Σκοπός - Μαθησιακά Αποτελέσματα

Το συγκεκριμένο μάθημα έχει ως σκοπό να εισάγει τους καταρτιζόμενους σε ***προχωρημένες έννοιες της περιγραφικής στατιστικής***, προκειμένου να καταστούν ικανοί να αξιοποιήσουν τις γνώσεις αυτές σε στοιχειώδεις εφαρμογές και στην ερμηνεία φαινομένων, στην καθημερινή λειτουργία των επιχειρήσεων ή των οργανισμών στους οποίους θα εργαστούν με το πέρας των σπουδών τους. Είναι δε μάθημα βασικής κατάρτισης με το οποίο οι καταρτιζόμενοι αποκτούν τις βασικές γνώσεις - ικανότητες - δεξιότητες για την ειδικότητά τους.

Από τον οδηγό σπουδών

Περιεχόμενο του Μαθήματος:

- Ασυμμετρία – κύρτωση
- Ροπές κατανομής συχνοτήτων
- Ροπές περί την αρχή (x=0)
- Ροπές περί το μέσο αριθμητικό
- Κύρτωση
- Παλινδρόμηση και συσχέτιση δυο μεταβλητών
- Μέθοδος Ελαχίστων τετραγώνων
- Ευθεία ελαχίστων τετραγώνων
- Μέσο τετραγωνικό σφάλμα – Δείκτης προσδιορισμού
- Καμπύλη ελαχίστων τετραγώνων – Β’ βαθμού παραβολή

Περιεχόμενο του Μαθήματος:

- Εκθετική παλινδρόμηση
- Συσχετισμένες μεταβλητές
- Γραμμική μεταβολή
- Συνδιακύμανση δύο μεταβλητών
- Συντελεστής συσχέτισης
- Χρονολογικές σειρές
- Οι κυριότερες συνιστώσες μιας χρονολογικής σειράς
- Αριθμοδείκτες
- Απλοί και σύνθετοι αριθμοδείκτες
- Ιδανικός τύπος Fisher
- Δείκτης τιμών καταναλωτή ή δείκτης κόστους ζωής

Εφαρμογή - Παράδειγμα της Γεωμετρικής ερμηνείας μέσης τιμής και τυπικής απόκλισης

Σε δείγμα 30 σπουδαστών ενός ΙΕΚ βρέθηκε μέσο ύψος 168,3 και τυπική απόκλιση 4,2 μονάδες.

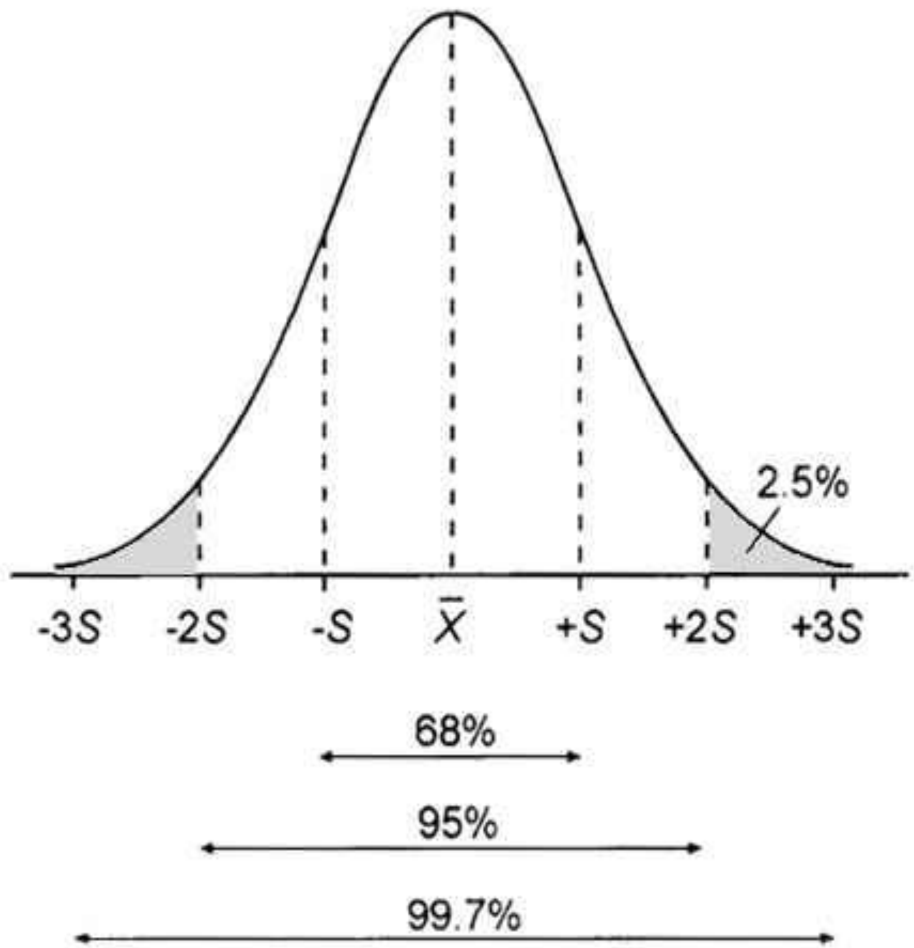
(α) Αν το δείγμα αυτό είναι αντιπροσωπευτικό του συνόλου των σπουδαστών του ΙΕΚ τότε να βρεθεί διάστημα συμμετρικό γύρω από τη μέση τιμή στο οποίο θα ανήκει το ύψος του 68% περίπου των σπουδαστών.

(β) **Αν το ΙΕΚ έχει 150 σπουδαστές** τότε πόσοι από αυτούς περιμένουμε να έχουν ύψος που θα ανήκει στο παραπάνω διάστημα;

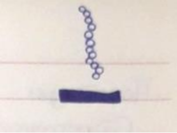
$(\bar{X} - \sigma \qquad \bar{X} + \sigma)$

(168,3 – 4,2   168,3 + 4,2)

(164,1   172,5)



ΜΕΤΡΑ   ΔΙΑΣΠΟΡΑΣ



<u>Ορισμός:</u>

Ο ***συντελεστής διακύμανσης CV*** (*coefficient of variation)* είναι ένα μέτρο (ένας δείκτης) που χρησιμοποιείται για την σύγκριση της διασποράς σε δύο κατανομές

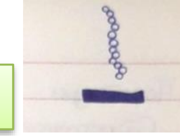
Συντελεστής διακύμανσης: CV = σ / $\bar{x}$ = σταθερή απόκλιση / μέσο

<u>Εφαρμογή:</u>

Συντελεστής διακύμανσης για το παράδειγμα με τα 60 χωράφια: 16,48 / 74,20   = 0,2221 ή 22,21%

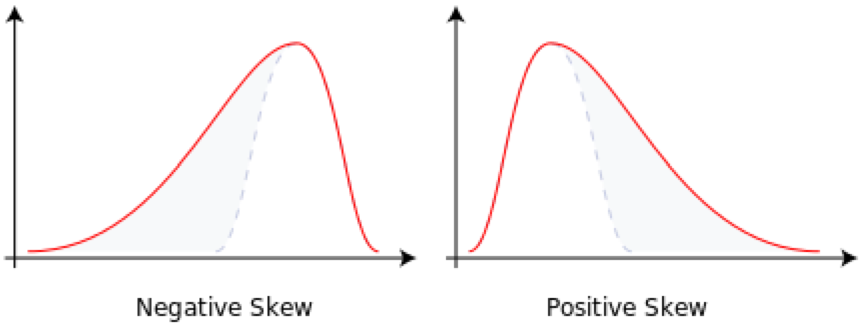
και ο Συντελεστής διακύμανσης για το παράδειγμα με τις προμήθειες 100 στελεχών:

10,36 / 50 = 0,2072 ή 20,72%

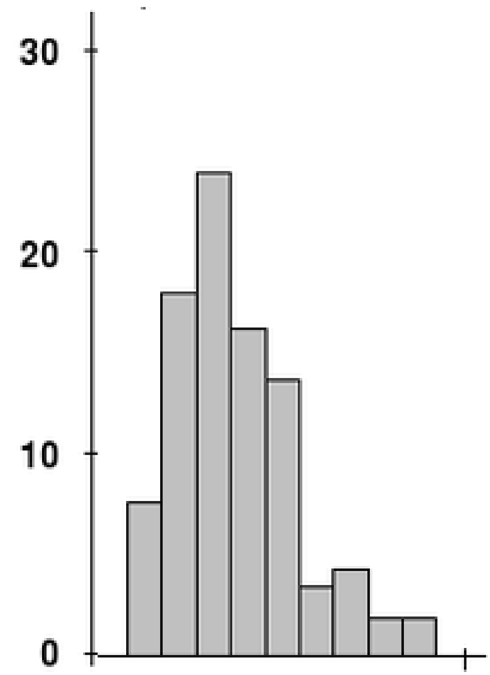


ΜΕΤΡΑ - ΑΣΥΜΜΕΤΡΙΑΣ

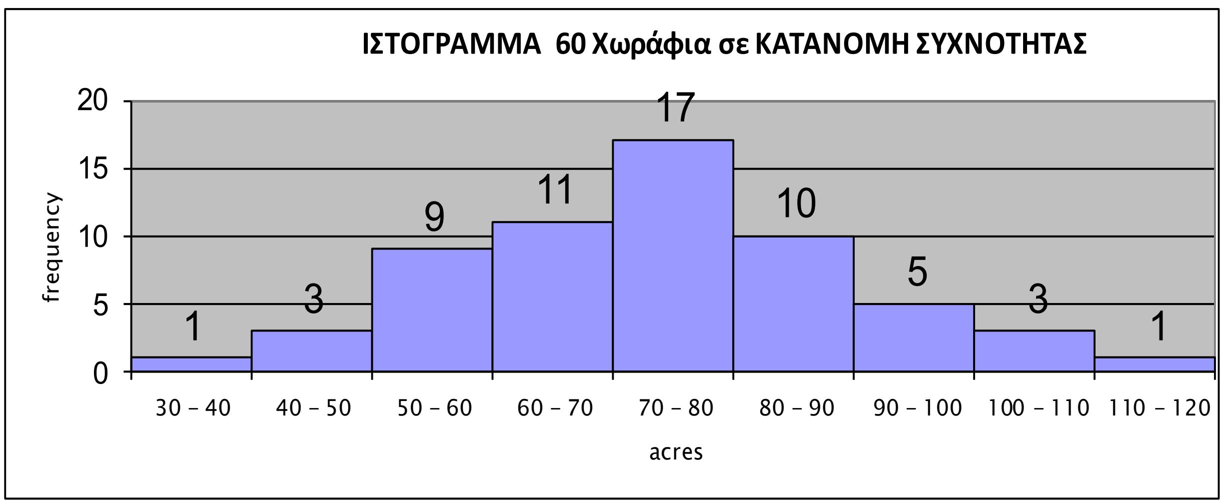
<u>Συμμετρικά και ασύμμετρα δεδομένα</u>



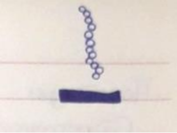
Αρνητική Ασυμμετρία   Θετική Ασυμμετρία



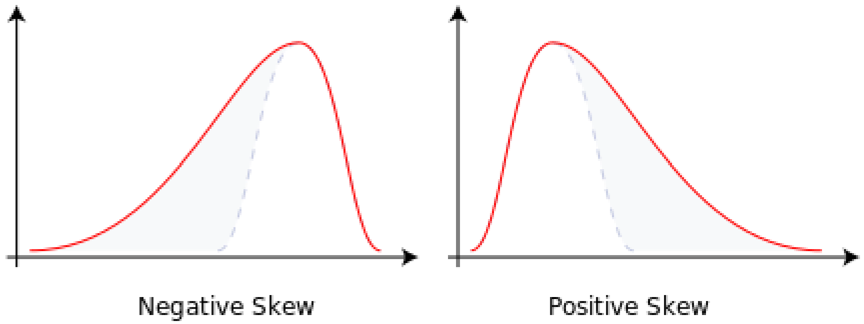
Θετική Ασυμμετρία



Συμμετρική κατανομή … ?



Συμμετρικά και ασύμμετρα δεδομένα - Κύρτωση / Skewness



Αρνητική Ασυμμετρία Θετική Ασυμμετρία

Αν ένας πληθυσμός είναι κατά προσέγγιση συμμετρικός, τότε σε δείγμα του μεγέθους ($\nu \geq 30$) οι τιμές του **μέσου $\bar{X}$** και της **διαμέσου Δ** δε διαφέρουν πολύ μεταξύ τους.

Αν τα δεδομένα έχουν και **επικρατούσα τιμή Μ** (λέγεται και ***κορυφή***), τότε η τιμή της είναι κοντά στην τιμή του μέσου και της διαμέσου.

Αν ένας πληθυσμός είναι απόλυτα συμμετρικός (*μηδενική ασυμμετρία*), τότε

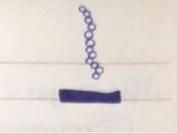
Μέσος $\bar{X}$ = **Διάμεσος Δ** = **Επικρατούσα τιμή Μ**.

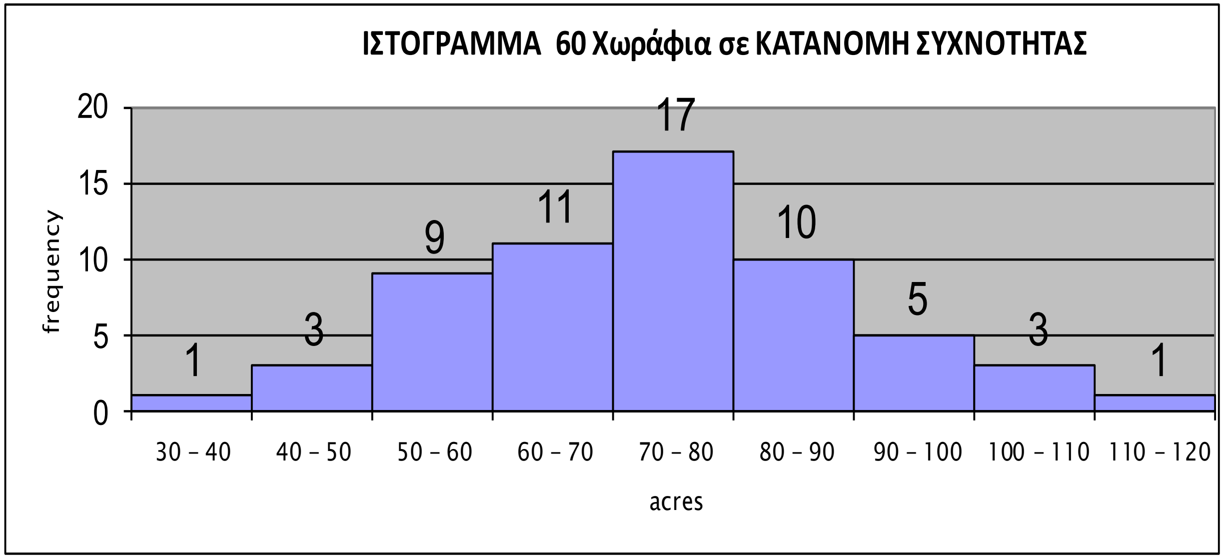
<u>Συντελεστής ασυμμετρίας του Pearson:</u>

(Μέσος – Επικρατούσα Τιμή) / Τυπική Απόκλιση δηλ. **$(\bar{X} - M) / \sigma$**

Εναλλακτικός συντελεστής **3(Μέσος $\bar{X}$ – Διάμεσος Δ) / σ**

ΜΕΤΡΑ - ΑΣΥΜΜΕΤΡΙΑΣ





Η τυπική απόκλιση είναι: σ = 16,48

Ο μέσος είναι: $\bar{X}$ = 74,20

Η επικρατούσα τιμή είναι: Μ = 75 (αυτή που έχει την μεγαλύτερη συχνότητα)

Συντελεστής ασυμμετρίας του Pearson:

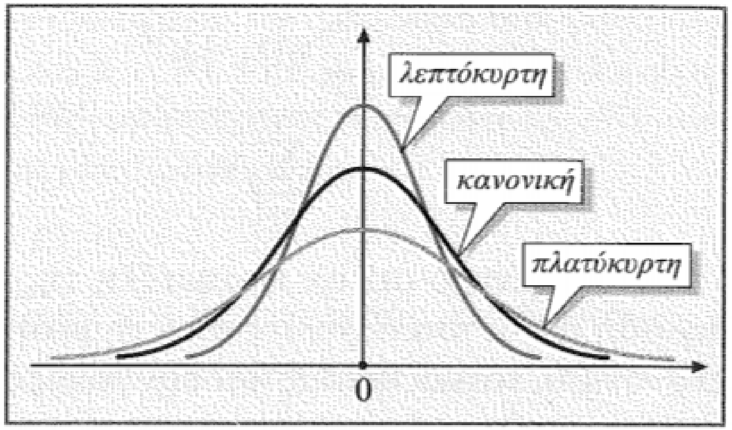
(Μέσος – Επικρατούσα Τιμή) / Τυπική Απόκλιση δηλ. **($\bar{X}$ – M) / σ**

(74,20 – 75) / 16,48 = - 0,80/16,48 = - 0,0485 δηλαδή ***πάρα πολύ μικρή αρνητική ασυμμετρία***

ΜΕΤΡΑ ΚΥΡΤΟΤΗΤΑΣ - ΡΟΠΕΣ

β) **Μέτρο κυρτότητας του Pearson**. Για να χαρακτηρίσουμε την κυρτότητα του διαγράμματος (της καμπύλης) μιας κατανομής συχνοτήτων, το συγκρίνουμε και πάλι με το διάγραμμα της κανονικής κατανομής. Έτσι, μια καμπύλη κατανομής συχνοτήτων η οποία παρουσιάζει «αιχμηρότητα» **μεγαλύτερη, ίση** ή **μικρότερη** από αυτή της κανονικής κατανομής χαρακτηρίζεται αντίστοιχα ως **λεπτόκυρτη**, **κανονική** ή **πλατύκυρτη**, όπως φαίνεται στο διπλανό σχήμα.

Ορίζουμε ως παράμετρο (μέτρο) κυρτότητας της κατανομής συχνοτήτων (δείκτη κυρτότητας του Pearson) τον αριθμό:



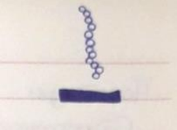
Σχ. 15

$$\beta_2 = \frac{\mu_4}{\sigma^4}, \quad \text{όπου} \quad \mu_4 = \frac{\sum_{i=1}^{N}(x_i-\mu)^4}{N} \quad \text{ή} \quad \mu_4 = \frac{\sum_{i=1}^{\kappa}\nu_i(x_i-\mu)^4}{\sum_{i=1}^{\kappa}\nu_i}$$

Αποδεικνύεται ότι ο συντελεστής (δείκτης) β_2 για κανονικές κατανομές είναι πάντοτε τρία (δηλαδή $\beta_2 = 3$), για λεπτόκυρτες κατανομές είναι $\beta_2 > 3$ και για πλατύκυρτες είναι $\beta_2 < 3$.

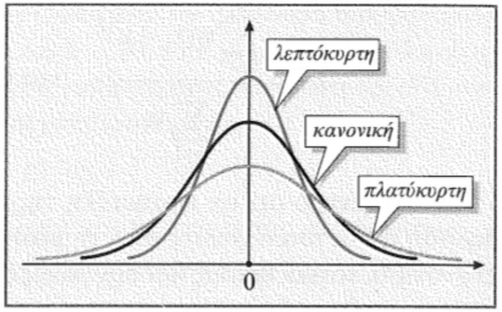
Είναι φανερό ότι όσο μεγαλύτερη είναι η διαφορά $\beta_2 - 3$, τόσο «αιχμηρότερη» είναι η καμπύλη της κατανομής συχνοτήτων.

ΜΕΤΡΑ ΚΥΡΤΟΤΗΤΑΣ - ΡΟΠΕΣ



Δείκτης κυρτότητας Pearson: $\beta_2 = \mu_4 / \sigma^4$

β) **Μέτρο κυρτότητας του Pearson**. Για να χαρακτηρίσουμε την κυρτότητα του διαγράμματος (της καμπύλης) μιας κατανομής συχνοτήτων, το συγκρίνουμε και πάλι με το διάγραμμα της κανονικής κατανομής. Έτσι, μια καμπύλη κατανομής συχνοτήτων η οποία παρουσιάζει «αιχμηρότητα» **μεγαλύτερη**, **ίση** ή **μικρότερη** από αυτή της κανονικής κατανομής χαρακτηρίζεται αντίστοιχα ως **λεπτόκυρτη**, **κανονική** ή **πλατύκυρτη**, όπως φαίνεται στο διπλανό σχήμα.



Σχ. 15

Ορίζουμε ως παράμετρο (μέτρο) κυρτότητας της κατανομής συχνοτήτων (δείκτη κυρτότητας του Pearson) τον αριθμό:

$$\beta_2 = \frac{\mu_4}{\sigma^4}, \quad \text{όπου} \quad \mu_4 = \frac{\sum_{i=1}^{N}(x_i - \mu)^4}{N} \quad \text{ή} \quad \mu_4 = \frac{\sum_{i=1}^{\kappa} v_i (x_i - \mu)^4}{\sum_{i=1}^{\kappa} v_i}$$

Αποδεικνύεται ότι ο συντελεστής (δείκτης) β_2 για κανονικές κατανομές είναι πάντοτε τρία (δηλαδή $\beta_2 = 3$), για λεπτόκυρτες κατανομές είναι $\beta_2 > 3$ και για πλατύκυρτες είναι $\beta_2 < 3$.

Είναι φανερό ότι όσο μεγαλύτερη είναι η διαφορά $\beta_2 - 3$, τόσο «αιχμηρότερη» είναι η καμπύλη της κατανομής συχνοτήτων.

Απλά δεδομένα:

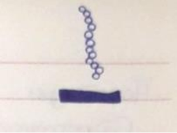
$$\mu_4 = [\, \Sigma (x_i - \bar{x})^4 \,] \; / \; n$$

Ομαδοποιημένα δεδομένα σε κατανομή συχνότητας:

$$\mu_4 = [\, \Sigma v_i (x_i - \bar{x})^4 \,] \; / \; \Sigma v_i$$

Kurtosis Formula

$$\text{Kurtosis} = n * \frac{\sum_i^n (Y_i - \bar{Y})^4}{\sum_i^n (Y_i - \bar{Y}^2)^2}$$



ΜΕΤΡΑ ΚΥΡΤΟΤΗΤΑΣ - ΡΟΠΕΣ

Σχόλιο: Τα σημαντικότερα από τα μέτρα θέσης και διασποράς (μέση τιμή και τυπική απόκλιση) καθώς και οι δείκτες ασυμμετρίας και κυρτότητας μιας κατανομής συχνοτήτων ανήκουν σε μια ευρύτερη οικογένεια παραμέτρων που ονομάζονται **ροπές**.

Οι ροπές ορίζονται είτε ως προς την αρχή των δεδομένων (δηλαδή το μηδέν) είτε ως προς το μέσο (κεντρική ροπή). Επίσης οι ροπές διαφοροποιούνται αν τα δεδομένα είναι ταξινομημένα σε κατανομές συχνοτήτων ή αταξινόμητα.

Ορίζουμε ως ροπή t-τάξης ως προς την αρχή μηδέν τον αριθμό $V_t = \frac{\sum_{i=1}^{N} x_i^t}{N}$, όταν τα δεδομένα είναι αταξινόμητα, και τον αριθμό $V_t = \frac{\sum_{i=1}^{\kappa} v_i x_i^t}{\sum_{i=1}^{\kappa} v_i}$, όταν τα δεδομένα είναι ταξινομημένα σε κατανομές συχνοτήτων.

Ορίζουμε ως ροπή t-τάξης ως προς τη μέση τιμή (t-κεντρική ροπή) τον αριθμό

$$\mu_t = \frac{\sum_{i=1}^{N} (x_i - \mu)^t}{N}$$

όταν τα δεδομένα είναι αταξινόμητα, και τον αριθμό $\mu_t = \frac{\sum_{i=1}^{\kappa} v_i (x_i - \mu)^t}{\sum_{i=1}^{\kappa} v_i}$, όταν τα δεδομένα είναι ταξινομημένα σε κατανομές συχνοτήτων.

Είναι φανερό ότι:

- Η μηδενική κεντρική ροπή $(t = 0)$ είναι $\mu_0 = 1$.
- Η πρώτη κεντρική ροπή $(t = 1)$ είναι $\mu_1 = 0$.
- Η δεύτερη κεντρική ροπή $(t = 2)$ είναι $\mu_2 = \sigma^2$, δηλαδή ίση με τη διακύμανση.
- Η τρίτη και η τέταρτη κεντρική ροπή είναι οι αριθμοί μ_3 και μ_4, οι οποίοι παρουσιάζονται στους συντελεστές ασυμμετρίας και κυρτότητας αντιστοίχως.