

## §4.4 Καρπύση του Lorenz

Η καρπύση συχέντρωσης ή καρπύση Lorenz (concentration curve or Lorenz curve) χρησιμοποιείται για τη σύγκριση κατανομών σχετικά με τα εισοδήματα, τους μισθούς, την απασχόληση ανθρώπινου δυναμικού σε οργανισμούς ή επιχειρήσεις κλπ.

Θα περιγράψουμε την τεχνική απεικόνισης της καρπύσης συχέντρωσης χρησιμοποιώντας επίσημα στατιστικά στοιχεία σχετικά με τις καταναλωτικές δαπάνες των νοικοκυριών που προέκυψαν από την έρευνα της ΕΣΥΕ για τους οικογενειακούς προϋπολογισμούς της περιόδου 1993-1994 (Πίνακας 4.15).

Ο σχεδιασμός της καρπύσης του Lorenz (Σχήμα 4.9) γίνεται με τη βοήθεια των ορθογωνίων αξόνων απολυνόμενων τα εξής βήματα:

- Στον οριζόντιο άξονα του συστήματος των ορθογωνίων συντεταγμένων απεικονίζουμε το ποσοστό του συνόλου των νοικοκυριών (Στήλη 4) και στον κάθετο άξονα το ποσοστό του συνόλου των δαπανών (Στήλη 5). Οι κλίμακες μέτρησης και των δύο αξόνων ξεκινούν από το 0 και καταλήγουν στο 100%.
- Χαράσουμε τη διαγώνιο ισοκατανομής του τετραγώνου ενώνοντας την κάτω αριστερή με την άνω δεξιά γωνία του σχήματος.
- Προσδιορίζουμε στο επίπεδο διαδοχικά σημεία τα

οποία έχουν ως τετμημένη τις αθροιστικές συχνότητες -43-  
της Στήλης 4 και ως τεταγμένη τις αθροιστικές  
συχνότητες της Στήλης 5.

- Η καρπύλη συσχέτισης σχηματίζεται συνδέοντας τα  
σημεία αυτά μεταξύ τους ώστε να δημιουργηθεί μία πο-  
λυγωνική ή μία σχετικά ομαλή γραμμή.

Η καρπύλη του Lorenz απεικονίζει γραφικά τον βαθμό  
συσχέτισης και τη μεταβλητότητα των μετρήσεων σε  
σχέση με το υπό μελέτη χαρακτηριστικό. Αν, για παρά-  
δειγμα, όλα τα νοικοκυριά δαπανούσαν το ίδιο ποσό για  
αγορές, τότε η καρπύλη συσχέτισης θα συνέπιπτε  
με τη διαγώνιο ισοκατανομής κάτι το οποίο θα σήμαινε  
απουσία ανισοτήτων ως προς την κατανόη της αξίας  
των καταναλωτικών αγορών. Όσο η ανισότητα επτείνεται,  
τόσο η καρπύλη συσχέτισης απομακρύνεται από  
τη διαγώνιο ισοκατανομής. Στην περίπτωση αυτή έχουμε  
αυξημένη μεταβλητότητα των δαπανών για τις αγορές μεταξύ  
των υπό μελέτη νοικοκυριών.

## Μέτρα κεντρικής τάσης και θέσης

### § 5.1 Μέτρα Περιγραφικής Στατιστικής

Στο κεφάλαιο αυτό θα εντοπίσουμε και θα υπολογίσουμε κάποιες τυπικές τιμές οι οποίες συμπυκνώνουν ποσοτικά τη συμπεριφορά των δεδομένων.

Όπως έχουμε αναφέρει συνήθως μελετάμε στατιστικές πληροφορίες που προέρχονται από κάποιο δείγμα, δηλαδή ένα τμήμα ενός ευρύτερου συνόλου μετρήσεων, του πληθυσμού. Ο πληθυσμός χαρακτηρίζεται από ιδιότητες που περιγράφουν τη φύση και τη δομή του και συνήθως εκφράζονται από κάποια αριθμητικά μεγέθη, τις πληθυσμιακές παραμέτρους οι οποίες υπολογίζονται στην περίπτωση που είναι γνωστές όλες οι μετρήσεις του πληθυσμού. Επιλέγοντας ένα δείγμα συλλέγουμε πληροφορίες που αναφέρονται στις ιδιότητες του πληθυσμού από τον οποίον προέρχονται και μπορούμε να επιτηρήσουμε από τα στοιχεία του δείγματος. Οι τιμές αυτές που συμπυκνώνουν τις πληροφορίες ενός πολυπληθούς συνόλου μετρήσεων καλούνται στατιστικά περιγραφικά μέτρα (descriptive measures).

Για ένα σύνολο ποσοτικών δεδομένων, το στατιστικό περιγραφικό μέτρο είναι ένας αριθμός που υπολογίζεται από τα διαθέσιμα στοιχεία και παρέχει πληροφορίες σχετικά με τη μορφή και τη δομή των δεδομένων.

Συνήθως τα δεδομένα εμφανίζουν τάση να συγκεντρώνονται γύρω από συγκεκριμένο σημείο της κλίμακας μέτρησης των μεταβλητών. Η έννοια του μέσου όρου (average

γε) είναι οιμεία. Το μέσο ημερολόγιο, το κατά κεφαλήν εισόδημα, η κατά κεφαλήν κατανάλωση αποτελούν παραδείγματα των πιο γνωστών περιγραφικών μέτρων αυτού του είδους, που χαρακτηρίζουν τα αντίστοιχα οικονομικά μεγέθη. Άλλα παραδείγματα γνώριμων μέτρων είναι η μέση διάρκεια ζωής, η μέση ηλικία γάμου και άλλα, θέματα με τα οποία ασχολούνται κυρίως οι δημογράφοι.

Διακρίνουμε πέντε είδη στατιστικών μέτρων που αντιστοιχούν σε ισάριθμους τρόπους περιγραφής της κατανομής των δεδομένων:

- (α) Τα δεδομένα εμφανίζουν τάση να έχουν τιμές γύρω από μία κεντρική τιμή η οποία εκφράζει τη μέση τιμή τους. Καλούνται μέτρα κεντρικής τάσης. Σ' αυτά υπάγονται ο αριθμητικός, ο γεωμετρικός και ο αρμονικός μέσος.
- (β) Ορισμένα μέτρα που έχουν σκοπό να εντοπίσουν την θέση της κατανομής κατά μήκος του άξονα των τιμών της μεταβλητής. Καλούνται μέτρα θέσης και σ' αυτά περιλαμβάνονται η επιμερατώσα τιμή, η διάμεσος και τα ποσοστημόρια.
- (γ) Μέτρα που αξιολογούν ποσοτικά την έκταση κατά την οποία οι τιμές μιας μεταβλητής συγκεντρώνονται ή διασκορπίζονται γύρω από τη κεντρική της τιμή. Καλούνται μέτρα διασποράς ή διαμόρφωσης και σ' αυτά περιλαμβάνεται η διαμόρφωση, η τυπική και η μέση απόκλιση, ο συντελεστής μεταβλητότητας.
- (δ) Το είδος και ο βαθμός θετικής ή αρνητικής ασυμμετρίας προσδιορίζεται από τα μέτρα λοξότητας ή

ασυμμετρίας. Γνωστότερο είναι ο συντελεστής ασυμμετρίας του Pearson.

(ε) Η κατανομή των τιμών της μεταβλητής γύρω από την μετριά της τιμή σε σχέση με τις ακραίες τιμές της προσδιορίζει τη μορφή της μονοκόρυφης καμπύλης ως προς την αιχμηρότητα της. Αυτή μετράται με τα μέτρα κύρτωσης.

§6.2 Απλός αριθμητικός Μέσος

Εάν σε πληθυσμό με μέγεθος  $N$  είναι διαθέσιμες οι τιμές  $x_1, \dots, x_N$  μιας ποσοτικής μεταβλητής  $X$ , τότε ο πληθυσμιακός μέσος ή μέσος του πληθυσμού είναι:

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$
 Στην περίπτωση  $n$  μετρήσεων,  $x_1, \dots, x_n$  που προέρχονται από ένα δείγμα, ο δειγματικός μέσος (sample mean) συμβολίζεται με  $\bar{x}$  και είναι: 
$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Η τιμή του εκφράζεται σε μονάδες μέτρησης της μεταβλητής. Παρουσιάζει το μειονέκτημα ότι επηρεάζεται από ακραίες - απορριπτόμενες τιμές (έκτροπες παρατηρήσεις).

Παράδειγμα: Βάρος 5 παιδιών σε κιλά 14, 6, 8, 10, 12

$$\bar{x} = \frac{\sum_{i=1}^5 x_i}{5} = 10.$$
 Οι αποκλίσεις των τιμών της μεταβλητής από τον αριθμητικό τους μέσο έχουν άθροισμα μηδέν. 
$$(14-10) + (6-10) + (8-10) + (10-10) + (12-10) = 4 + (-4) + (-2) + 0 + 2 = 0$$

(1) Αν  $x_1 = x_2 = \dots = x_n = a \Rightarrow \bar{x} = a$

Διότι  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{a + a + \dots + a}{n} = \frac{na}{n} = a$

(2)  $x_{(\min)} \leq \bar{x} \leq x_{(\max)}$

Είναι για κάθε  $i$ ,  $x_i \leq x_{(\max)} \Rightarrow \sum_{i=1}^n x_i \leq \sum_{i=1}^n x_{(\max)}$

$\Rightarrow \sum_{i=1}^n x_i \leq n x_{(\max)} \Rightarrow \frac{\sum_{i=1}^n x_i}{n} \leq \frac{n x_{(\max)}}{n} \Rightarrow \bar{x} \leq x_{(\max)}$

Ομοίως  $x_{(\min)} \leq \bar{x}$ .

(3)  $\sum_{i=1}^n (x_i - \bar{x}) = 0$  Διότι  $\sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = \sum_{i=1}^n x_i - n\bar{x} =$

$= n\bar{x} - n\bar{x} = 0$

(4)  $\min \sum_{i=1}^n (x_i - a)^2 \Rightarrow a = \bar{x}$ , διότι

$\frac{\partial}{\partial a} \sum_{i=1}^n (x_i - a)^2 = 2(-1) \sum_{i=1}^n (x_i - a) = 0 \Rightarrow$

$\Rightarrow \sum_{i=1}^n (x_i - a) = 0 \Rightarrow \sum_{i=1}^n x_i - na = 0 \Rightarrow a = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

(5) Αν οι πετρήσεις  $x_1, x_2, \dots, x_n$  έχουν μέσο  $\bar{x}$  οι τιμές

$x_1 + a, x_2 + a, \dots, x_n + a$  έχουν μέσο  $\bar{x} + a$ , διότι

$\frac{\sum_{i=1}^n (x_i + a)}{n} = \frac{\sum_{i=1}^n x_i + \sum_{i=1}^n a}{n} = \frac{\sum_{i=1}^n x_i}{n} + \frac{na}{n} = \bar{x} + a$

(6) Αν οι πετρήσεις  $x_1, x_2, \dots, x_n$  έχουν μέσο  $\bar{x}$  τότε οι τιμές

$b x_1, b x_2, \dots, b x_n$  έχουν μέσο  $b \bar{x}$ .

(7) Έστω  $K$  σύνολο δεδομένων. Το πρώτο,  $j=1$ , περιέχει  $n_1$  μετρήσεις με αριθμητικό μέσο  $\bar{x}_1, \dots$ , το τελευταίο,  $j=K$ , περιέχει  $n_K$  μετρήσεις με αντίστοιχο αριθμητικό μέσο  $\bar{x}_K$ . Ο αριθμητικός μέσος των τιμών όλων των μετρήσεων  $n = n_1 + \dots + n_K$  είναι:

$$\bar{x} = \frac{n_1 \bar{x}_1 + \dots + n_K \bar{x}_K}{n_1 + n_2 + \dots + n_K} = \frac{\sum_{j=1}^K n_j \bar{x}_j}{n}, \text{ διότι}$$

$$\bar{x} = \frac{\sum_{i=1}^{n_1 + \dots + n_K} x_i}{n_1 + \dots + n_K} = \frac{\sum_{i=1}^{n_1} x_i + \dots + \sum_{i=1}^{n_K} x_i}{n_1 + \dots + n_K} =$$

$$= \frac{n_1 \frac{\sum_{i=1}^{n_1} x_i}{n_1} + \dots + n_K \frac{\sum_{i=1}^{n_K} x_i}{n_K}}{n_1 + n_2 + \dots + n_K} \Rightarrow$$

$$\Rightarrow \bar{x} = \frac{n_1 \bar{x}_1 + \dots + n_K \bar{x}_K}{n}$$

Προφανώς οι παραπάνω ιδιότητες ισχύουν και για τον πληθυσμιακό μέσο.

Θα περιγράψουμε την τεχνική απεικόνισης της καμπύλης συγκέντρωσης χρησιμοποιώντας επίσημα στατιστικά στοιχεία σχετικά με τις καταναλωτικές δαπάνες των νοικοκυριών, όπως αυτές προέκυψαν από την Έρευνα Οικογενειακών Προϋπολογισμών της περιόδου 1993/1994. Την έρευνα διεξήγαγε η ΕΣΥΕ σε δείγμα που κάλυπτε περίπου το 96% των νοικοκυριών της χώρας (Πίνακας 4.15).

**ΠΙΝΑΚΑΣ 4.15**

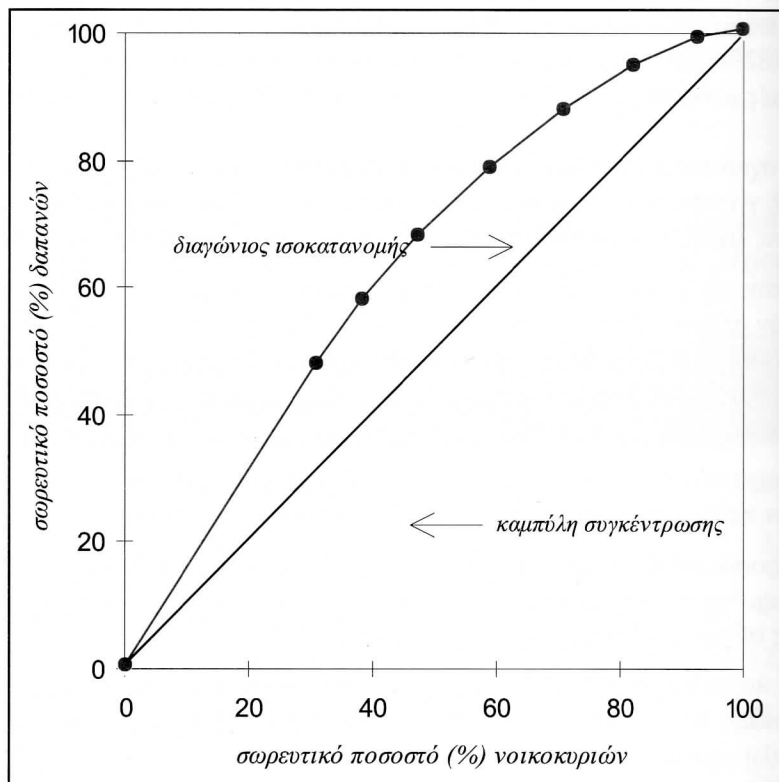
*Μηνιαίες δαπάνες καταναλωτικών ειδών των νοικοκυριών κατά κλιμάκια αξίας αγορών: Έρευνα Οικογενειακών Προϋπολογισμών 1993/1994*

| <i>Αξία<br/>Αγορών<br/>(χιλ. δραχ.)</i> | <i>Αριθμός<br/>Νοικοκυριών</i> | <i>Αγορές<br/>(εκατ. δραχ.)</i> | <i>Αθροιστικό<br/>Ποσοστό (%)<br/>Νοικοκυριών</i> | <i>Αθροιστικό<br/>Ποσοστό (%)<br/>Αγορών</i> |
|-----------------------------------------|--------------------------------|---------------------------------|---------------------------------------------------|----------------------------------------------|
| (1)                                     | (2)                            | (3)                             | (4)                                               | (5)                                          |
| - < 100                                 | 502                            | 25,1                            | 7,43                                              | 1,27                                         |
| 100-150                                 | 686                            | 85,7                            | 17,58                                             | 5,62                                         |
| 150-200                                 | 774                            | 135,5                           | 29,04                                             | 12,49                                        |
| 200-250                                 | 796                            | 179,1                           | 40,82                                             | 21,57                                        |
| 250-300                                 | 784                            | 215,6                           | 52,43                                             | 32,51                                        |
| 300-350                                 | 612                            | 198,9                           | 61,49                                             | 42,59                                        |
| 350-400                                 | 517                            | 193,9                           | 69,14                                             | 52,42                                        |
| 400 +                                   | 2085                           | 938,2                           | 100,0                                             | 100,00                                       |
| Σύνολο                                  | 6756                           | 1972                            |                                                   |                                              |

Η πρώτη στήλη του πίνακα περιέχει τις τάξεις μηνιαίων αγορών η αξία των οποίων εκφράζεται σε χιλιάδες δραχμές. Για κάθε τάξη δίδεται ο αριθμός των νοικοκυριών (στήλη 2) και οι αντίστοιχες δαπάνες αγορών του κλιμακίου σε εκατομμύρια δραχμές (στήλη 3). Υπολογίζουμε τις δεξιόστροφες σχετικές αθροιστικές συχνότητες των νοικοκυριών (στήλη 4)

**ΣΧΗΜΑ 4.9**

Καμπύλη συγκέντρωσης των στοιχείων του Πίνακα 4.15

**4.7 Ειδικά Θέματα**

4.1 Στον πίνακα που ακολουθεί δίδεται η κατά ηλικία κατανομή 1275 ατόμων:

| Ηλικίες | $f_i$ | $f_i^* = 1,02f_i$ |
|---------|-------|-------------------|
| 15 - 24 | 70    | 71                |
| 25 - 34 | 150   | 153               |
| 35 - 44 | 230   | 235               |

45 - 54

55 - 64

Δεν Δήλωσαν

Σύνολο

Επί συνόλου 1275

Το φαινόμενο αυτό π  
 την οποία η συχνότη  
 χαρακτηριστικό είναι μι  
 χιότητας αυτής σε όλ

- Ορίζουμε τον διορ  
 θους των παρατημ  
 οποίες είναι γνωσ

όπου  $n$  το συνολικ  
 που δεν δήλωσαν  
 του ανωτέρω πίνα

- οι διορθωμένες  $c$   
 των αρχικών συχ  
 κάθε περίπτωση

Οι διορθωμένες  $c$   
 στην τελευταία στήλ  
 νομής των  $f_{(ΔΔ)}$  μπορε  
 περιγράψαμε δίνει  $c$   
 για των  $f_{ΔΔ}$  δεν είναι

4.2 Στην περίπτω  
 τα κάτω ή προς τα ά