληρο τον πληθυσμό

Τύποι δειγματοληψίας

- Μη πιθανοθεωρητικά δείγματα
- Πιθανοθεωρητικά δείγματα

Δειγματοληψία 1/2

Όμως αν δεν πραγματοποιηθεί **με σωστές μεθόδους** τότε υπάρχει μεγάλος κίνδυνος εξαγωγής λανθασμένων συμπερασμάτων.

- ✓ Ένα χαρακτηριστικό παράδειγμα λανθασμένης εκλογής δείγματος ήταν η Αμερικάνικες προεδρικές εκλογές του 1936 όπου η πρόβλεψη οδήγησε σε πλήρη αποτυχία γιατί το δείγμα επιλέχθηκε από τους τηλεφωνικούς καταλόγους οι οποίοι την εποχή εκείνη περιείχαν μόνο τις “ανώτερες” κοινωνικές τάξεις που διέθεταν τηλέφωνο.

Δειγματοληψία 2/2

- Από τις πιο συχνά εφαρμόσιμες μεθόδους συλλογής δεδομένων είναι η **δειγματοληψία**, η οποία ορίζεται ως η μερική απογραφή σε ένα υποσύνολο του πληθυσμού το δείγμα.

Μη πιθανοθεωρητικά δείγματα 1/2

- Βολικά δείγματα (ευκολία πρόσβασης)
- Δειγματοληψία χιονοστοιβάδας (το φίλο του φίλου κλπ)
- Σκόπιμη δειγματοληψία (Επιλέγουμε ότι νομίζουμε κατάλληλο για την έρευνα)
 - Δειγματοληψία ποσόστωσης

Μη πιθανοθεωρητικά δείγματα 2/2

- Η πιθανότητα επιλογής παραμένει άγνωστη.
- Είναι φθηνότερα αλλά αδυνατούν να δώσουν γενικευμένα συμπεράσματα.
- Δυνητικά μεροληπτικά.

Πιθανοθεωρητικά δείγματα

- Τυχαία δειγματοληψία
 - Κάθε μονάδα έχει μια γνωστή πιθανότητα επιλογής.
- Επιτρέπουν:
 - Γενικεύσεις.
 - Δοκιμασία υποθέσεων.

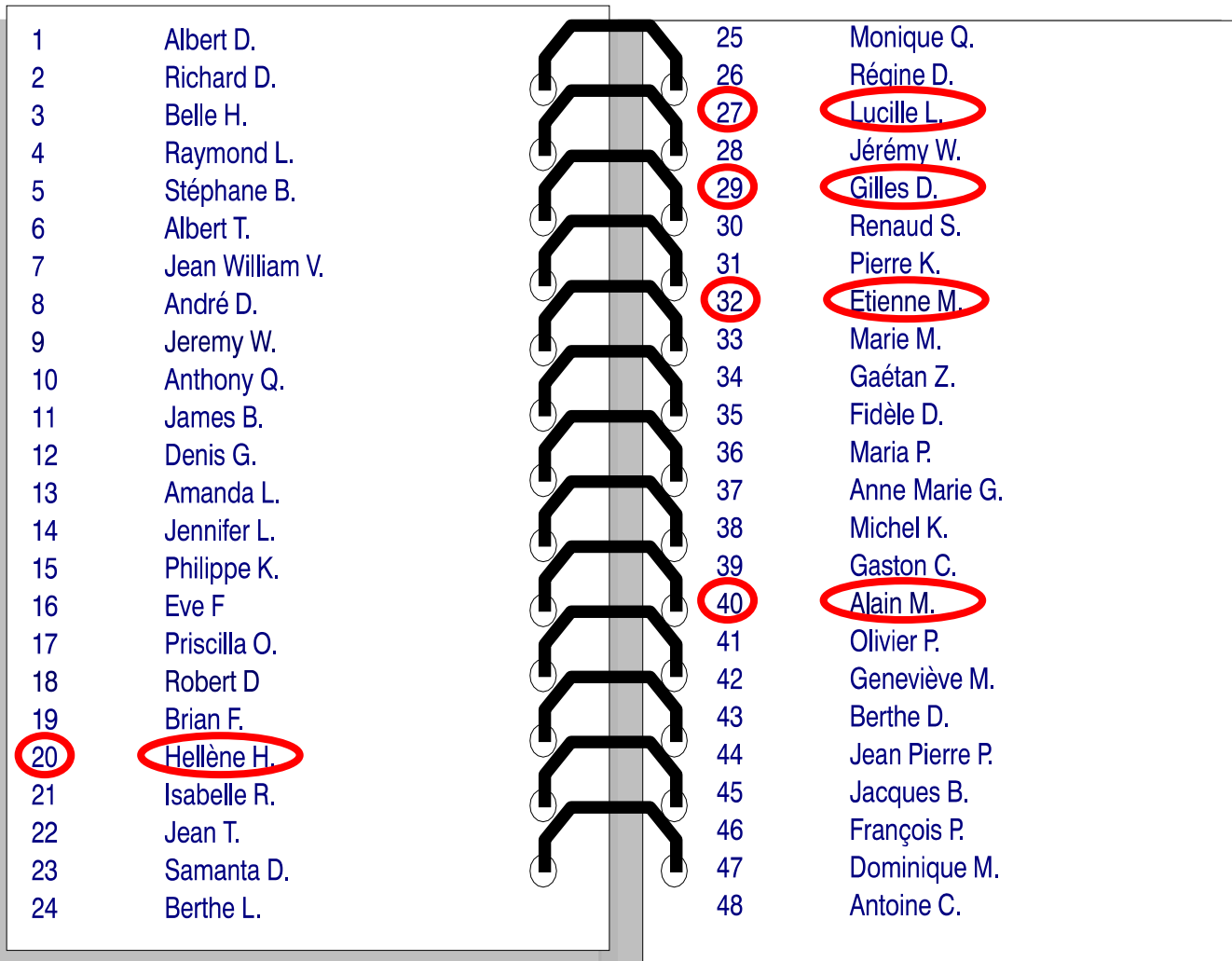
Συμπερασματικά

- Τα πιθανοθεωρητικά δείγματα είναι τα καλύτερα
- Εξασφαλίζουν:
 - Αντιπροσωπευτικότητα
 - Ακρίβεια

Μέθοδοι Δειγματοληψίας (πιθανοθεωρητικής)

- Απλή τυχαία δειγματοληψία
- Συστηματική δειγματοληψία
- Στρωματοποιημένη δειγματοληψία
- Πολυσταδιακή δειγματοληψία

Simple random sampling



1	Albert D.	25	Monique Q.
2	Richard D.	26	Réçine D.
3	Belle H.	27	Lucille L.
4	Raymond L.	28	Jérémy W.
5	Stéphane B.	29	Gilles D.
6	Albert T.	30	Renaud S.
7	Jean William V.	31	Pierre K.
8	André D.	32	Etienne M.
9	Jeremy W.	33	Marie M.
10	Anthony Q.	34	Gaétan Z.
11	James B.	35	Fidèle D.
12	Denis G.	36	Maria P.
13	Amanda L.	37	Anne Marie G.
14	Jennifer L.	38	Michel K.
15	Philippe K.	39	Gaston C.
16	Eve F	40	Alain M.
17	Priscilla O.	41	Olivier P.
18	Robert D	42	Geneviève M.
19	Brian F.	43	Berthe D.
20	Hellène H.	44	Jean Pierre P.
21	Isabelle R.	45	Jacques B.
22	Jean T.	46	François P.
23	Samanta D.	47	Dominique M.
24	Berthe L.	48	Antoine C.

Table of random numbers

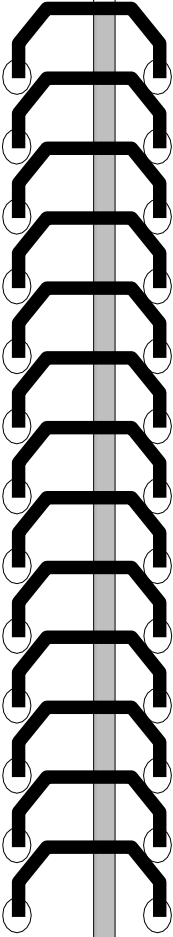
6 8 4 2 5 7 9 5 4 1 2 5 6 3 2 1 4 0

5 8 2 0 3 2 1 5 4 7 8 5 9 6 2 0 2 4

3 6 2 3 3 3 2 5 4 7 8 9 1 2 0 3 2 5

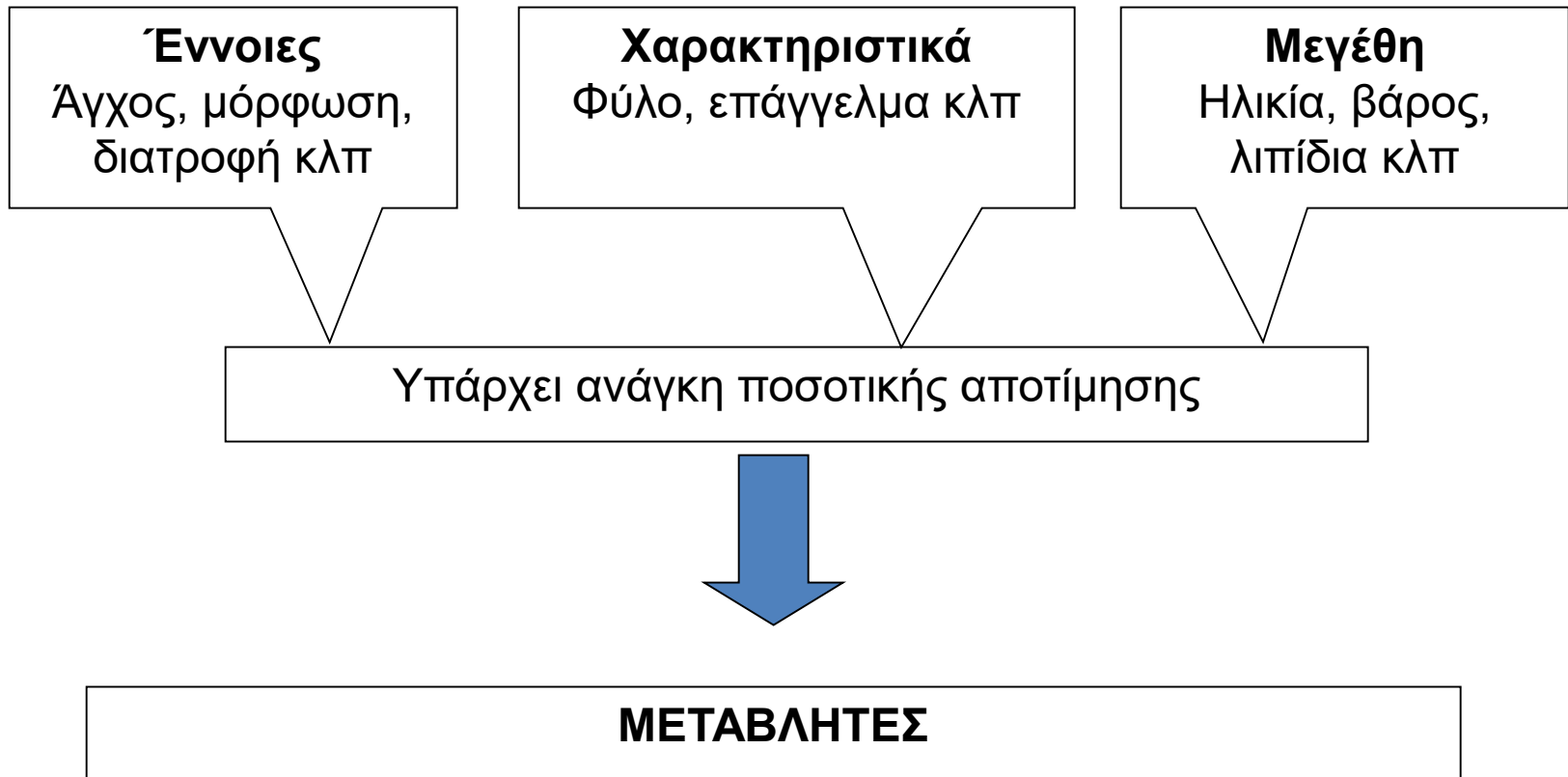
9 8 5 2 6 3 0 1 7 4 2 4 5 0 3 6 8 6

Systematic sampling

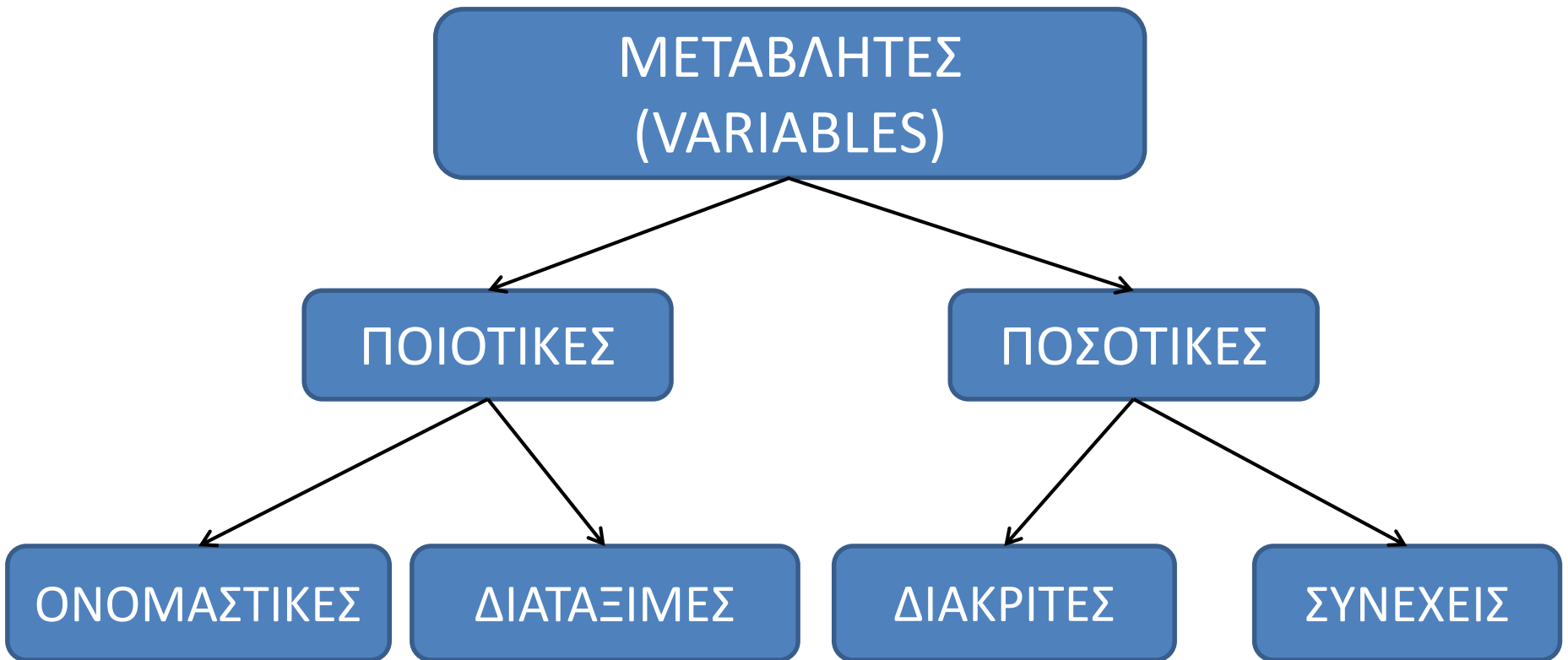


1	Albert D.	25	Monique Q.
2	Richard D.	26	Régine D.
3	Belle H.	27	Lucille L.
4	Raymond L.	28	Jérémy W.
5	Stéphane B.	29	Gilles D.
6	Albert T.	30	Renaud S.
7	Jean William V.	31	Pierre K.
8	André D.	32	Etienne M.
9	Jeremy W.	33	Marie M.
10	Anthony Q.	34	Gaétan Z.
11	James B.	35	Fidèle D.
12	Denis G.	36	Maria P.
13	Amanda L.	37	Anne-Marie G.
14	Jennifer L.	38	Michel K.
15	Philippe K.	39	Gaston C.
16	Eve F.	40	Alain M.
17	Priscilla O.	41	Olivier P.
18	Robert D.	42	Geneviève M.
19	Brian F.	43	Berthe D.
20	Hellène H.	44	Jean Pierre P.
21	Isabelle R.	45	Jacques B.
22	Jean T.	46	François P.
23	Samanta D.	47	Dominique M.
24	Berthe L.	48	Antoine C.

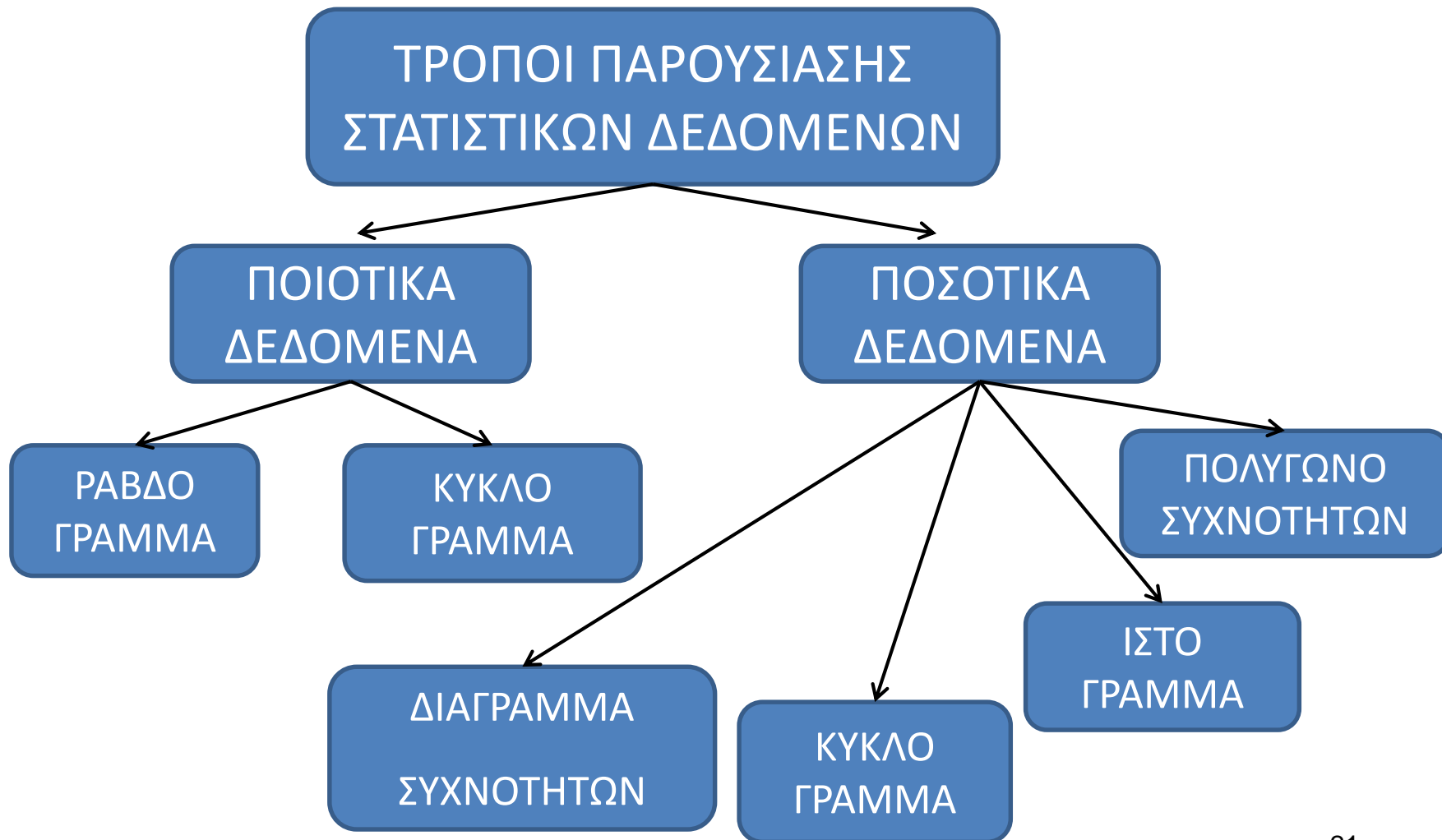
Η ανάγκη χρήσης μεταβλητών



Περιγραφική Στατιστική



Περιγραφική Στατιστική



Περιγραφική Στατιστική

- **Παράδειγμα:** Σε ένα δείγμα 20 οικογενειών από μια περιοχή της Αθήνας, το επάγγελμα του πατέρα, το ημερομίσθιο του πατέρα και ο αριθμός παιδιών της οικογένειας ήταν:

επάγγελμα	ημερομίσθιο	αρ_ παιδιών
εργάτης	70	0
οδηγός	75	1
εργάτης	80	0
δημ_υπόδητης	70	2
δημ_υπόδητης	80	2
δημ_υπόδητης	50	2
δάσκαλος	90	3
ιερέας	100	2
οδηγός	60	4
εργάτης	60	1
δάσκαλος	70	1
εργάτης	60	2
εργάτης	80	3
δημ_υπόδητης	70	4
ιερέας	90	1
δάσκαλος	100	2
εργάτης	90	2
δημ_υπόδητης	65	2
δάσκαλος	75	2
δημ_υπόδητης	80	2

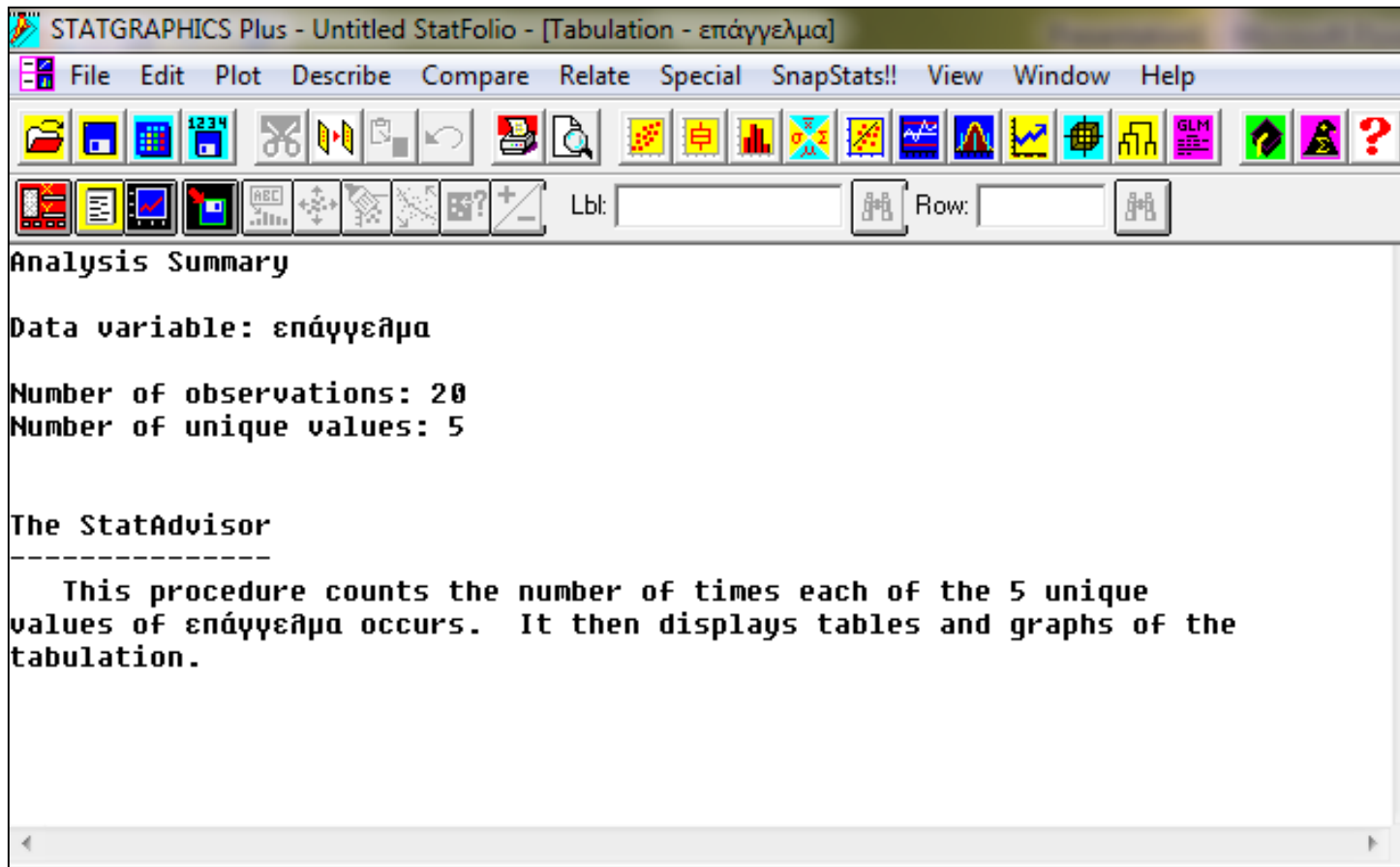
Περιγραφική Στατιστική

- Στο παράδειγμα, για να εισάγουμε τη μεταβλητή «επάγγελμα», επιλέγουμε μεταβλητή τύπου character αφού πρόκειται για ποιοτική μεταβλητή, την ονομάζουμε «επάγγελμα» και στη συνέχεια πληκτρολογούμε τα δεδομένα μας:

	επάγγελμα	Col_2	Col_3	Col_4	Col_5
1	εργάτης				
2	οδηγός				
3	εργάτης				
4	δημ_υπόδηλος				
5	δημ_υπόδηλος				
6	δημ_υπόδηλος				
7	δάσκαλος				
8	ιερέας				
9	οδηγός				
10	εργάτης				
11	δάσκαλος				
12	εργάτης				
13	εργάτης				
14	δημ_υπόδηλος				
15	ιερέας				
16	δάσκαλος				
17	εργάτης				
18	δημ_υπόδηλος				
19	δάσκαλος				
20	δημ_υπόδηλος				

Περιγραφική Στατιστική

- Η ανάλυση που προκύπτει άμεσα είναι η ακόλουθη:



STATGRAPHICS Plus - Untitled StatFolio - [Tabulation - επάγγελμα]

File Edit Plot Describe Compare Relate Special SnapStats!! View Window Help

Analysis Summary

Data variable: επάγγελμα

Number of observations: 20
Number of unique values: 5

The StatAdvisor

This procedure counts the number of times each of the 5 unique values of επάγγελμα occurs. It then displays tables and graphs of the tabulation.

Περιγραφική Στατιστική

- Και ο πίνακας συχνοτήτων:

Frequency Table for επάγγελμα

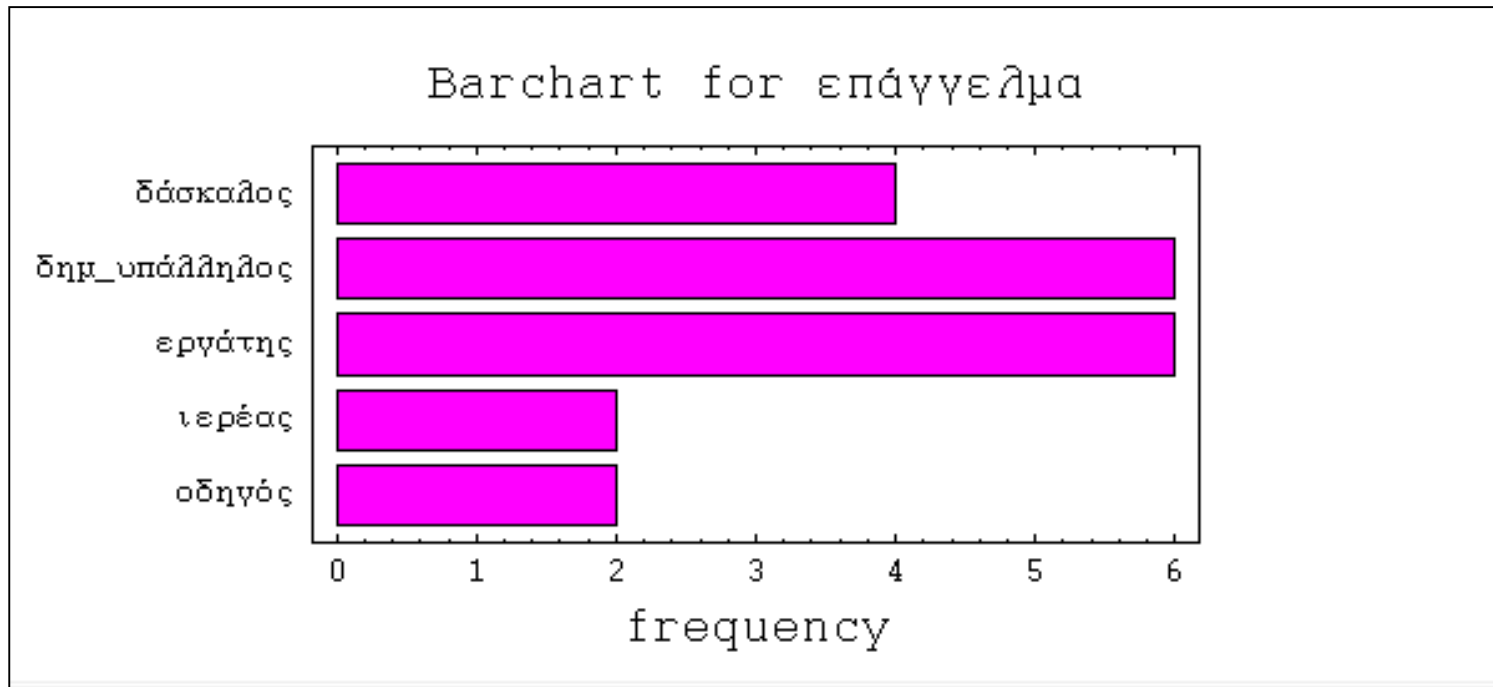
Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	δάσκαλος	4	0,2000	4	0,2000
2	δημ_ υπάλληλος	6	0,3000	10	0,5000
3	εργάτης	6	0,3000	16	0,8000
4	ιερέας	2	0,1000	18	0,9000
5	οδηγός	2	0,1000	20	1,0000

The StatAdvisor

This table shows the number of times each value of επάγγελμα occurred, as well as percentages and cumulative statistics. For example, in 4 rows of the data file επάγγελμα equaled δάσκαλος. This represents 20,0% of the 20 values in the file. The rightmost two columns give cumulative counts and percentages from the top of the

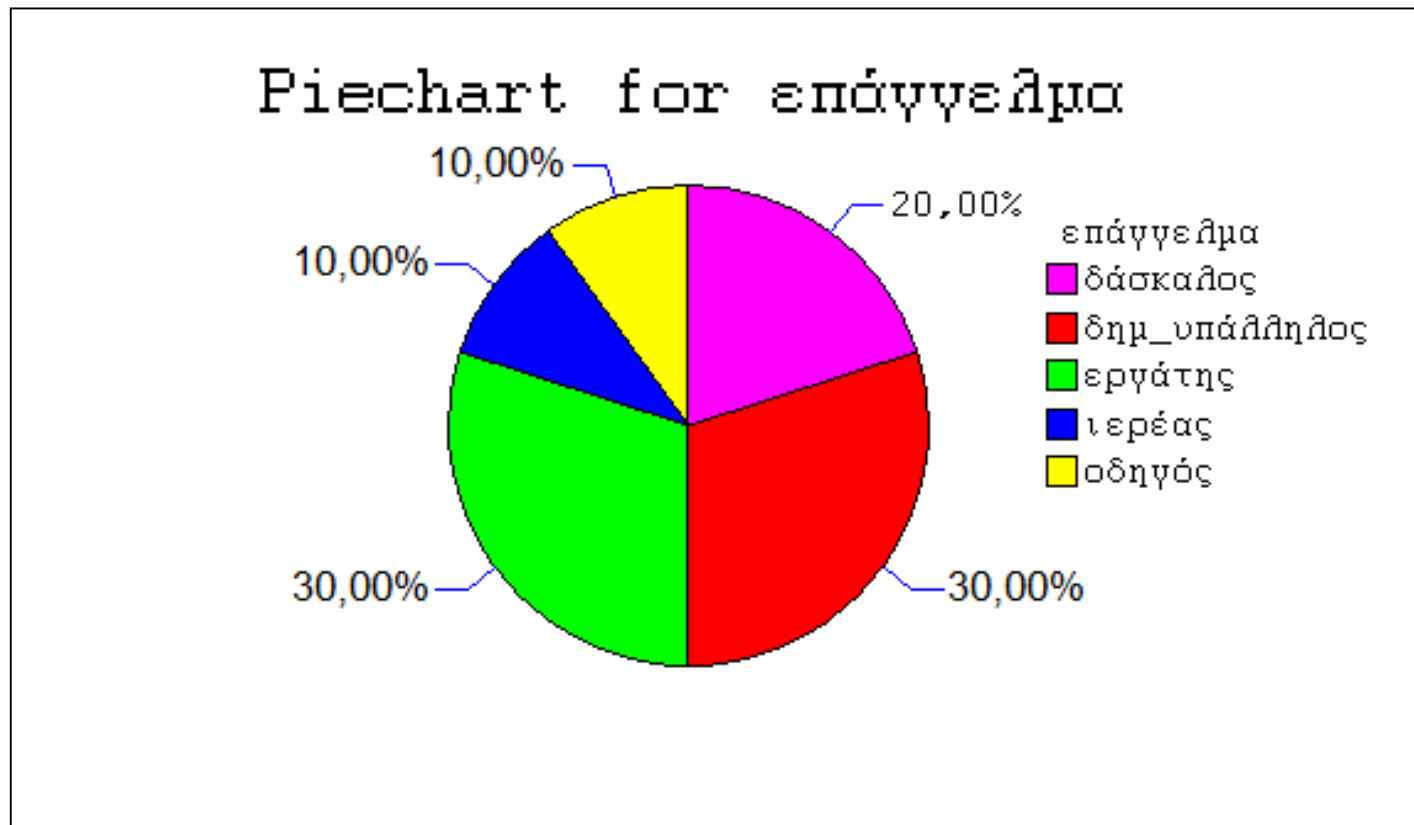
Περιγραφική Στατιστική

- Και το ραβδόγραμμα συχνοτήτων:



Περιγραφική Στατιστική

- Και το κυκλόγραμμα (πίτα) συχνοτήτων:



Περιγραφική Στατιστική

- Συνεπώς έχουμε εισάγει τις δύο μεταβλητές και έχουμε αυτή την εικόνα:

	επάγγελμα	ημερομίσθιο	
1	εργάτης	70	
2	οδηγός	75	
3	εργάτης	80	
4	δημ_υπάλληλος	70	
5	δημ_υπάλληλος	80	
6	δημ_υπάλληλος	50	
7	δάσκαλος	90	
8	ιερέας	100	
9	οδηγός	60	
10	εργάτης	60	
11	δάσκαλος	70	
12	εργάτης	60	
13	εργάτης	80	
14	δημ_υπάλληλος	70	
15	ιερέας	90	
16	δάσκαλος	100	
17	εργάτης	90	
18	δημ_υπάλληλος	65	
19	δάσκαλος	75	
20	δημ_υπάλληλος	80	

STATGRAPHICS Plus - Untitled StatFolio

File Edit Plot Describe Compare Relate Special SnapStats!! View Window Help

One-Variable Analysis - ημερομίσθιο

Frequency Tabulation for ημερομίσθιο

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		47,0		0	0,0000	0	0,0000
1	47,0	57,0	52,0	1	0,0500	1	0,0500
2	57,0	67,0	62,0	4	0,2000	5	0,2500
3	67,0	77,0	72,0	6	0,3000	11	0,5500
4	77,0	87,0	82,0	4	0,2000	15	0,7500
5	87,0	97,0	92,0	3	0,1500	18	0,9000
6	97,0	107,0	102,0	2	0,1000	20	1,0000
above	107,0			0	0,0000	20	1,0000

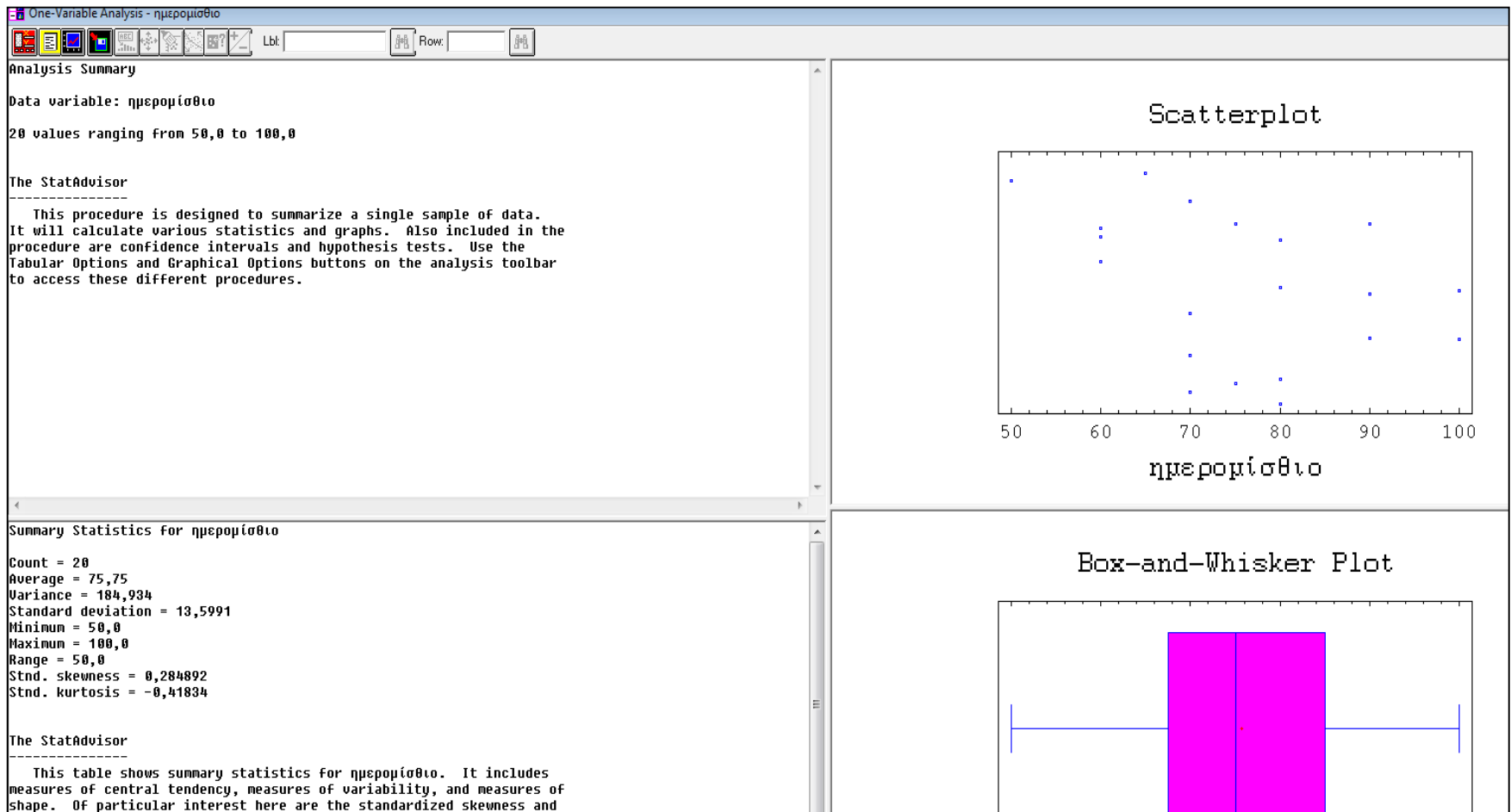
Mean = 75,75 Standard deviation = 13,5991

The StatAdvisor

This option performs a frequency tabulation by dividing the range of ημερομίσθιο into equal width intervals and counting the number of data values in each interval. The frequencies show the number of data values in each interval, while the relative frequencies show the proportions in each interval. You can change the definition of the intervals by pressing the alternate mouse button and selecting Pane Options. You can see the results of the tabulation graphically by selecting Frequency Histogram from the list of Graphical Options.

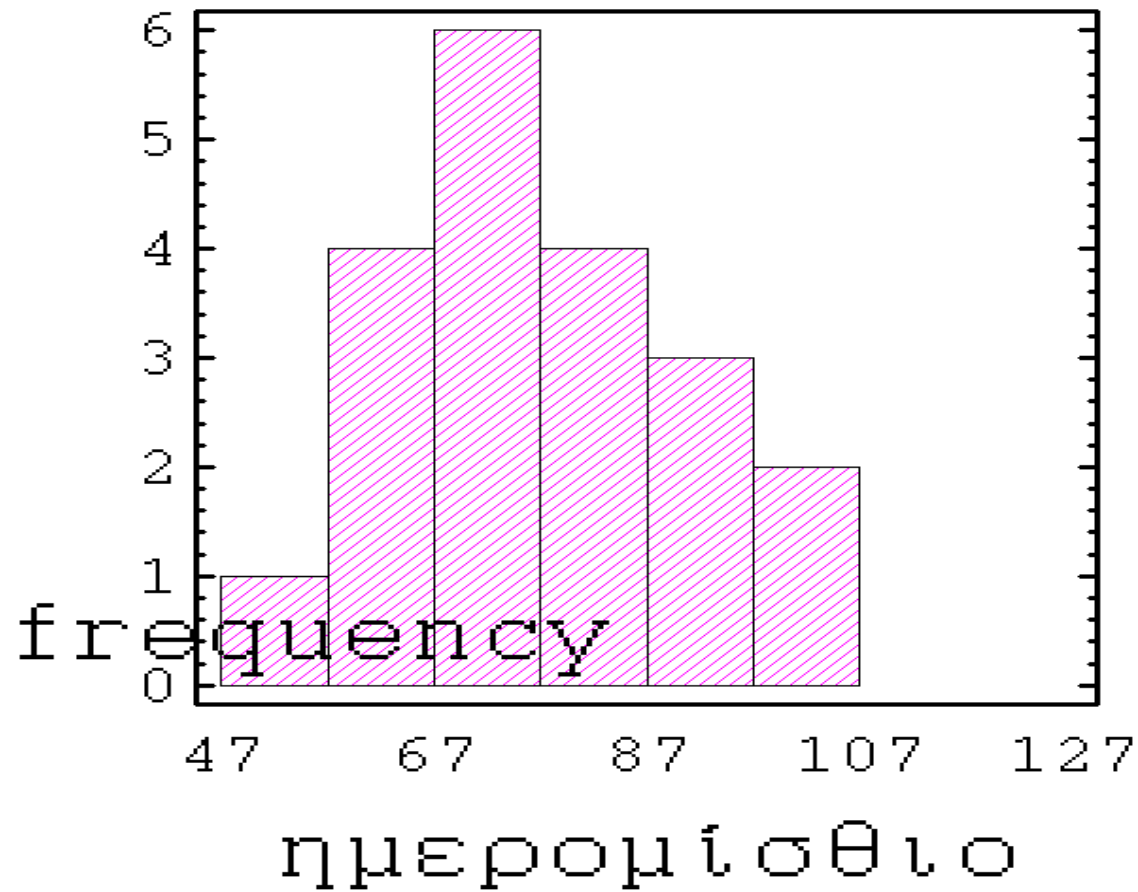
Περιγραφική Στατιστική

- By default εμφανίζεται η παρακάτω ανάλυση:



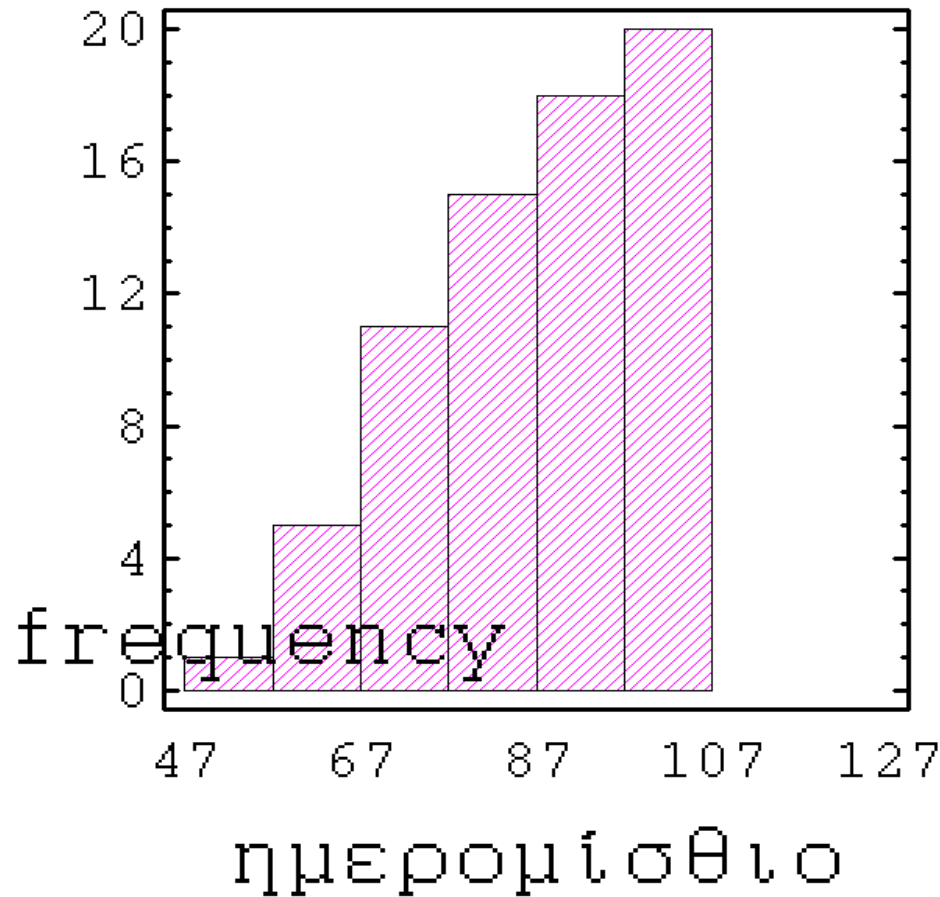
Περιγραφική Στατιστική

Histogram



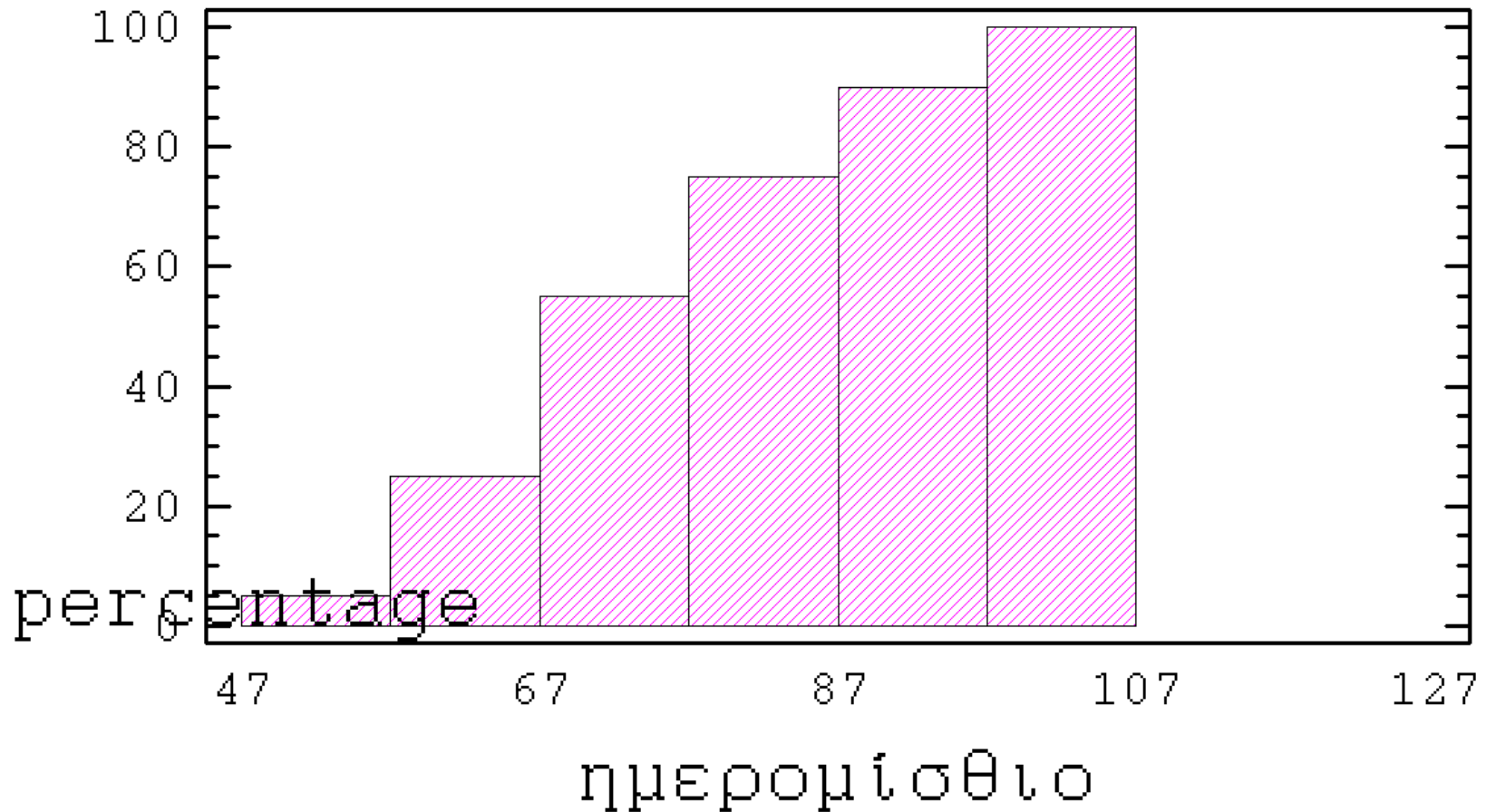
Περιγραφική Στατιστική

Histogram



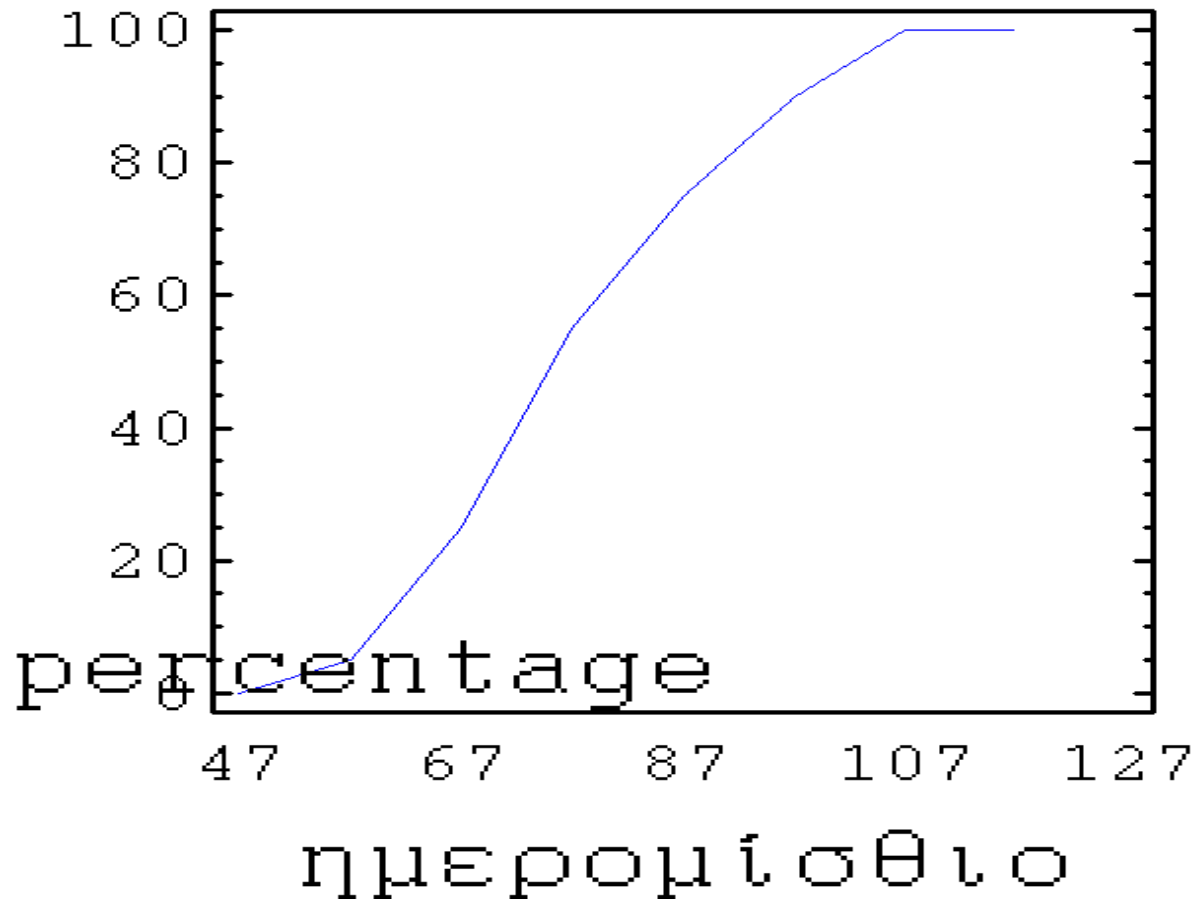
Περιγραφική Στατιστική

Histogram



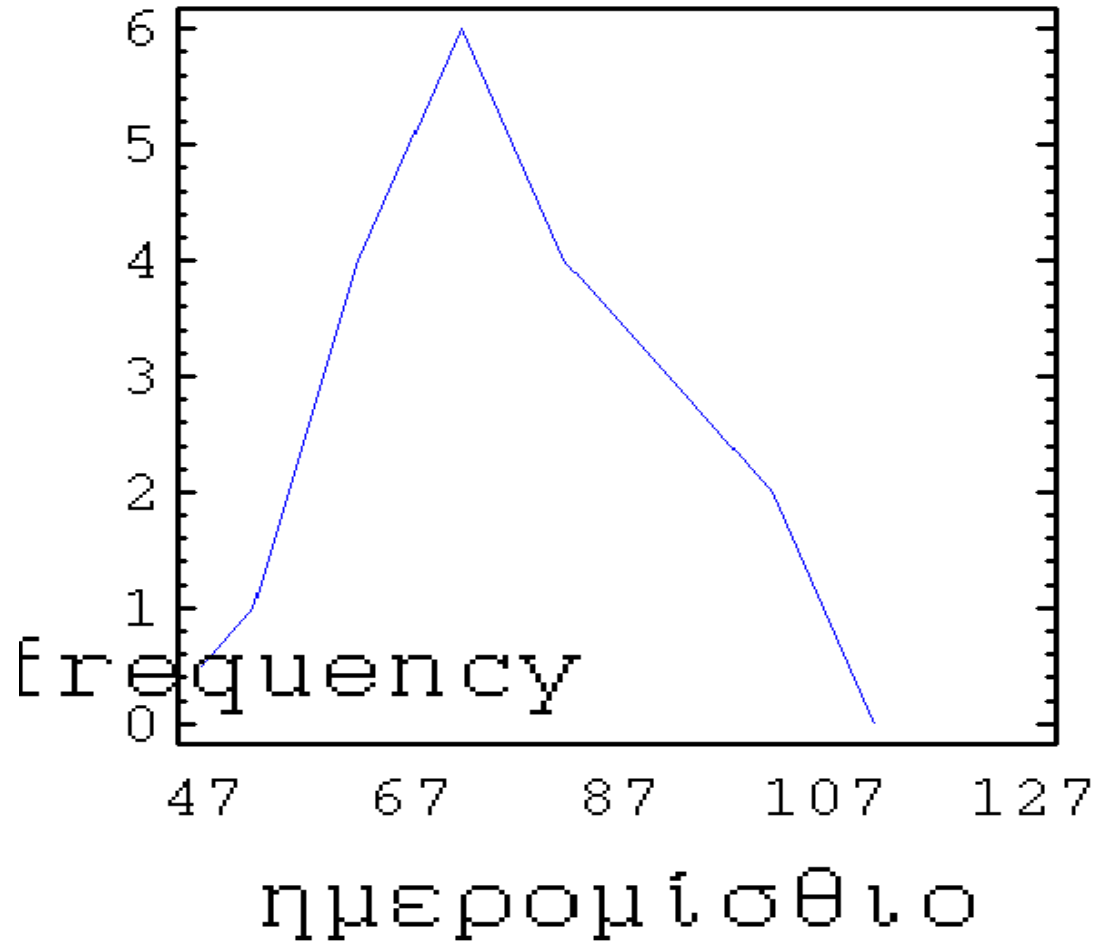
Περιγραφική Στατιστική

Histogram

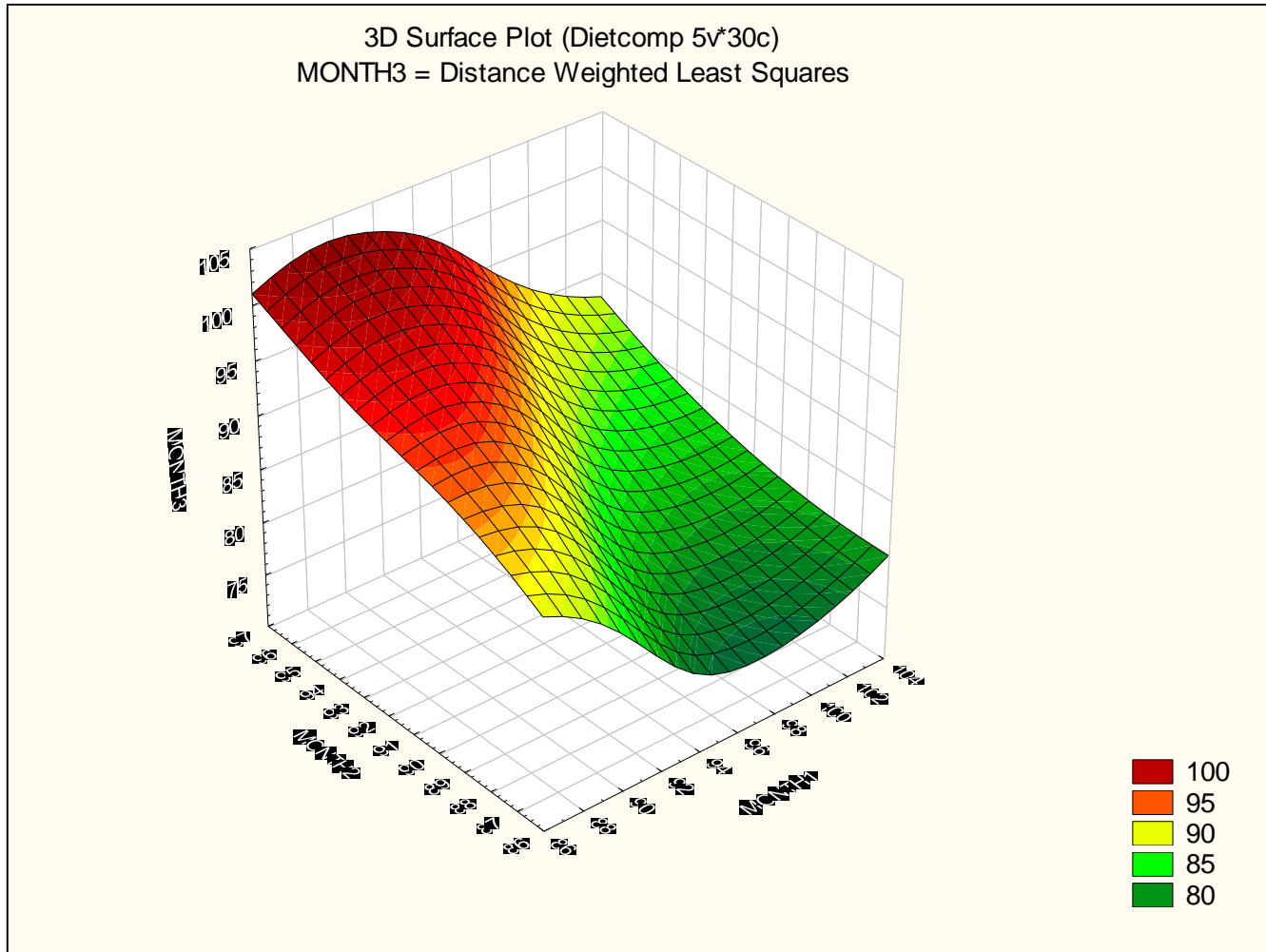


Περιγραφική Στατιστική

Histogram



Διαγράμματα επιφανείας



Περιγραφική Στατιστική

- Ως περιγραφικά μέτρα θέσης εννοούμε τα εξής:
 - Μέση Τιμή
 - Κορυφή ή Επικρατούσα Τιμή
 - Διάμεσο
 - Ποσοστημόρια

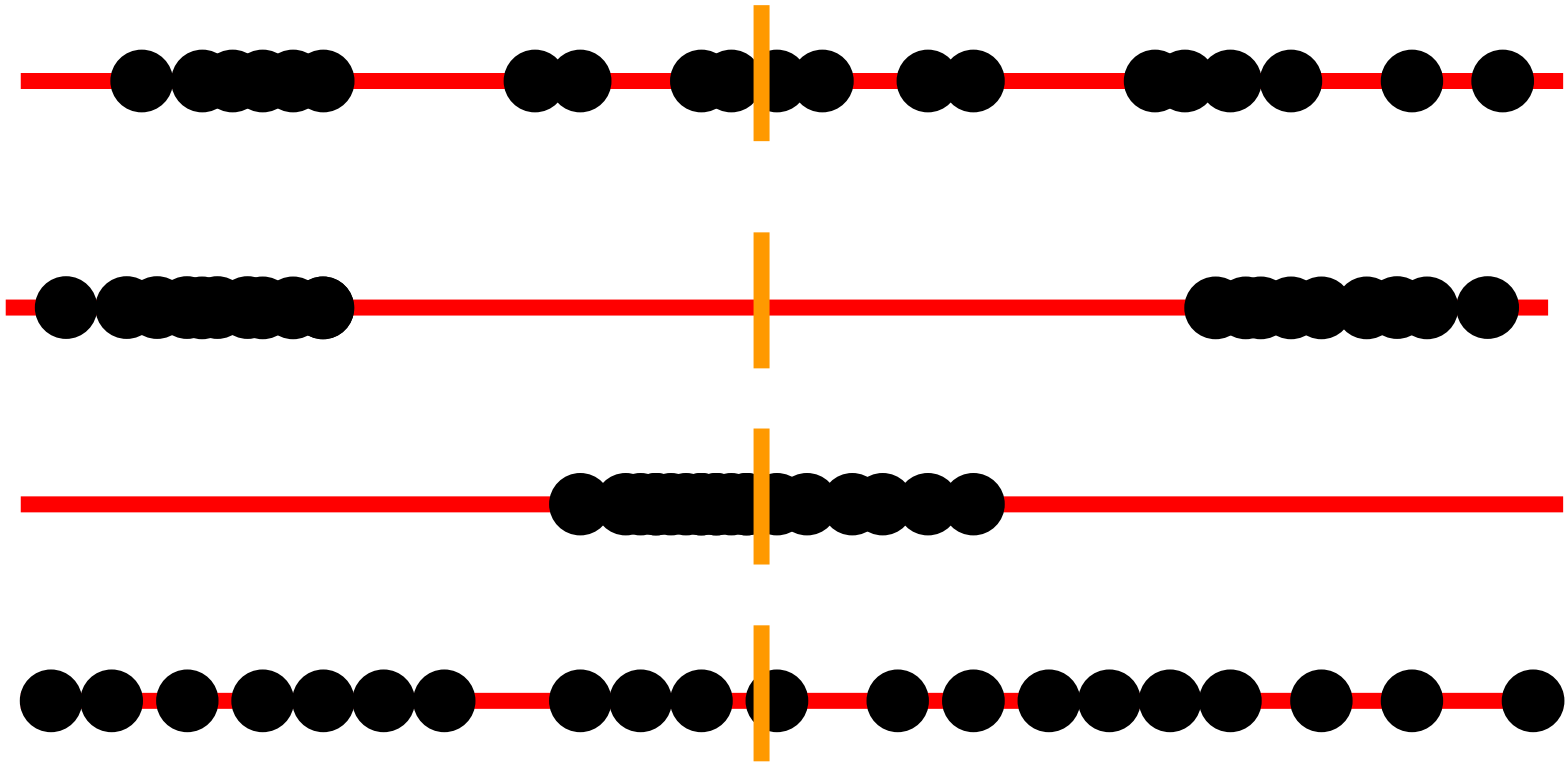
Η ερμηνεία των περιγραφικών μέτρων

- **Αριθμητικός μέσος**

- Η αναμενόμενη τιμή που θα έχει η ποσοτική μεταβλητή σε ένα τυχαία επιλεγμένο άτομο του δείγματος.

- ✓ Πόσο αξιόπιστο μέτρο είναι; (όταν στο δείγμα υπάρχει ανισοκατανομή).

Αριθμητικός μέσος

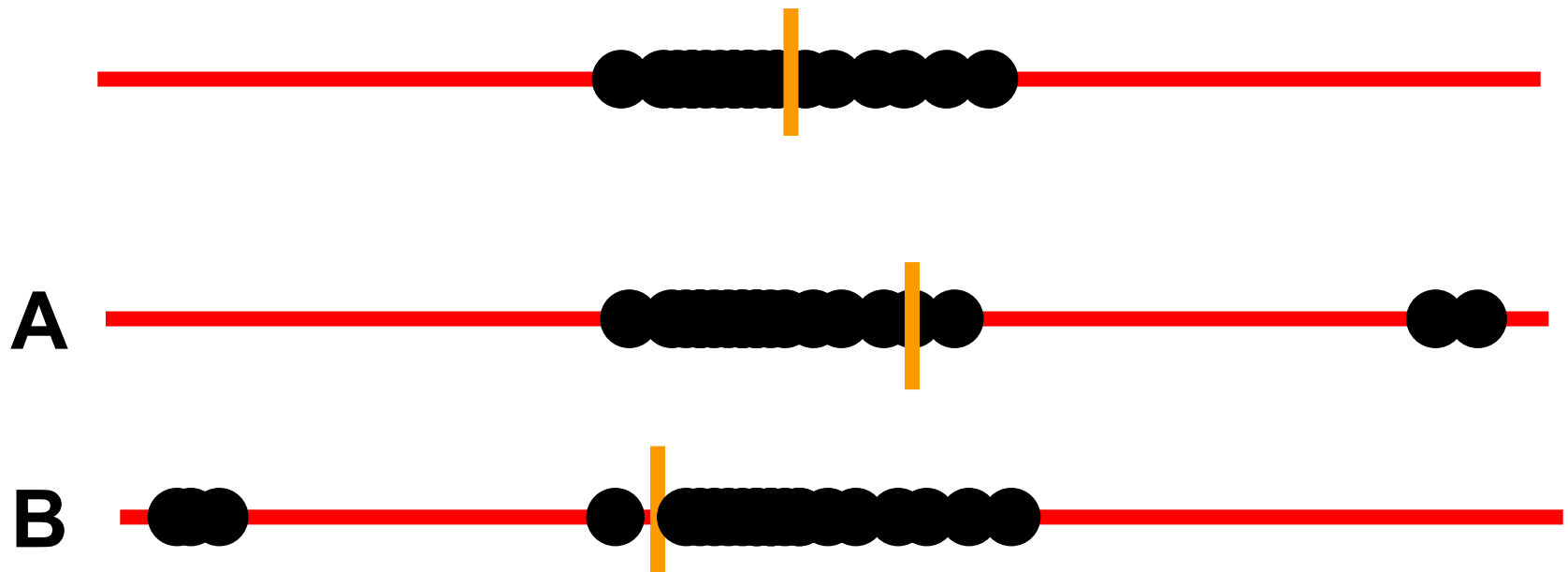


Αριθμητικός
μέσος

Η ερμηνεία των περιγραφικών μέτρων

- Πότε χρειαζόμαστε τη **διάμεσο**;
 - Όταν έχουμε ακραίες τιμές στην κατανομή της ποσοτικής μεταβλητής,
 - ✓ και ειδικότερα όταν είναι **ασύμμετρα** κατανεμημένες.

Αριθμητικός μέσος και ακραίες τιμές



Περιγραφική Στατιστική

- Ως πιο διαδεδομένα περιγραφικά μέτρα διασποράς εννοούμε τα εξής:
 - Εύρος
 - Ενδοτεταρτημοριακή απόκλιση
 - Μέση απόκλιση
 - Διασπορά ή Διακύμανση
 - Τυπική απόκλιση

Η ερμηνεία των περιγραφικών μέτρων

- **Τυπική απόκλιση**
 - Ένας δείκτης μεταβλητότητας των τιμών της ποσοτικής μεταβλητής.
 - ✓ Όσο μικρότερες τιμές λαμβάνει, τόσο πιο ομοιογενές το δείγμα.
 - ✓ **Επηρεάζεται από τις μονάδες μέτρησης.**

Μέτρηση της μεταβλητότητας

- Συντελεστής μεταβλητότητας
 - Ένας δείκτης μεταβλητότητας των τιμών της ποσοτικής μεταβλητής, που λαμβάνει υπόψη την μέση τιμή και δεν επηρεάζεται από τις μονάδες μέτρησης.

STATGRAPHICS Plus - Untitled StatFolio

File Edit Plot Describe Compare Relate Special SnapStats!! View Window Help

One-Variable Analysis - ημερομίσθιο

Summary Statistics for ημερομίσθιο

Count = 20
 Average = 75,75
 Median = 75,0
 Mode =
 Minimum = 50,0
 Maximum = 100,0
 Range = 50,0
 Lower quartile = 67,5
 Upper quartile = 85,0

The StatAdvisor

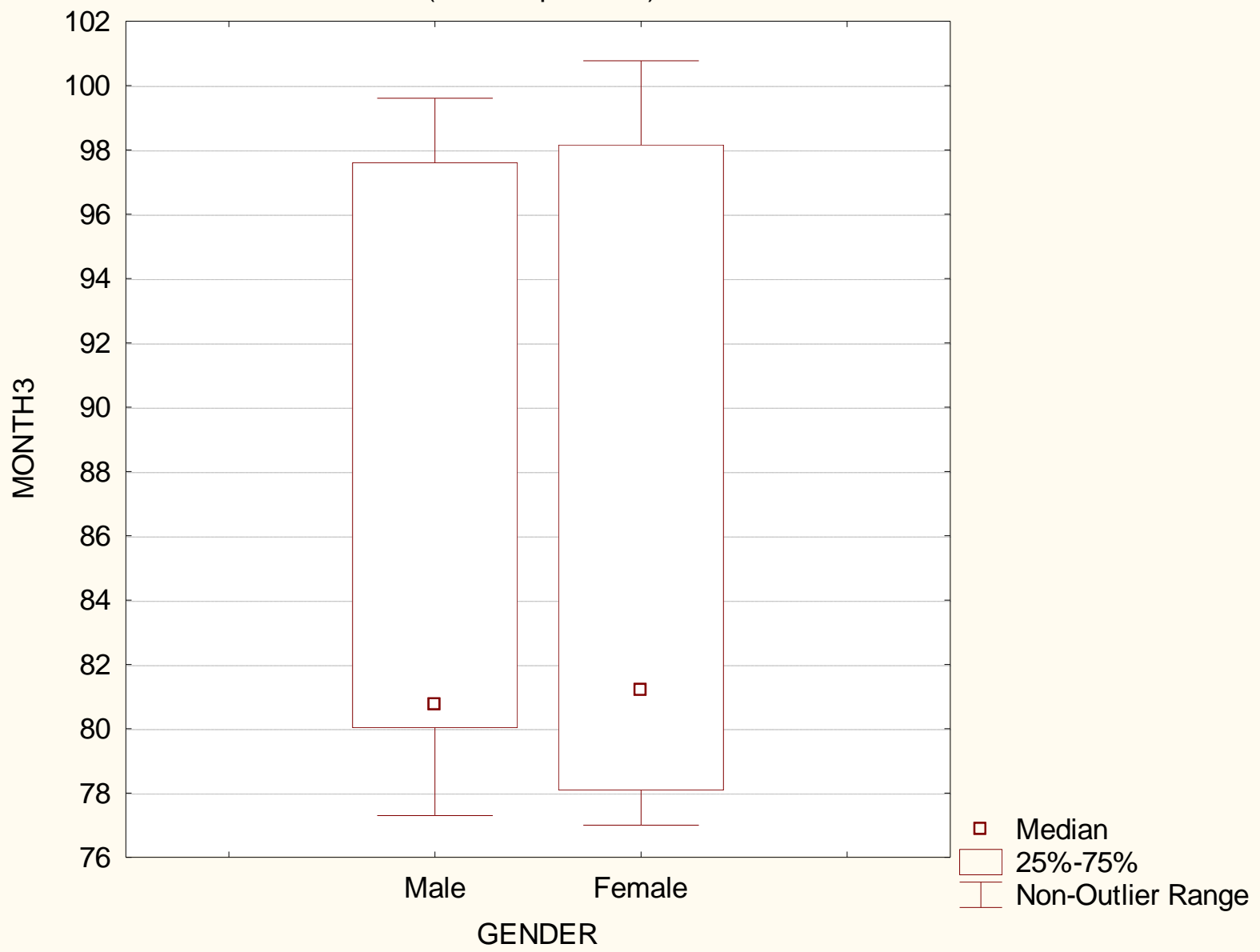
This table shows summary statistics for measures of central tendency, measures of shape. Of particular interest here are the standardized kurtosis, which can be used to determine if the sample comes from a normal distribution. Values outside the range of -2 to +2 indicate signs of non-normality, which would tend to invalidate a test regarding the standard deviation. In this case, the standardized skewness value is within the range expected for data from a normal distribution. The standardized kurtosis value is within the range expected for data from a normal distribution.

Summary Statistics Options

<input checked="" type="checkbox"/> Average	<input checked="" type="checkbox"/> Min.	<input type="checkbox"/> Skewness
<input checked="" type="checkbox"/> Median	<input checked="" type="checkbox"/> Max.	<input type="checkbox"/> Std. Skewness
<input checked="" type="checkbox"/> Mode	<input checked="" type="checkbox"/> Range	<input type="checkbox"/> Kurtosis
<input type="checkbox"/> Geo. Mean	<input checked="" type="checkbox"/> Lower Quartile	<input type="checkbox"/> Std. Kurtosis
<input type="checkbox"/> Variance	<input checked="" type="checkbox"/> Upper Quartile	<input type="checkbox"/> Coeff. of Var.
<input type="checkbox"/> Std. Deviation	<input type="checkbox"/> Interquartile Range	<input type="checkbox"/> Sum
<input type="checkbox"/> Std. Error		

OK Cancel All Help

Box Plot (Dietcomp 5v*30c)

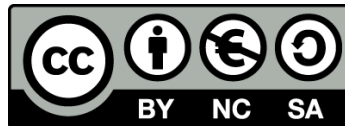


Τέλος Ενότητας

Σημειώματα

Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση.



[1] <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.