

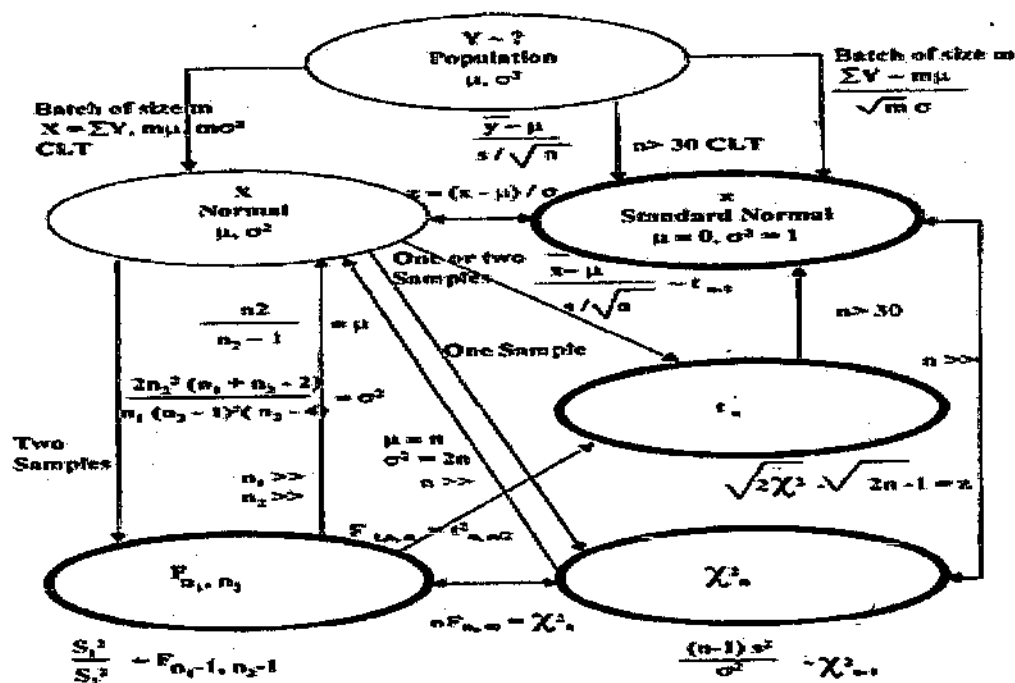
Α.Τ.Ε.Ι ΠΑΤΡΩΝ

ΣΧΟΛΗ: ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ: ΛΟΓΙΣΤΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΘΕΜΑ: ΣΤΑΤΙΣΤΙΚΗ : ΕΡΓΑΛΕΙΟ ΓΙΑ ΤΗΝ ΑΝΑΠΤΥΞΗ ΕΠΙΧΕΙΡΗΣΕΩΝ



Relationship Among Statistical Tables and Their Applications

ΚΑΘΗΓΗΤΗΣ : ΦΩΤΟΠΟΥΛΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ

ΣΠΟΥΔΑΣΤΕΣ : ΣΚΟΤΑΔΗ ΜΑΡΙΑ
ΠΟΥΛΟΥ ΣΠΥΡΙΔΟΥΛΑ

ΠΑΤΡΑ 2004

ΑΡΙΘΜΟΣ ΕΙΣΑΓΩΓΗΣ	6072
----------------------	------

Αντί προλόγου

Η παρούσα εργασία έχει ως αντικείμενο ένα θέμα πολύ γενικό και έτσι είχαμε πολλές συζητήσεις σχετικά με το τι ακριβώς θα καταπιανόταν.

Μετά από πολύ σκέψη επειδή θεωρήσαμε ότι το σπλοστάσιο ως προς την στατιστική με το οποίο ένας σπουδαστής του τμήματος εφοδιάζεται μέσα σ' ένα εξάμηνο διδασκαλίας της στατιστικής είναι πολύ ελλιπές σκεφτήκαμε να ασχοληθούμε με την παρουσίαση (εξ απαλών ονύχων βέβαια) κάποιων τεχνικών και μεθόδων που έχει η στατιστική στην υπηρεσία τόσο της επιχείρησης όσο και της ερευνάς γενικότερα .

έτσι η παρούσα εργασία αποτελείται ουσιαστικά από τρία μέρη 1) τα γενικά περί στατιστικής 2) Τις τεχνικές και μεθόδους της 3) Την παρουσίαση ενός στατιστικού πακέτου (SPSS) από τα πολλά που υπάρχουν στην χρήση του στατιστικού η του οποιουδήποτε κάνει χρήση της στατιστικής σαν εργαλείο λήψης αποφάσεων το πρόγραμμα που επελέγη είναι το δημοφιλέστερο της αγοράς.

Τέλος ελπίζουμε αυτή η εργασία να βοηθήσει όποιον θα ήθελε να ενημερωθεί για αυτό το πολύ χρήσιμο εργαλείο ,την στατιστική



ΜΕΡΟΣ ΠΡΩΤΟΝ

1) ΣΤΑΤΙΣΤΙΚΗ: ΕΝΝΟΙΑ

Ανάλογα με τη σκοπιά της θεώρησης και τους σκοπούς της ανάλυσης έχουν διατυπωθεί διάφοροι ορισμοί της Στατιστικής ως επιστήμης. Ο ορισμός της Στατιστικής πρέπει να περιλαμβάνει τουλάχιστον τα εξής συστατικά στοιχεία: 1) τη συλλογή, επεξεργασία και παρουσίαση πληροφοριών (δεδομένων), 2) την αναγνώριση ότι αντικείμενο μελέτης της Στατιστικής αποτελούν φαινόμενα και καταστάσεις της πραγματικής ζωής που χαρακτηρίζονται από αβεβαιότητα, 3) την ανάπτυξη θεωρητικών μοντέλων (κατασκευών) τα οποία βοηθούν στη διατύπωση προτάσεων (συμπερασμάτων) με γενική ισχύ αναφορικά με τη συμπεριφορά των φαινομένων που εξετάζονται και 4) τα συμπεράσματα της ανάλυσης πρέπει να αποτελούν χρήσιμο οδηγό στον τομέα λήψης αποφάσεων.

Στο πλαίσιο που διαμορφώνεται από τις παραπάνω βασικές συνιστώσες του ορισμού, θα ορίσουμε την Στατιστική ως την επιστήμη που έχει ως αντικείμενο: α) Την ανάπτυξη τεχνικών για τη συγκέντρωση, επεξεργασία και κατάλληλη παρουσίαση πληροφοριών (δεδομένων) σχετικών με τις ιδιότητες και τη συμπεριφορά πραγματικών φαινομένων, β) τη διαμόρφωση μεθοδολογίας για την ανάλυση των πληροφοριών αυτών, γ) το πάντρεμα των εμπειρικών δεδομένων και των θεωρητικών μοντέλων με τελικό στόχο την εξαγωγή συμπερασμάτων χρήσιμων για την κατανόηση των νόμων που διέπουν τα φαινόμενα της φύσης και της ζωής, αλλά και για τη λήψη αποφάσεων με πρακτικό περιεχόμενο.

Πιο αναλυτικά η επιστήμη της Στατιστικής ασχολείται με τα ακόλουθα θέματα:

1. Με τη συλλογή, την επεξεργασία και τη συμπυκνωμένη παρουσίαση πληροφοριών (δεδομένων) που αναφέρονται στις ιδιότητες και στην εξέλιξη φαινομένων τα οποία ενδιαφέρουν τον άνθρωπο.

2. Τη μεθοδολογία σχεδιασμού και διεξαγωγής πειραμάτων και ειδικών ερευνών (π.χ. πειράματα τύχης, δειγματοληπτικές έρευνες, έρευνες προσδοκίων κ.λ.π.).

3. Τη μέτρηση της μεταβλητότητας στις τιμές των δεδομένων που συγκεντρώνονται με πειραματικές μεθόδους ή ειδικές έρευνες.

4. Την ανάπτυξη μεθοδολογίας για την εκτίμηση των παραμέτρων του πληθυσμού με την βοήθεια στοιχείων δείγματος και την κατάρτιση δεικτών (μέτρων) αξιολόγησης της ακρίβειας των εκτιμήσεων. Ο όρος **πληθυσμός** στη Στατιστική εκφράζει το σύνολο των δυνατών απαριθμήσεων ή μετρήσεων που αφορούν τις επιμέρους μονάδες ενός καλά ορισμένου πλήθους διακεκριμένων όντων ή αντικειμένων, τα οποία έχουν ένα ή περισσότερα κοινά χαρακτηριστικά. Με άλλα λόγια, ο όρος **πληθυσμός** δεν αναφέρεται στις ίδιες τις μονάδες του πλήθους των όντων ή των πραγμάτων που είναι αντικείμενο στατιστικής μελέτης, αλλά στις μετρήσεις ή τις εν γένει παρατηρήσεις τις σχετικές με κάποιο χαρακτηριστικό ή κάποια ιδιότητα των μονάδων αυτών. Για παράδειγμα, αν αντικείμενο της έρευνας είναι η οικιακή κατανάλωση ηλεκτρικού ρεύματος στην περιοχή της Αθήνας, στατιστικό πληθυσμό αποτελεί το σύνολο των μετρήσεων (των αριθμητικών δεδομένων) που αναφέρονται στην κατανάλωση ρεύματος από τα επιμέρους νοικοκυριά στη διάρκεια ενός μηνός ή ενός έτους (ανάλογα με τη χρονική μονάδα μέτρησης που έχει επιλεγεί). Τα ίδια τα νοικοκυριά ως τέτοια δεν αποτελούν μέρος του πληθυσμού, αποτελούν τις μονάδες του πλήθους που μελετάται. Τα δεδομένα (μετρήσεις ή παρατηρήσεις) που χρησιμοποιούνται στις στατιστικές έρευνες σπάνια καλύπτουν το σύνολο των μονάδων ενός πεπερασμένου πολυπληθούς πληθυσμού. Οι λόγοι ανάγονται κυρίως στις υψηλές δαπάνες και στο μεγάλο χρόνο που απαιτούνται για τη συγκέντρωση παρατηρήσεων που να αναφέρονται σε όλες τις μονάδες ενός πολυάριθμου πληθυσμού. Βεβαίως, στις περιπτώσεις των απείρων στατιστικών πληθυσμών η πλήρης κάλυψη είναι αδύνατη. Για τους λόγους αυτούς η εφαρμοσμένη έρευνα, κατά κανόνα, περιορίζεται σε ένα υποσύνολο του συνόλου των μονάδων που αποτελούν τον πληθυσμό. Το υποσύνολο αυτό ονομάζεται **δείγμα** και αντιπροσωπεύει συνήθως πολύ μικρό ποσοστό του αρχικού συνόλου (π.χ. 10%, 5%, 2% ή μικρότερα). Έτσι με βάση τις πληροφορίες που περιέχονται στο δείγμα προσπαθούμε να εξαγάγουμε συμπεράσματα για το σύνολο του πληθυσμού. Είναι συνήθης η πρακτική να επιχειρείται η εξαγωγή συμπερασμάτων αναφορικά με την εκλογική συμπεριφορά ενός πληθυσμού με τη βοήθεια πολύ μικρού δείγματος (τα γνωστά *gallor*). Επειδή οι παρατηρήσεις του δείγματος αποτελούν την μόνη πληροφορία που διαθέτει ο ερευνητής για τον πληθυσμό, το δείγμα πρέπει να είναι καλά σχεδιασμένο και αντιπροσωπευτικό. Για το σκοπό αυτό έχει αναπτυχθεί ειδική μεθοδολογία που είναι γνωστή στη Στατιστική ως **δειγματοληψία**. Αξίζει να τονιστεί ότι η αντιπροσωπευτικότητα και το μέγεθος του δείγματος παίζουν κρίσιμο ρόλο στην εξαγωγή αξιόπιστων συμπερασμάτων για τον πληθυσμό.

5.Τη διαμόρφωση μεθοδολογίας ελέγχου υποθέσεων (θεωριών) σχετικών με τον πληθυσμό που ερευνάται και τη διατύπωση προτάσεων πιθανότητας με γενική ισχύ.

6.Την ανάπτυξη μεθόδων για την ταυτόχρονη μελέτη της εξέλιξης των τιμών δύο ή περισσότερων μεταβλητών.

7.Τη διαμόρφωση μεθοδολογίας για τη συνδυασμένη χρησιμοποίηση στοιχείων δείγματος και εκ των προτέρων πληροφόρησης (prior information).

8.Την ανάπτυξη κανόνων (rules) για την λήψη αποφάσεων σε καταστάσεις που χαρακτηρίζονται από συνθήκες αβεβαιότητας.

Από τα παραπάνω είναι προφανές ότι η Στατιστική καλύπτει ένα μεγάλο φάσμα γνώσεων και εφοδιάζει τον ερευνητή με πλούσιο απόθεμα αναλυτικών εργαλείων, τα οποία του επιτρέπουν να μελετά δύσκολα προβλήματα και να καταλήγει σε προσδιορίσιμης αξιοπιστίας συμπεράσματα αναφορικά με τους παράγοντες που προσδιορίζουν τη συμπεριφορά διαφόρων φαινομένων. Ακόμη, η γνώση αυτή βοηθάει τους ενδιαφερόμενους να παίρνουν αποφάσεις σε καθεστώς αβεβαιότητας.

Οι εφαρμογές της Στατιστικής επεκτείνονται σε όλους τους τομείς της ανθρώπινης δράσης που για τη λύση προβλημάτων και την κατανόησή της εξέλιξης των φαινομένων απαιτείται η χρησιμοποίηση δεδομένων και η εκτέλεση πειραμάτων. Έτσι, πέρα από τα φαινόμενα του οικονομικο-διοικητικού χώρου, όπου οι εφαρμογές της Στατιστικής είναι πολυάριθμες, σημαντικές εφαρμογές έχουμε στην ιατρική, στη βιολογία, στη φυσική και στη χημεία, στην κτηνοτροφία, στη γεωργία, στη μετεωρολογία, στην αστρονομία, στις τεχνολογικές επιστήμες, στη δημόσια διοίκηση, στη διερεύνηση της στάσης και των προθέσεων των ανθρώπων, κ.λπ.

Δεδομένου ότι οι μέθοδοι και οι τεχνικές της Στατιστικής, σωστά αξιοποιούμενες, είναι σε θέση να βοηθήσουν αποτελεσματικά όλους τους επιστήμονες που ασχολούνται με την εφαρμοσμένη κυρίως έρευνα, καθώς και όσους λαμβάνουν αποφάσεις στο επίπεδο των επιχειρήσεων και των δημόσιων οργανισμών, είναι χρήσιμο όλοι τους να εξοικιώνονται με τη βασική στατιστική μεθοδολογία και τις τεχνικές ανάλυσης των δεδομένων.

Είναι παρήγορο το γεγονός ότι αυτό συμβαίνει σήμερα σε σημαντικό βαθμό. Πράγματι, από απλή επισκόπηση των προγραμμάτων σπουδών πολλών επιστημονικών κλάδων (π.χ. γιατρών, μηχανικών, γεωπόνων, ψυχολόγων, κοινωνιολόγων, κ.λπ.) συνάγεται ότι σε αυτά περιλαμβάνεται ένα ή περισσότερα μαθήματα Στατιστικής. Εκεί όμως που η διδασκαλία της Στατιστικής κατέχει κεντρική θέση είναι στα προγράμματα εκπαίδευσης των οικονομικο-διοικητικών επιστημών και γενικότερα των επιστημών λήψης αποφάσεων.

Σε σχέση με τη συμμετοχή της Στατιστικής στην εφαρμοσμένη έρευνα, είναι χρήσιμο να τονιστούν τα εξής :

Πρώτον, λόγω της ιδιαιτερότητας που εμφανίζει η έρευνα στους επιμέρους τομείς, σε πολλές περιπτώσεις έχουν αναπτυχθεί εξειδικευμένες προσεγγίσεις και τεχνικές για την αντιμετώπιση ειδικών προβλημάτων. Έτσι, η έμφαση που δίνεται στη διδασκαλία της Στατιστικής συχνά διαφοροποιείται ανάλογα με τις ειδικότερες απαιτήσεις του επιστημονικού χώρου στον οποίο απευθύνεται. Ωστόσο, ο κορμός των βασικών γνώσεων της Στατιστικής επιστήμης είναι κοινός για όλους τους επιστήμονες.

Δεύτερον, επειδή η Στατιστική αποτελεί σημαντικό αναλυτικό εργαλείο, συχνά γίνεται κατάχρηση της εφαρμογής στατιστικών μεθόδων στην προσπάθεια των ατόμων να ερμηνεύσουν τα αποτελέσματα μη καλά προσδιορισμένων και πλημελώς οργανωμένων ερευνών ή πειραμάτων.

Τρίτον, απαιτείται γνώση του χώρου εφαρμογής, μεγάλη προσοχή και ειλικρίνεια για τη σωστή ερμηνεία και τη χρησιμοποίηση των αποτελεσμάτων της στατιστικής ανάλυσης. Υπάρχουν περιπτώσεις που οι ερευνητές εφαρμόζουν με καλή πίστη στατιστικές μεθόδους, χωρίς όμως να προσέχουν στο βαθμό που χρειάζεται την ισχύ και τη σημασία των υποθέσεων που βρίσκονται πίσω από τις μεθόδους αυτές.

Οι πιθανότητες και η στατιστική έχουν γνωρίσει μεγάλη ανάπτυξη τα τελευταία 30-40 χρόνια και δεν υπάρχει σήμερα τομέας της ανθρώπινης δραστηριότητας (επιστημονικός ή επαγγελματικός) που να μην τις χρησιμοποιεί.

2) ΠΡΟΕΛΕΥΣΗ ΚΑΙ ΕΝΑΛΛΑΚΤΙΚΕΣ ΧΡΗΣΕΙΣ ΤΟΥ ΟΡΟΥ «ΣΤΑΤΙΣΤΙΚΗ»

Ο όρος «στατιστική» (statistica) προέρχεται από την λατινική λέξη status που σημαίνει κράτος (πολιτεία) και καθιερώθηκε στη βιβλιογραφία από τον Γερμανό καθηγητή στο πανεπιστήμιο της Γοττίνγης Gottfried Achenwall (1719-1772). Αρχικά ο όρος σήμαινε τη συλλογή αριθμητικών δεδομένων, τα οποία αναφέρονταν στη λειτουργία του κράτους (π.χ. πληθυσμός, έκταση της χώρας και χρήση αυτής, στην αριθμητική δύναμη του στρατού, στην είσπραξη φόρων, στην παραγωγή διάφορων προϊόντων, κ.λπ). Οι πρώτες εφαρμογές της Στατιστικής άρχισαν από το χώρο της δημογραφίας (εξέλιξη πληθυσμού, γεννήσεων, θανάτων, κατανομή πληθυσμού κατά ηλικία και περιφέρεια, κ.λπ). Η απόδοση του όρου Στατιστική σε άλλες γλώσσες γίνεται με τις λέξεις statistics (Αγγλικά), statistique (Γαλλικά), statistik (Γερμανικά) και statistica (Ιταλικά).

Αξίζει να αναφερθεί, ότι συλλογή αριθμητικών δεδομένων για τις ανάγκες του κράτους έκαναν όλοι οι αρχαίοι λαοί (Αιγύπτιοι, Κινέζοι, Ασσύριοι, Έλληνες, Ρωμαίοι, κ.λπ). Όμως, η εμφάνιση της Στατιστικής ως ξεχωριστού κλάδου της επιστήμης χρονολογείται από τα μέσα περίπου του

17^ο αιώνα και τις αρχές του 18^ο. Αρχικά η στατιστική ανάλυση είχε περιγραφικό χαρακτήρα. Αργότερα, με την ανάπτυξη του Λογισμού των Πιθανοτήτων και την εισαγωγή των Μαθηματικών ως εργαλείου ανάλυσης στατιστικών δεδομένων, άρχισε να αναπτύσσεται σταδιακά ο κλάδος της στατιστικής συμπερασματολογίας (στατιστική επαγωγή = statistical inference) και να διαμορφώνεται η σύγχρονη στατιστική θεωρία. Βεβαίως, αν μελετήσει κανείς τη διαχρονική εξέλιξη της Στατιστικής, σύντομα θα διαπιστώσει ότι το μεγαλύτερο απόθεμα γνώσεων του τομέα της Στατιστικής θεωρίας δημιουργήθηκε κατά τον 19^ο αιώνα και κυρίως κατά τον 20^ο αιώνα, παρουσιάζοντας αιχμή στη μεταπολεμική περίοδο.

Ο όρος Στατιστική (statistics) χρησιμοποιείται στη βιβλιογραφία με τρεις διαφορετικές έννοιες. Η πρώτη χρήση του όρου αφορά στη Στατιστική ως επιστήμη, η οποία έχει ως περιεχόμενο αυτό που αναφέρθηκε πιο πάνω. Όμως, η ίδια λέξη χρησιμοποιείται για να υποδηλώσει σειρά αριθμητικών δεδομένων. Έτσι δημιουργήθηκαν οι εκφράσεις «Στατιστική του Πληθυσμού» που σημαίνει συλλογή στατιστικών δεδομένων σχετικών με την κίνηση του πληθυσμού, «Στατιστική Εξωτερικού Εμπορίου» που υποδηλώνει σειρές αριθμητικών δεδομένων, τα οποία αναφέρονται στις εισαγωγές και εξαγωγές μιας χώρας, «Τουριστική Στατιστική» που σημαίνει ταξινομημένη συγκέντρωση αριθμητικών δεδομένων σχετικών με την τουριστική κίνηση μιας χώρας (π.χ. αριθμός τουριστών, προέλευση τουριστών ανά χώρα, χρόνος παραμονής στη χώρα, ποσό συναλλάγματος που δαπανάται κ.λπ.).

Τέλος, ο όρος Στατιστική χρησιμοποιείται ως δείκτης ή μέτρο (measure) το οποίο εκφράζει κατά τρόπο συνοπτικό κάποιο χαρακτηριστικό ενός πληθυσμού και υπολογίζεται με στοιχεία δείγματος. Για παράδειγμα το μέγεθος \bar{X} , που αποτελεί εκτίμηση του μέσου αριθμητικού ενός πληθυσμού, ονομάζεται **Στατιστική** (statistic). Ο όρος Στατιστική με την έννοια αυτή υποδηλώνει τον μαθηματικό τύπο από τον οποίο εκτιμάται μία άγνωστη παράμετρος του πληθυσμού με στοιχεία δείγματος. Στην περίπτωση του μέσου ο σχετικός τύπος είναι:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Όπου τα X_i ($i = 1, 2, 3, \dots, n$) αποτελούν τις παρατηρήσεις, το δε n παριστάνει το μέγεθος του δείγματος. Ο τύπος είναι γνωστός και ως μια Στατιστική (statistic). Οι στατιστικές με την παρούσα σημασία του όρου κατέχουν κεντρική θέση στην στατιστική επαγωγή.

3) ΒΑΣΙΚΕΣ ΔΙΑΚΡΙΣΕΙΣ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ

Το γνωστικό πεδίο της Στατιστικής καλύπτει ένα ευρύτατο φάσμα αντικειμένων με αποτέλεσμα να είναι αδύνατη η συνολική και σύμμετρη διαπραγμάτευση αυτών στα πλαίσια ενός συγγράμματος. Από την άλλη μεριά, ο επιδιωκόμενος σκοπός και το είδος των πληροφοριών που πρόκειται να αναλυθούν αποτελούν παράγοντα διαφοροποίησης της μεθοδολογικής προσέγγισης που είναι απαραίτητη για τη στατιστική διερεύνηση επιμέρους προβλημάτων και φαινομένων. Οι λόγοι αυτοί οδήγησαν στην ανάπτυξη ειδικών κλάδων στον ευρύτερο χώρο της Στατιστικής.

Έτσι, ανάλογα με τον επιδιωκόμενο σκοπό η Στατιστική διακρίνεται σε **περιγραφική** και σε **επαγωγική**. Η **περιγραφική Στατιστική** (descriptive statistics) έχει ως αντικείμενο την ανάπτυξη μεθοδολογίας και τεχνικών για τη συλλογή, την επεξεργασία, την αποθήκευση και τη συστηματοποιημένη παρουσίαση του πρωτογενούς στατιστικού υλικού. Ακόμη, ασχολείται με την εξαγωγή συμπερασμάτων περιγραφικού χαρακτήρα, τα οποία αναφέρονται αποκλειστικά στο ερευνούμενο τμήμα του πληθυσμού χωρίς να προχωρεί στη διατύπωση προτάσεων γενικότερης εφαρμογής, δηλαδή προτάσεων (συμπερασμάτων), οι οποίες να αφορούν στο σύνολο του πληθυσμού ή να σχετίζονται με την πρόβλεψη της μελλοντικής πορείας του φαινομένου που μελετάται.

Η **επαγωγική Στατιστική** (inductive statistics) ή **Στατιστική συμπερασματολογίας** (statistical inference) ασχολείται με την ανάπτυξη μεθόδων και τεχνικών, οι οποίες επιτρέπουν τη διατύπωση συμπερασμάτων με γενική ισχύ αναφορικά με τη συμπεριφορά ενός φαινομένου με βάση πληροφορίες που περιέχονται σε ένα μικρό τμήμα (δείγμα) του πληθυσμού. Με άλλα λόγια, στην περίπτωση της επαγωγικής στατιστικής προσπαθούμε να διατυπώσουμε νόμους με γενικότερη ισχύ, στηριζόμενοι σε ένα πολύ μικρό τμήμα των δεδομένων που προσδιορίζουν την εξέλιξη ενός φαινομένου. Στο ίδιο εννοιολογικό πλαίσιο εντάσσεται και η προσπάθεια για τη διενέργεια προβλέψεων σχετικών με τη μελλοντική πορεία του φαινομένου και τη λήψη αποφάσεων σε συνθήκες αβεβαιότητας.

Εξάλλου, ανάλογα με το είδος των προς επεξεργασία πληροφοριών και του σκοπού της ανάλυσης, η επαγωγική στατιστική διακρίνεται στους εξής μεγάλους τομείς (α) στην **κλασική στατιστική συμπερασματολογία** (classical statistical inference), (β) στη **στατιστική επαγωγή κατά Bayes** (Bayesian inference) και (γ) στη **θεωρία αποφάσεων** (decision theory). Στην κλασική στατιστική επαγωγή ως πηγής πληροφοριών για την εξαγωγή συμπερασμάτων χρησιμοποιούνται μόνο τα δεδομένα του δείγματος, ενώ η προσέγγιση της επαγωγής κατά Bayes είναι προσανατολισμένη στην ταυτόχρονη επεξεργασία δεδομένων δείγματος και εκ των προτέρων πληροφόρησης (prior information).

Τέλος, η μεθοδολογία της θεωρίας αποφάσεων είναι σχεδιασμένη κατά τρόπο που να επιτρέπει, πέρα από τη χρησιμοποίηση στοιχείων δείγματος και προυπάρχουσας πληροφόρησης, την ενσωμάτωση στην ανάλυση και των **συνεπειών** που εκτιμάται ότι θα έχουν στο τελικό αποτέλεσμα οι εναλλακτικοί τρόποι δράσης (εναλλακτικές προτάσεις ή αποφάσεις).

Προτού κλείσουμε την παράγραφο αυτή είναι σημαντικό να αναφερθεί ότι οι βάσεις του μεθοδολογικού οπλοστασίου της Στατιστικής βρίσκονται στα Μαθηματικά και ειδικότερα στη θεωρία των Πιθανοτήτων. Ωστόσο, πρέπει να τονιστεί ότι παρά την εκτεταμένη χρήση Μαθηματικών, η **στατιστική δεν είναι Μαθηματικά**. Είναι ξεχωριστή επιστήμη εφαρμοσμένου χαρακτήρα με δικό της περιεχόμενο και μέθοδο ανάλυσης.

4) ΠΕΙΡΑΜΑΤΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Η χρησιμοποίηση της Βιοστατιστικής στην εξήγηση και κατανόηση βιολογικών φαινομένων αναμφισβήτητα έχει συμβάλει αρκετά στην ανάπτυξη των βιολογικών επιστημών. Η θέση αυτή στηρίζεται σε στατιστικά στοιχεία σύμφωνα με τα οποία το ποσοστό των δημοσιευμένων ερευνητικών εργασιών που περιλαμβάνουν και στατιστική ανάλυση των δεδομένων, παρουσιάζει δραματική αύξηση. Στη βιολογία, τα περισσότερα φαινόμενα επηρεάζονται από πολλούς παράγοντες που δεν μπορούν πολλές φορές να ελεγχθούν ή ακόμα και να επιβεβαιωθούν. Η βιοστατιστική είναι έτσι, απαραίτητη για τη μελέτη τέτοιων φαινομένων, με προκαθορισμένες αποκλίσεις, ουσιαστικοποιώντας ποιες από τις αποκλίσεις – διαφορές, έστω και μικρές, έχουν σημασία ή είναι τυχαίες, αναβαθμίζοντας έτσι την αξιολόγηση των ευρημάτων, η οποία βασιζόταν στην προσωπική πείρα.

Η προέλευση της σύγχρονης στατιστικής χρονολογείται από τον 17^ο αιώνα όπου δύο διαφορετικές πηγές γεγονότων έπαιξαν καθοριστικό ρόλο στη διαμόρφωση τους. Η πρώτη κατηγορία, που έγινε γνωστή και ως πολιτική αριθμητική, σχετίζεται με την πολιτική επιστήμη και αναπτύχθηκε ως μια ποσοτική περιγραφή των διαφόρων στοιχείων που αφορούσαν τις δραστηριότητες των κυβερνήσεων ή των πολιτειών (States, γι' αυτό και στατιστική) των ΗΠΑ. Φόροι, ασφάλειες κ.λπ. έκαναν τους πολίτες να ενδιαφέρονται για προβλήματα, μακροβιότητας, θνησιμότητας κ.α. Η δεύτερη ρίζα της σύγχρονης στατιστικής βρίσκεται στην ανάπτυξη της μαθηματικής πιθανοθεωρίας, από ενδιαφέρον για τα τυχερά παιχνίδια.

Μεγάλη πρόοδος στη στατιστική έγινε στο 19^ο αιώνα από τους μαθηματικούς. Ο Francis Galton (1822-1911) εξάδελφος του Κάρολου Δαρβίνου, είχε τιτλοδοτηθεί σαν ο πατέρας της βιοχημείας και της Ευγονικής. Ο Galton παρακινήθηκε στις έρευνες του μετά τη δημοσίευση της θεωρίας του Δαρβίνου για την καταγωγή των ειδών. Στη θεωρία αυτή υπάρχει ένα

μειονέκτημα που αφορά τη γενετική εξήγηση του μηχανισμού της φυσικής επιλογής. Ο ίδιος ο Δαρβίνος κατανοώντας αυτό το πρόβλημα, προσπάθησε να αναπτύξει δική του θεωρία κληρονομικότητας, αν και ο σύγχρονος του Mendel είχε ήδη δημοσιεύσει τους βασικούς νόμους της Γενετικής. Η γνώση αυτή δεν είχε φτάσει φαίνεται μέχρι το Δαρβίνο, ο οποίος τελικά πρότεινε μια δική του θεωρία που ήταν λαθεμένη. Έτσι, ο Galton προσπάθησε να λύσει αυτά τα περίπλοκα προβλήματα της κληρονομικότητας αρχίζοντας με πάρα πολύ δύσκολο υλικό και κάνοντας στη συνέχεια λαθεμένες υποθέσεις.

Η συμβολή του Galton, ωστόσο, στη Βιολογία είναι μεγάλη και διαμορφώθηκε μέσα από την εφαρμογή της στατιστικής μεθοδολογίας στην ανάλυση της ποικιλότητας και στη ανάπτυξη μελετών σχετικών με τη συσχέτιση βιολογικών παραμέτρων. Ο Pearson (1857-1936) προσπάθησε να αποδείξει με στατιστική μεθοδολογία την ύπαρξη της φυσικής επιλογής και θεωρείται ο ιδρυτής της περιγραφικής στατιστικής. Η επικρατούσα όμως μορφή της Στατιστικής και Βιομετρίας στον αιώνα μας, θεωρείται ο Fisher (1890-1962).

5) ΑΝΤΙΚΕΙΜΕΝΟ ΤΗΣ ΠΕΡΙΓΡΑΦΙΚΗΣ ΣΤΑΤΙΣΤΙΚΗΣ

Η περιγραφική Στατιστική πραγματεύεται τις τεχνικές και τις μεθόδους που χρησιμοποιούνται για τη συλλογή, την επεξεργασία και τη συμπυκνωμένη παρουσίαση των πληροφοριών που υπάρχουν στα στατιστικά δεδομένα. Τα δεδομένα μπορεί να αναφέρονται σε πληθυσμό ή σε δείγμα. Βασικός σκοπός της ανάλυσης είναι η συστηματοποιημένη παρουσίαση των δεδομένων σε πίνακες και διαγράμματα και η συμπύκνωση των πληροφοριών που αυτά περιέχουν σε απλούς αριθμητικούς δείκτες (μέτρα), έτσι ώστε να γίνεται γρήγορη και απλή η επικοινωνία με τους χρήστες των πληροφοριών αυτών. Η περιγραφική στατιστική δεν επεκτείνεται σε θέματα αξιολόγησης των πληροφοριών με σκοπό τη διατύπωση γενικού χαρακτήρα προτάσεων αναφορικά με τη συμπεριφορά του φαινομένου που εξετάζεται, ούτε στην αξιοποίηση των πληροφοριών για τη διενέργεια προβλέψεων σε σχέση με τη μελλοντική εξέλιξη των φαινομένων.

Η ανάλυση αρχίζει με τη συγκέντρωση και επεξεργασία μεγάλου αριθμού παρατηρήσεων που αφορούν στην εξέλιξη κάποιου φυσικού, κοινωνικού ή οικονομικού φαινομένου και σταματά όταν έχει επιτευχθεί η μεγαλύτερη δυνατή συμπύκνωση των πληροφοριών σε απλούς δείκτες. Αυτό επιτυγχάνεται όταν το σύνολο ή ο κύριος όγκος των πληροφοριών που περιέχουν τα δεδομένα εκφράζονται με πολύ περιορισμένο αριθμό δεικτών ποσοτικής φύσεως, δηλαδή με απλούς αριθμούς.

Τα δεδομένα μπορεί να αφορούν οποιοδήποτε φαινόμενο ή θέμα που ενδιαφέρει τα άτομα, τις επιχειρήσεις ή το κοινωνικό σύνολο. Ενδεικτικά αναφέρουμε ως παραδείγματα τα εξής : το βάρος και το ύψος των

ατόμων μιας κοινωνίας, τις απόψεις του πληθυσμού της Αθήνας για τους Έλληνες πολιτικούς ή τα ναρκωτικά, τη βιομηχανική παραγωγή, την κατανάλωση νερού από τους κατοίκους του λεκανοπεδίου της Αττικής, το εξωτερικό εμπόριο της Ελλάδας, τις επιδόσεις των υποψηφίων φοιτητών κατά τις γενικές εξετάσεις, τη ρύπανση του περιβάλλοντος, τους θανάτους από αυτοκινητιστικά ατυχήματα ή άλλη αιτία, τα κέρδη των 200 μεγαλύτερων εταιρειών της Ελλάδας, την εξέλιξη του πληθωρισμού στην Ελλάδα και στις χώρες της ΕΟΚ, τα εργατικά ατυχήματα, την ανεργγία, το εθνικό εισόδημα μιας χώρας, τις αμοιβές των εργαζομένων, τους σεισμούς, την εξέλιξη της θερμοκρασίας, τις επιδόσεις των αθλητών, την ακροαματικότητα των τηλεοπτικών σταθμών και γενικά οποιοδήποτε θέμα ή φαινόμενο ή δραστηριότητα που εμφανίζει για κάποιον ενδιαφέρον.

6) ΜΕΘΟΔΟΛΟΓΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ ΚΑΙ ΕΡΓΑΛΕΙΑ ΑΝΑΛΥΣΗΣ

Για να είναι χρήσιμες και αξιοποιήσιμες οι πληροφορίες που περιέχονται στα στατιστικά δεδομένα είναι απαραίτητο αυτά να ταξινομούνται και να επεξεργάζονται σύμφωνα με ορισμένους κανόνες και αρχές. Οι κανόνες και οι αρχές αποτελούν μεθοδολογικά εργαλεία που επιτρέπουν την κατά ενιαίο, συστηματικό και αποτελεσματικό τρόπο παρουσίαση και επεξεργασία των πληροφοριών των δεδομένων.

Οι πιο σημαντικές μεθοδολογικές προσεγγίσεις που χρησιμοποιούνται στην περιγραφική στατιστική είναι :

(α) Η παρουσίαση των δεδομένων σε μορφή κατανομών συχνοτήτων. Οι κατανομές αυτές εμφανίζουν κατά τρόπο συνοπτικό τη δομή του πληθυσμού στον οποίο αυτά αναφέρονται και επιτρέπουν την άμεση αξιοποίησή τους. Ανάλογα με τις διαστάσεις του προβλήματος ή τα χαρακτηριστικά του πληθυσμού που εξετάζονται κάθε φορά, οι κατανομές συχνοτήτων διακρίνονται σε **μονομεταβλητές** και σε **διμεταβλητές** ή **πολυμεταβλητές** κατανομές. Στην περίπτωση των μονομεταβλητών κατανομών συχνοτήτων οι μονάδες του πληθυσμού κατατάσσονται ως προς ένα κριτήριο ή ένα χαρακτηριστικό, ενώ στις περιπτώσεις των διμεταβλητών ή πολυμεταβλητών κατανομών οι μονάδες του πληθυσμού κατατάσσονται κατά δύο ή περισσότερα χαρακτηριστικά συγχρόνως.

(β) Ο υπολογισμός ορισμένων χαρακτηριστικών μεγεθών ή μέτρων τα οποία ονομάζονται παράμετροι κατανομών συχνοτήτων και εκφράζουν με απλούς αριθμούς τα βασικά χαρακτηριστικά του πληθυσμού που μελετάται και παρέχουν πληροφορίες για τη μορφολογία της κατανομής του. Εδώ ανήκουν οι παράμετροι κεντρικής τάσης και θέσης, διασποράς και προσδιορισμού της μορφολογίας της κατανομής. Στις περιπτώσεις των διμεταβλητών ή πολυμεταβλητών κατανομών υπολογίζονται παράμετροι που οι τιμές τους

είναι ενδεικτικές του βαθμού συµµεταβολής δυο ή περισσοτέρων µεταβλητών, όπως είναι η συνδιακύµανση και ο συντελεστής συσχέτισης.

(γ) Η κατάρτιση αριθµοδεικτών, οι οποίοι δείχνουν τη διαχρονική ή διαπεριφερειακή ή διακλαδική εξέλιξη διαφορών µεγεθών ή φαινοµένων (π.χ. δείκτες τιµών του καταναλωτή, όγκου, εξωτερικού εµπορίου, τιµών των µετοχών στο χρηµατιστήριο κ.λπ.).

Εξάλλου, δύο από τα σηµαντικότερα αναλυτικά εργαλεία που χρησιµοποιεί ο στατιστικός είναι οι **πίνακες** και τα **διαγράµµατα**. Οι πίνακες εµφανίζουν τα δεδοµένα διατεταγµένα σε γραµµές και στήλες. Οι επικεφαλίδες των γραµµών και των στηλών δείχνουν το χαρακτηριστικό ή τη διάσταση σύµφωνα µε την οποία έχουν καταγραφεί τα δεδοµένα. Τα διαγράµµατα δείχνουν κατά τρόπο παραστατικό (δηλαδή υπό µορφή εικόνας) τις ίδιες πληροφορίες που περιέχονται στα δεδοµένα των πινάκων.

Ωστόσο, µεταξύ πινάκων και διαγραµµάτων υπάρχουν ουσιώδεις διαφορές. Τα διαγράµµατα αναδεικνύουν κατά τρόπο παραστατικό και άµεσο τα κυριότερα χαρακτηριστικά των δεδοµένων. Υπάρχουν πολλά είδη διαγραµµάτων. Κάθε ένα από αυτά εξυπηρετεί διαφορετικούς σκοπούς.

7) ΔΙΑΡΘΡΩΣΗ ΚΑΙ ΠΕΡΙΕΧΟΜΕΝΟ

Σκοπός της ανάλυσης είναι να εισαγάγει τον αναγνώστη σε βασικές έννοιες της Στατιστικής και να τον εξοπλίσει µε ένα *minimum* αναλυτικών εργαλείων, τα οποία θα του επιτρέψουν να προβαίνει σε αναλύσεις στατιστικών δεδοµένων και να εξάγει χρήσιµα συµπεράσµατα για τα φαινόµενα που αυτά περιγράφουν, χωρίς να είναι απαραίτητο να εισέλθει στα βαθιά νερά της Στατιστικής.

Στις κατηγορίες και στις µεθόδους συλλογής των στατιστικών στοιχείων υποδεικνύονται τρόποι οµαδοποιηµένης παρουσίασης των δεδοµένων µε τη βοήθεια συνοπτικών πινάκων και ειδικών διαγραµµάτων. Οι συνοπτικοί πίνακες ονοµάζονται ειδικότερα **κατανοµές συχνοτήτων** και κατέχουν κυρίαρχη θέση όχι µόνο στα πλαίσια της περιγραφικής στατιστικής, αλλά και στη στατιστική θεωρία γενικότερα. Οι κατανοµές συχνοτήτων διακρίνονται σε µονοµεταβλητές και σε πολυµεταβλητές κατανοµές, ανάλογα µε αν το φαινόµενο που περιγράφουν είναι µονοδιάστατο ή πολυδιάστατο. Στην πρώτη περίπτωση τα επιµέρους χαρακτηριστικά του πληθυσµού εξετάζονται κεχωρισµένα, ενώ στη δεύτερη έχουµε συνεξέταση δύο ή περισσοτέρων χαρακτηριστικών του ίδιου πληθυσµού ή φαινοµένου.

Οι τεχνικές που επιτρέπουν τη µεγαλύτερη δυνατή συµπύκνωση των πληροφοριών που περιέχουν τα δεδοµένα αναπτύσσονται, ώστε να είναι εύκολη η αξιοποίηση τους για την κατανόηση της συµπεριφοράς των φαινοµένων που αυτά περιγράφουν, καθώς και στη

λήψη αποφάσεων. Η συμπύκνωση αυτή επιτυγχάνεται με την εκτίμηση ορισμένων δεικτών, οι οποίοι εκφράζουν τις σχετικές πληροφορίες με απλούς αριθμούς. Οι δείκτες αυτοί είναι γνωστοί ως **παράμετροι** κατανομών συχνοτήτων.

Στο χώρο των μονοδιάστατων κατανομών συχνοτήτων, οι πιο σημαντικοί από τους δείκτες αυτούς είναι : οι παράμετροι κεντρικής τάσης και θέσης, οι παράμετροι διασποράς και τα μέτρα που προσδιορίζουν την μορφολογία της κατανομής. Στο χώρο των διμεταβλητών και πολυμεταβλητών κατανομών αναζητούμε παραμέτρους που να μετρούν τον βαθμό συμμεταβολής δύο ή περισσότερων μεταβλητών. Οι πιο σημαντικές παράμετροι στο χώρο αυτό είναι : η συνδιακύμανση, ο συντελεστής συσχέτισης και ο συντελεστής παλινδρόμησης. Άλλο ισχυρό αναλυτικό εργαλείο στη διάθεση του ερευνητή, κυρίως για τη συνεξέταση δύο συσχετιζόμενων μεταβλητών, είναι τα **διαγράμματα διασποράς**.

8) ΠΑΡΑΠΕΡΑ ΣΥΜΠΥΚΝΩΣΗ ΤΩΝ ΠΛΗΡΟΦΟΡΙΩΝ

Η παραπέρα συμπύκνωση (Summarization) των πληροφοριών των δεδομένων ώστε να είναι εύκολη και άμεση η αξιοποίησή τους για την παρακολούθηση της εξέλιξης του φαινομένου και τη λήψη αποφάσεων. Αυτό επιτυγχάνεται με την ανάπτυξη μεθοδολογικών προσεγγίσεων και τεχνικών που επιτρέπουν να εκφραστούν με απλούς δείκτες και συγκεκριμένους αριθμούς τα κυριότερα χαρακτηριστικά του ερευνώμενου πληθυσμού. Στην ορολογία της Στατιστικής οι δείκτες αυτοί ονομάζονται **παράμετροι κατανομών συχνοτήτων** όταν αναφέρονται στον πληθυσμό και **στατιστικά** (statistics) όταν αναφέρονται στο δείγμα. Οι μαθηματικοί τύποι από τους οποίους υπολογίζονται οι σχετικοί δείκτες, είτε αναφέρονται στον πληθυσμό είτε στο δείγμα, είναι οι ίδιοι.

9) Η ΑΝΑΓΚΑΙΟΤΗΤΑ ΤΗΣ ΠΑΡΑΠΕΡΑ ΣΥΜΠΥΚΝΩΣΗΣ ΤΩΝ ΠΛΗΡΟΦΟΡΙΩΝ

Οι κατανομές συχνοτήτων συχνά δεν κρίνονται επαρκείς για την πλήρη περιγραφή και ανάλυση πολυπληθών στατιστικών δεδομένων, για τους παρακάτω λόγους :

(α) Παρά το συνοπτικό τους χαρακτήρα, οι κατανομές συχνοτήτων εξακολουθούν να περιλαμβάνουν μεγάλο αριθμό μεγεθών, γεγονός που δυσχεραίνει τόσο την συγκρατησή τους από τα ενδιαφερόμενα άτομα όσο και την κατανόηση της σημασίας τους για την ερμηνεία της συμπεριφοράς του φαινομένου που ερευνάται.

(β) Στην αρχική τους μορφή οι κατανομές συχνοτήτων δεν διευκολύνουν τις παντοειδείς συγκρίσεις (διαχρονικές, διαπεριφερειακές, διαστρωματικές, διακλαδικές κ.λπ.) που συχνά αποτελούν ένεαν από τους κυρίους σκοπούς της

στατιστικής ανάλυσης των φαινομένων της ζωής (φυσικών, οικονομικών, κοινωνικών κ.λπ.).

(γ) Δυσχεραίνουν τη διενέργεια προβλέψεων και τη λήψη αποφάσεων, γεγονός που περιορίζει τη χρησιμότητα της στατιστικής ανάλυσης.

(δ) Σε πολλές περιπτώσεις δεν είναι απαραίτητη η παρουσίαση των πληροφοριών στην έκταση που δίνουν οι κατανομές συχνοτήτων. Στις περιπτώσεις αυτές η λεπτομερής παρουσίαση των δεδομένων αποτελεί περιττή και δαπανηρή πολυτέλεια.

Για τους παραπάνω λόγους, στην στατιστική έχουν αναπτυχθεί διαδικασίες και τεχνικές που συνοψίζουν τις πληροφορίες ενός συνόλου δεδομένων σε απλά αριθμητικά μεγέθη, τα οποία, όπως ήδη αναφέρθηκε, ονομάζονται παράμετροι του πληθυσμού ή της αντίστοιχης κατανομής. Τα μεγέθη αυτά είναι γνωστά και ως **περιγραφικά μέτρα (descriptive measures)** της κατανομής.

10) ΕΙΔΗ ΠΑΡΑΜΕΤΡΩΝ ΜΟΝΟΜΕΤΑΒΛΗΤΩΝ ΚΑΤΑΝΟΜΩΝ

Σε σχέση με τις μονομεταβλητές κατανομές το ενδιαφέρον επικεντρώνεται στον προσδιορισμό : α) του «κέντρου» της κατανομής, β) της διασποράς των τιμών της μεταβλητής γύρω από το «κέντρο», γ) στη σύγκριση της θέσης μιας τιμής με εκείνη των υπολοίπων τιμών και δ) στον προσδιορισμό της μορφολογίας της κατανομής. Έτσι διακρίνουμε τις ακόλουθες κατηγορίες (είδη) παραμέτρων :

Παράμετροι κεντρικής τάσης. Αυτές εκφράζουν το σημείο γύρω από το οποίο τείνουν να συγκεντρωθούν οι επιμέρους τιμές του πληθυσμού (της μεταβλητής). Οι βασικότερες παράμετροι της κατηγορίας αυτής είναι : ο μέσος, η διάμεσος και το σημείο μέγιστης συχνότητας ή τύπος.

Παράμετροι διασποράς. Αυτές μετράνε τη διασπορά των επιμέρους τιμών γύρω από το «κέντρο» της κατανομής. Οι πιο σημαντικές παράμετροι διασποράς είναι : το εύρος, η απόλυτη μέση απόκλιση, η διακύμανση και η τυπική απόκλιση.

Παράμετροι θέσης. Πρόκειται για μεγέθη που δείχνουν το βαθμό που μια τιμή (ή μια ομάδα τιμών) συμβαδίζει (ταιριάζει) με τις υπόλοιπες τιμές ή ομάδες. Εδώ ανήκουν τα τεταρτημόρια, τα δεκατημόρια και τα εκατοστημόρια.

Παράμετροι προσδιορισμού της μορφολογίας της κατανομής. Πρόκειται για δείκτες μέτρησης του βαθμού ασυμμετρίας και κύρτωσης μιας κατανομής.

Σημειώνεται ότι ο υπολογισμός των ανωτέρων παραμέτρων αφορά στις ποσοτικές μεταβλητές (συνεχείς και ασυνεχείς). Στις ποιοτικές και κατηγορικές μεταβλητές κάτι τέτοιο είναι αδύνατο. Εξαιρέση αποτελούν η διάμεσος, η οποία υπολογίζεται και στις περιπτώσεις των ποιοτικών

μεταβλητών και το σημείο μέγιστης συχνότητας που μπορεί να υπολογιστεί και για τις τρεις κατηγορίες μεταβλητών (ποσοτικές, ποιοτικές, κατηγορικές).

11) ΤΡΟΠΟΙ ΕΜΦΑΝΙΣΗΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Τα δεδομένα που χρησιμοποιούνται για την εκτίμηση των παραμέτρων των κατανομών συχνοτήτων μπορεί να εμφανίζονται είτε σε απλή είτε σε ομαδοποιημένη μορφή. Απλή μορφή έχουμε στις περιπτώσεις που τα δεδομένα είναι αταξινόμητα. Ο συμβολισμός που χρησιμοποιείται για την αναπαράσταση των τιμών των δεδομένων είναι : X_1, X_2, \dots, X_N , όπου N = το πλήθος των μονάδων του πληθυσμού ή του δείγματος. Αντίθετα, όταν οι τιμές της υπο εξέταση μεταβλητής είναι ταξινομημένες σε ομάδες, τα δεδομένα εμφανίζονται σε ομαδοποιημένη μορφή . Μια άλλη περίπτωση που εμφανίζεται στην πράξη είναι η εξής : Οι τιμές της μεταβλητής (έστω X) είναι μεμονωμένες (δηλαδή αταξινόμητες), πλην όμως στις επιμέρους τιμές δίνεται διαφορετική βαρύτητα.

Η μορφή υπό την οποία εμφανίζονται τα δεδομένα επηρεάζει την επιλογή του κατάλληλου μαθηματικού τύπου από τον οποίο θα υπολογίζονται οι διάφοροι παράμετροι.

12) ΠΑΡΑΜΕΤΡΟΙ ΔΙΜΕΤΑΒΛΗΤΩΝ ΚΑΤΑΝΟΜΩΝ

Στις περιπτώσεις που αντικείμενο έρευνας είναι η μελέτη διμεταβλητών κατανομών το πρόβλημα ανάγεται στην αναζήτηση δεικτών (παραμέτρων) που να εκφράζουν κατά το πλέον συμπυκνωμένο τρόπο τις πληροφορίες που αφορούν ταυτόχρονα τις δύο μεταβλητές. Η πιο βασική παράμετρος που μετράει τη συγκατανομή δυο μεταβλητών είναι η συνδιακύμανση. Άλλες παράμετροι που μετρούν το βαθμό και τον τρόπο κατά τον οποίο μεταβάλλονται δυο συνεξεταζόμενες μεταβλητές είναι ο **συντελεστής συσχέτισης** και ο **συντελεστής παλινδρόμησης**. Στη βάση του ορισμού και των δύο αυτών παραμέτρων βρίσκεται η έννοια της συνδιακύμανσης .

13) ΠΑΡΑΜΕΤΡΟΙ ΚΕΝΤΡΙΚΗΣ ΘΕΣΗΣ ΚΑΙ ΤΑΣΗΣ

Οι παράμετροι κεντρικής τάσης είναι αριθμητικοί δείκτες (μέτρα) της τυπικής ή της μέσης τιμής μιας κατανομής, ενδεικτικής της συγκέντρωσης των επιμέρους τιμών της μεταβλητής. Αυτές εκφράζουν τις πλέον αντιπροσωπευτικές τιμές της κατανομής με κριτήριο το βαθμό συγκέντρωσης και χρησιμοποιούνται τόσο για την περιγραφή των χαρακτηριστικών των κατανομών όσο και για τη σύγκριση δύο ή περισσότερων κατανομών. Όπως αναφέρθηκε πιο πάνω, οι σημαντικότερες παράμετροι κεντρικής τάσης είναι ο **μέσος**, η **διάμεσος** και το **σημείο μέγιστης συχνότητας**, ενώ ως μέτρα θέσης αναφέρονται τα **τεταρτημόρια**, τα **δεκατημόρια** και τα **εκατοστημόρια**.

14) ΠΑΡΑΜΕΤΡΟΙ ΔΙΑΣΠΟΡΑΣ

Τα μέτρα (παράμετροι) βοηθούν στον ποσοτικό προσδιορισμό της κεντρικής τάσης και της θέσης των τιμών μιας κατανομής. Όμως, δυο κατανομές ενδέχεται να έχουν την ίδια κεντρική τάση τιμών και παρά ταύτα να διαφέρουν αισθητά μεταξύ τους. Έτσι, ένα σύνολο χαρακτηριστικών που διαφοροποιούν τις κατανομές είναι οι παράμετροι διασποράς, οι οποίες μετρούν το βαθμό που οι τιμές της κατανομής αποκλίνουν από την τιμή της κεντρικής της τάσης. Η διασπορά των τιμών μιας μεταβλητής γύρω από τον μέσο της κατανομής μπορεί να μετρηθεί είτε σε όρους της απόστασης μεταξύ ζευγών διατεταγμένων τιμών αυτής, είτε σε όρους της μέσης απόστασης των επιμέρους παρατηρήσεων από την κεντρική τιμή. Στην πρώτη κατηγορία των μέτρων ανήκουν το εύρος και το ημιενδοτεταρτημοριακό εύρος. Στη δεύτερη κατηγορία εντάσσονται η μέση απόκλιση, η διακύμανση, η τυπική απόκλιση και ο συντελεστής μεταβλητότητας.

Η επαγωγική στατιστική (statistical induction) αποτελεί τον κορμό της στατιστικής θεωρίας. Αντικείμενο της στατιστικής επαγωγής είναι η ανάπτυξη μεθόδων και τεχνικών, οι οποίες βοηθούν τον ερευνητή στη διατύπωση γενικής ισχύος προτάσεων (νόμων) σε καθεστώς περιορισμένης πληροφόρησης. Η σχετική προσπάθεια διευκολύνεται τα μέγιστα με τη χρήση υποδειγμάτων πιθανότητας. Ο όρος «περιορισμένη πληροφόρηση» εκφράζει το γεγονός ότι ο ερευνητής έχει στη διάθεση του μικρό μόνο τμήμα (δείγμα) του συνόλου των δεδομένων, που προσδιορίζουν τη συμπεριφορά του φαινομένου. Με άλλα λόγια, στην επαγωγική στατιστική η αναλυτική πορεία που διατρέχει ο νους έχει ως αφετηρία το μέρος (το συγκεκριμένο, το δείγμα) και καταλήγει στη διατύπωση νόμων ή γενικών κανόνων αναφορικά με την εξέλιξη του υπό εξέταση φαινομένου. Στη διαδρομή αυτή ως εισροές χρησιμοποιούνται τα δεδομένα που έχουν προετοιμαστεί σύμφωνα με τις αρχές της περιγραφικής στατιστικής, καθώς και τα υποδείγματα πιθανότητας.

Αξίζει να τονιστεί ότι η στατιστική επαγωγή αποτελεί τη σημαντικότερη μέθοδο παραγωγής νέας γνώσης σε συνθήκες σχετικής αβεβαιότητας. Ως παράδειγμα αναφέρουμε τη διατύπωση του νόμου του Engel στα οικονομικά σύμφωνα με το οποίο, αυξανόμενου του εισοδήματος μιας κοινωνίας το ποσοστό αυτού που δαπανάται για είδη διατροφής βαίνει φθίνον. Η διατύπωση του νόμου αυτού έγινε πριν από έναν και πλέον αιώνα από τον Γερμανό στατιστικό E. Engel μετά από επεξεργασία δεδομένων της Γερμανικής οικονομίας εκείνης της εποχής. Και μέχρι σήμερα γίνεται αποδεκτό ότι ισχύει για κάθε οικονομία. Βεβαίως, τα παραδείγματα ανακάλυψης νέων γνώσεων στους διάφορους τομείς της επιστήμης (ιατρική, βιολογία, οικονομικά, ψυχολογία, κ.λ.π.) είναι «άπειρα».

Ωστόσο, η γνώση που δημιουργείται με βάση τη στατιστική επαγωγή είναι ατελής με την έννοια ότι δεν γίνεται ανεπιφύλακτα δεκτή, ούτε έχει καθολική και μη ανατρέψιμη ισχύ. Γι' αυτό τα «προϊόντα» της στατιστι-

κής επαγωγής (νόμοι, κανόνες, σχέσεις) δεν εκφράζονται κατά τρόπο απόλυτο, αλλά με την μορφή προτάσεων πιθανότητας. Αυτό σημαίνει ότι τα διάφορα συμπεράσματα έχουν υψηλή πιθανότητα να είναι σωστά. Ουδέποτε όμως μπορούμε να πούμε με βεβαιότητα ότι αυτά είναι όντως σωστά. Έτσι, η νέα γνώση που «παράγεται» με το μηχανισμό τη στατιστικής επαγωγής υπόκειται σε διαρκή αναθεώρηση, καθώς εισρέουν νέες πληροφορίες (δεδομένα) για το φαινόμενο που μελετάται και είναι δυνατό να ανατραπεί μόλις εμφανιστούν πολλές (όχι μία) περιπτώσεις που δεν συμφωνούν με τον κανόνα. Είναι προφανές ότι η στατιστική επαγωγή διαφέρει ριζικά από τη μαθηματική επαγωγή, η οποία έχει τη δύναμη της τέλει λογικής επαγωγής.

Παρά τις αδυναμίες που τη χαρακτηρίζουν, η στατιστική επαγωγή αποτελεί εξαιρετικά χρήσιμο νοητικό εργαλείο παραγωγής νέων γνώσεων, αφού αποτελεί το μηχανισμό που μετασχηματίζει την περιορισμένη πληροφόρηση (μικρό δείγμα) σε γνώσεις για το σύνολο οι οποίες δεν είναι απόλυτα βέβαιες, είναι όμως σε υψηλό βαθμό πιθανές. Λόγω της αδυναμίας αυτής, η γνώση που δημιουργείται με τη μεθοδολογία της στατιστικής επαγωγής πρέπει να χρησιμοποιείται με μεγάλη προσοχή. Ειδικότερα, οι χρήστες της νέας γνώσης πρέπει να έχουν συνείδηση του γεγονότος ότι : (α) αυτό ισχύει με υψηλή πιθανότητα (π.χ. 95%) όχι όμως με βεβαιότητα, πράγμα που σημαίνει ότι είναι δυνατό να υπάρξουν περιπτώσεις, οι οποίες δεν υπακούουν στον κανόνα, (β) η νέα γνώση πρέπει να συμπληρώνεται και να διορθώνεται ή ακόμη και να εγκαταλείπεται, όταν κάποιος βεβαιώνεται ότι αυτή είναι ανεπαρκής.

Η προσφυγή στη μέθοδο της στατιστικής επαγωγής είναι αναγκαία σε πολλές περιπτώσεις που έχουμε πλήρη άγνοια ή περιορισμένη γνώση των νόμων που διέπουν τα διάφορα φυσικά, οικονομικά, κοινωνικά και λοιπά φαινόμενα. Από την άλλη μεριά, ακόμη και στις περιπτώσεις που είναι θεωρητικά δυνατή η απόκτηση όλων των πληροφοριών (δεδομένων), που προσδιορίζουν τη συμπεριφορά των φαινομένων, το κόστος σε χρόνο και χρήμα είναι συνήθως απαγορευτικό. Έτσι, περιοριζόμαστε αναγκαστικά σε μικρό μόνο τμήμα του συνόλου των πληροφοριών και προσπαθούμε με βάση αυτές και τη βοήθεια της θεωρίας πιθανοτήτων να θεμελιώσουμε γενικούς κανόνες ή νόμους, οι οποίοι να ερμηνεύουν τα φαινόμενα που ενδιαφέρουν τον άνθρωπο.

15)ΘΕΩΡΙΑ ΚΑΙ ΚΑΤΑΝΟΜΕΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Η βασικότερη μέθοδος συγκέντρωσης στατιστικών δεδομένων για την ποσοτική διερεύνηση των φαινομένων είναι η μέθοδος της δειγματοληψίας. Η προσέγγιση αυτή συνιστάται στην επιλογή από τον ερευνώμενο πληθυσμό ενός δείγματος μικρού σχετικά μεγέθους, το οποίο στη συνέχεια χρησιμοποιείται ως βάση για την εξαγωγή συμπερασμάτων αναφορικά με τα χαρα-

κτηριστικά του πληθυσμού. Με τον όρο δείγμα, νοείται ένα υποσύνολο του συνόλου των μονάδων του πληθυσμού, που επιλέγεται σύμφωνα με κάποιο κανόνα ή με ορισμένο σχέδιο.

Αξίζει να τονιστεί ότι η ραγδαία ανάπτυξη της στατιστικής θεωρίας οφείλεται κατά κύριο λόγο στην πιεστική ανάγκη εξαγωγής συμπερασμάτων προσδιορίσιμης αξιοπιστίας για τον πληθυσμό, με βάση τις πληροφορίες που περιέχονται στο δείγμα. Πράγματι, από τη σκοπιά της στατιστικής θεωρίας το θέμα μελέτης των χαρακτηριστικών του πληθυσμού, όταν διαθέτουμε ολόκληρο τον πληθυσμό, είναι απλό και στη φύση του είναι κυρίως περιγραφικό (descriptive). Προβλήματα θεωρητικά και μάλιστα δύσκολα δημιουργούνται όταν πρόκειται να διατυπωθούν συμπεράσματα για τον πληθυσμό που στηρίζονται σε δεδομένα ενός μικρού σχετικά υποσυνόλου αυτού. Έτσι, αναπτύχθηκε ο κλάδος της στατιστικής συμπερασματολογίας που αποτελεί τη «ραχοκοκαλιά» της στατιστικής επιστήμης.

Η προσφυγή στη λύση του δείγματος, αντί της χρησιμοποίησης του ίδιου του πληθυσμού, υπαγορεύεται τόσο από λόγους αντικειμενικής αδυναμίας, όσο και από τον υπερβολικό χρόνο και το απαγορευτικό κόστος που συνεπάγεται συνήθως η επεξεργασία του συνόλου των στοιχείων του πληθυσμού στις περιπτώσεις που αυτά είναι διαθέσιμα. Συχνά, ο πληθυσμός που ερευνάται είναι αδύνατο να προσδιοριστεί αριθμητικά,

16) ΕΦΑΡΜΟΓΗ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ ΣΤΙΣ ΕΠΙΧΕΙΡΗΣΕΙΣ

Ακολουθούν μερικά παραδείγματα εφαρμογής της στατιστικής στις επιχειρήσεις

Μια εταιρεία που ετοιμάζεται να "ρίξει" στην αγορά ένα νέο προϊόν χρειάζεται να εκτιμήσει τις προτιμήσεις των καταναλωτών που ανήκουν στην σχετική αγορά. Αυτό μπορεί να το πετύχει συνήθως με μια ερευνά μάρκετινγκ η οποία θα βασίζεται σε συνεντεύξεις κάποιων τυχαία επιλεγμένων νοικοκυριών. Έτσι, τα αποτελέσματα της ερευνάς θα μπορούσαν να χρησιμοποιηθούν στη συνέχεια για να εκτιμηθούν οι προτιμήσεις ολόκληρου του πληθυσμού.

Στατιστικές τεχνικές είναι απαραίτητες για το "ξεμπέρδεμα" των επιδράσεων πολλών διαφορετικών παραγόντων. Για παράδειγμα, η ζήτηση για παγωτό σε μια κοινότητα ενδέχεται να εξαρτάται από την τιμή του παγωτού, το επίπεδο του μέσου εισοδήματος, τον αριθμό των παιδιών αυτής της κοινότητας, και τη μέση θερμοκρασία. Αν έχουμε στην διάθεση μας παρατηρήσεις όλων των διαφορετικών παραγόντων που εμπλέκονται σε αυτή την περίπτωση, μπορούμε να χρησιμοποιήσουμε ανάλυση πα-

λινδρομήσεις για να βρούμε ποιοι παράγοντες επιδρούν περισσότερο στην ζήτηση παγωτού .

Ένας ορκωτός λογιστής έχει Αποστολή να ελέγξει τα βιβλία μιας εταιρείας για να βεβαιωθεί ότι ανταποκρίνονται απόλυτα στην οικονομική κατάσταση της συγκεκριμένης εταιρείας . Ο ορκωτός λογιστής θα πρέπει να ελέγξει ολόκληρους σωρούς πρωτότυπων παραστατικών όπως Δελτία Αποστολής , έντυπα παραγγελιών , και τιμολόγια . Είναι προφανές ότι για να ελέγξει ένα προς ένα όλα τα παραστατικά θα χρειαζόταν παρά πολλές ημέρες , αν όχι εβδομάδες . Αντί γι' αυτό , μπορεί να ελέγξει ένα δείγμα των παραστατικών , το οποίο θα επιλέξει με τυχαίο τρόπο , και να καταλήξει σε ένα συμπέρασμα για ολόκληρο τον πληθυσμό των παραστατικών με βάση αυτό το δείγμα .

Πριν κυκλοφορήσει στην αγορά ένα νέο φάρμακο είναι απαραίτητο να γίνουν εκτεταμένα πειράματα για να εξασφαλιστεί ότι το φάρμακο είναι ασφαλές και αποτελεσματικό . Ο καλύτερος τρόπος για να ελέγξει κανείς ένα φάρμακο είναι να πάρει δυο ομάδες ανθρώπων που μοιάζουν μεταξύ τους όσο το δυνατό περισσότερο , να χορηγήσει το φάρμακο μόνο στη μια από τις δυο ομάδες , και μετά να ελέγξει αν τα αποτελέσματα των δυο ομάδων διαφέρουν . Η ομάδα στην οποία χορηγείται το φάρμακο ονομάζεται πειραματική ομάδα και η άλλη ομάδα ονομάζεται ομάδα έλεγχου . Για να εξακριβωθεί αν οι οποίες διαφορές που παρατηρήθηκαν οφείλονται πράγματι στη χορήγηση του φαρμάκου ή μπορεί να προκλήθηκαν από άλλους παράγοντες απαιτείται στατιστική ανάλυση .

Αν ένας έμπορος πρόκειται να παραλάβει μια μεγάλη ποσότητα Αγάθων από κάποιων προμηθευτή του , θα θελήσει οπωσδήποτε να βεβαιωθεί ότι έχουν ικανοποιηθεί όλες οι προδιαγραφές που έχει συμφωνήσει με τον προμηθευτή . Δε χωρά αμφιβολία ότι το να κάνει ποιοτικό έλεγχο στα αγαθά που παρέλαβε ένα προς ένα θα σήμαινε υπέρογκες δαπάνες . Και σε αυτή την περίπτωση οι στατιστικές τεχνικές έρχονται να σώσουν την κατάσταση , επιτρέποντας στον έμπορο να εξαγάγει συμπεράσματα για την ποιότητα ολόκληρης της παρτίδας ελέγχοντας μόνο ένα τυχαία επιλεγμένο δείγμα Αγάθων που θα πάρει από την παρτίδα .

ΜΕΡΟΣ ΔΕΥΤΕΡΟΝ

1) ΕΚΤΙΜΗΤΙΚΗ (Η ΕΥΡΕΣΗ ΕΚΤΙΜΗΤΡΙΩΝ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ)

ΙΔΙΟΤΗΤΕΣ ΤΩΝ ΕΚΤΙΜΗΤΡΙΩΝ

αμερόληπτη εκτιμήτρια (unbiased estimator): μια εκτιμήτρια της οποίας η αναμενόμενη τιμή ισούται με την πραγματική τιμή της παραμέτρου της οποίας αποτελεί εκτίμηση

εκτιμήτρια (estimator): μια ποσότητα που βασίζεται σε παρατηρήσεις ενός δείγματος και της οποίας η τιμή λαμβάνεται ως ένδειξη της τιμής μιας άγνωστης παραμέτρου του πληθυσμού (για παράδειγμα , ο μέσος του δείγματος , \bar{x} , χρησιμοποιείται συχνά ως εκτιμήτρια του άγνωστου μέσου του πληθυσμού , μ)

εκτιμήτρια μέγιστης πιθανότητας (maximum likelihood estimator): μια εκτιμήτρια με την εξής ιδιότητα : αν η πραγματική τιμή της άγνωστης παραμέτρου έχει αυτή την τιμή , τότε η πιθανότητα να πάρουμε το δείγμα που πράγματι παρατηρήθηκε μεγιστοποιείται

επαγωγική στατιστική (statistical inference) : η διαδικασία κατά την οποία χρησιμοποιούνται οι παρατηρήσεις ενός δείγματος για την εκτίμηση των ιδιοτήτων του πληθυσμού

συνεπής εκτιμήτρια (consistent estimator) : μια εκτιμήτρια που , όσο μεγαλώνει το μέγεθος του δείγματος , τείνει να συγκλίνει προς την πραγματική τιμή

διάστημα εμπιστοσύνης (confidence interval) : ένα διάστημα που βασίζεται σε παρατηρήσεις ενός δείγματος και είναι καθορισμένο με τέτοιο τρόπο ώστε να υπάρχει μια συγκεκριμένη πιθανότητα ότι θα περιέχει την άγνωστη πραγματική τιμή μιας παραμέτρου (για παράδειγμα ,

συνηθίζεται ο καθορισμός διαστημάτων εμπιστοσύνης που έχουν 95% πιθανότητα να περιέχουν την πραγματική τιμή)

επίπεδο εμπιστοσύνης (confidence level) : ο βαθμός εμπιστοσύνης που σχετίζεται με ένα διάστημα εμπιστοσύνης , η πιθανότητα ότι το διάστημα περιέχει την πραγματική τιμή της παραμέτρου

2) ΔΗΜΟΣΚΟΠΗΣΕΙΣ (ΜΕΘΟΔΟΙ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ)

δειγματοληψία κατά συστοιχίες (cluster sampling) : μια μέθοδος δειγματοληψίας κατά την οποία ο πληθυσμός διαιρείται σε συστοιχίες , επιλέγονται μερικές συστοιχίες τυχαία , και στη συνέχεια μερικά μέλη των επιλεγμένων συστοιχιών για να αποτελέσουν το δείγμα

στρωματοποιημένη δειγματοληψία (stratified sampling): μια μέθοδος δειγματοληψίας κατά την οποία ο πληθυσμός διαιρείται σε στρώσεις που μοιάζουν όσο το δυνατό περισσότερο μεταξύ τους

Πως μπορούμε να διαπιστώσουμε πόσοι άνθρωποι από κάποιων πληθυσμό έχουν ένα συγκεκριμένο χαρακτηριστικό ; Για παράδειγμα , μπορεί να θέλουμε να μάθουμε πόσοι ψηφοφόροι υποστηρίζουν τον υποψήφιο δήμαρχο που προτιμάμε , ή μπορεί να μας ενδιαφέρουν ορισμένα γενικά χαρακτηριστικά των κατοίκων ενός συγκεκριμένου νόμου - όπως , για παράδειγμα , το πλήθος των παιδιών ανά οικογένεια , ο πληθυσμός των πόλεων , ή το πλήθος των εργαζόμενων .

Ένας τρόπος για να απαντηθούν αυτές οι ερωτήσεις είναι να τους ρωτήσουμε όλους . Η μέθοδος αυτή θα ήταν , προφανώς , πολύ ακριβής . Επειδή θα ρωτήσουμε τους πάντες , θα μας δοθεί η δυνατότητα να πάρουμε μια λεπτομερή εικόνα ολόκληρου του πληθυσμού . Υπάρχουν φορές που χρησιμοποιείται αυτή η μέθοδος . Κάθε τέσσερα χρόνια προκηρύσσονται στη χώρα μας εκλογές για την εύρεση των κομματικών προτιμήσεων όλων των ψηφοφόρων , και κάθε δέκα χρόνια διεξάγεται απογραφή για την συγκέντρωση πληροφοριών για όλους τους κάτοικους της χώρας . Ωστόσο υπάρχουν μειονεκτήματα σε αυτή τη μέθοδο κατά την οποία ερωτώνται όλοι . Το βασικό μειονέκτημα είναι ότι κοστίζει πολύ . Οι εκλογές και οι απογραφές στοιχίζουν . Επίσης , σε πολλές περιπτώσεις θέλουμε να πάρουμε πληροφορίες για τον πληθυσμό

αλλά δεν μπορούμε να περιμένουμε μέχρι την επόμενη απογραφή ή τις επόμενες εκλογές .

Μια άλλη μέθοδος που μπορούμε να εφαρμόσουμε είναι να ρωτήσουμε μερικά μέλη ενός δείγματος . Αν το δείγμα είναι αντιπροσωπευτικό του πληθυσμού μπορούμε να χρησιμοποιήσουμε τα χαρακτηριστικά των ανθρώπων του δείγματος για να εκτιμήσουμε τα χαρακτηριστικά των ανθρώπων ολόκληρου του πληθυσμού . Για παράδειγμα , στις σφυγμομετρήσεις που γίνονται στη χώρα μας , συνήθως γίνεται καταγραφή των απόψεων περίπου 1,000 ανθρώπων σε μια προσπάθεια να εκτιμηθούν οι γνώμες 10 εκατομμυρίων κατοίκων .

Τι πιθανότητα υπάρχει να είναι ακριβή τα αποτελέσματα ; Με την πρώτη μάτια , ίσως να δυσπιστήσετε . Κατά τις σφυγμομετρήσεις λαμβάνεται η γνώμη μόνο ενός ανθρώπου για κάθε 10,000 ανθρώπους . Ίσως να θεωρείται ότι είναι απίθανο κάθε πρόσωπο της σφυγμομέτρησης να έχει τις ίδιες απόψεις με τους 10,000 γείτονες του . Όμως , οι σφυγμομετρήσεις φαίνεται ότι είναι αρκετά ακριβείς . Οι προβλέψεις των σφυγμομετρήσεων για εκλογικά αποτελέσματα συνήθως "πέφτουν " κοντά στα πραγματικά αποτελέσματα (με μερικές αξιοσημείωτες εξαιρέσεις) .

Θα ασχοληθούμε τώρα με την θεωρία που εξηγεί γιατί αυτά τα αποτελέσματα τείνουν να είναι ακριβή . Θα χρησιμοποιήσουμε το γράμμα N για να παραστήσουμε τον αριθμό των ανθρώπων του πληθυσμού που ερευνούμε , και θα υποθέσουμε ότι M από αυτούς τους ανθρώπους υποστηρίζουν τον υποψήφιο δήμαρχο της προτιμήσεως μας . Σκοπός μας είναι να εκτιμήσουμε το M/N -το ποσοστό των ανθρώπων που υποστηρίζουν τον υποψήφιο μας . Αν η σφυγμομέτρηση μας είναι καλή , το X/n θα βρίσκεται κοντά στο M/N . Έστω λοιπόν , $p=X/n$.

Για παράδειγμα , ας υποθέσουμε ότι προσπαθούμε να εκτιμήσουμε τις προτιμήσεις των κατοίκων μιας πόλης με πληθυσμό 30,000 ανθρώπους . Θα ρωτήσουμε και τους 500 ανθρώπους ενός δείγματος ποιον υποψήφιο υποστηρίζουν . Ας υποθέσουμε ότι στην πραγματικότητα υπάρχουν 16,500 υποστηρικτές του υποψήφιου μας (οι 16,500 αντιστοιχούν στο 55% του πληθυσμού) . Αν αποδειχθεί ότι 270 άνθρωποι (54%) του δείγματος υποστηρίζουν τον υποψήφιο μας , τότε το δείγμα μας αντιπροσώπευσε ικανοποιητικά ολόκληρο τον πληθυσμό . Από την άλλη μεριά , αν ο υποψήφιος μας υποστηριχτεί από 330 ανθρώπους (66%) του δείγματος , τότε το δείγμα δεν είναι καθόλου αντιπροσωπευτικό και η σφυγμομέτρηση μας θα μας δόση πολύ παραπλανητικά αποτελέσματα

Ο τρόπος επιλογής του δείγματος έχει μεγάλη σημασία , καθώς η τιμή του X θα εξαρτάται από το ποιοι ακριβώς απαρτίζουν το δείγμα . Θα πρέπει να επινοήσουμε κάποιο έξυπνο σύστημα για την επιλογή του

δείγματος , έτσι ώστε αυτό να αντιπροσωπεύει ολόκληρο τον πληθυσμό . Δεν μπορούμε να αρχίσουμε να ρωτάμε τους φίλους μας , γιατί το πιθανότερο είναι να έχουν τις ίδιες απόψεις με εμάς . Ούτε μπορούμε να επιλέξουμε μια μόνο γειτονία για να πάρουμε τη γνώμη όλων των κατοίκων της , καθώς οι κάτοικοι μιας συγκεκριμένης γειτονίας είναι απίθανο να αντιπροσωπεύουν τα ποικίλα χαρακτηριστικά όλων των ανθρώπων της πόλης .

Πάντως , και με αλλά συστήματα υπάρχουν περίπλοκα προβλήματα . Δεν μπορούμε απλώς να ταχυδρομήσουμε κάρτες και να ζητάμε από τους παραλήπτες να τις επιστρέψουν συμπληρωμένες , μιας και αυτοί που συμφωνούν με την άποψη είναι πολύ πιο πιθανό να μπουν στον κόπο να επιστρέψουν τις κάρτες . Ίσως να αποφασίσουμε ότι μπορούμε να κάνουμε το δείγμα μας αντιπροσωπευτικό αποφασίζοντας προκαταβολικά ότι θέλουμε να περιέχει κάποια συγκεκριμένα ποσοστά ανθρώπων με συγκεκριμένα χαρακτηριστικά . Για παράδειγμα , ενδέχεται να καταλήξουμε ότι θέλουμε το δείγμα μας να περιέχει 50% γυναίκες 15% άτομα από μειονότητες , και 0,5% κτηνίατρους . Όμως , ούτε αυτή η μέθοδος άπαντα στην ερώτηση για το πως θα επιλέξουμε το δείγμα , καθώς δεν μας λέει ποιες μειονότητες , ποιες γυναίκες , ή ποιους κτηνίατρους πρέπει να συμπεριλάβουμε στο δείγμα . Είναι προφανές ότι δεν μπορούμε να καθορίσουμε προκαταβολικά ένα ποσοστό για τους ανθρώπους του δείγματος που υποστηρίζουν τον υποψήφιο μας , καθώς δεν γνωρίζουμε αυτό το ποσοστό πριν εξετάσουμε το δείγμα . Αποδεικνύεται ότι το καλύτερο σύστημα επιλογής του δείγματος είναι να μην έχουμε κανένα σύστημα -με αλλά λόγια , είναι καλύτερα να επιλέξουμε το δείγμα εντελώς τυχαία . Θα πρέπει να σχεδιάσουμε το σύστημα δειγματοληψίας έτσι ώστε κάθε άτομο να έχει την ίδια ευκαιρία να επιλέγει . Και όχι μόνο αυτό , αλλά θα πρέπει να σχεδιάσουμε το σύστημα έτσι ώστε κάθε δυνατό δείγμα που θα μπορούσαμε να φανταστούμε να έχει την ίδια ευκαιρία να αποτελέσει το δείγμα που πραγματικά θα επιλέξουμε .

Πως θα το κάνουμε αυτό ; Ένας τρόπος είναι να γράψουμε το όνομα του καθενός σε ένα χαρτάκι , να βάλουμε τα χαρτάκια μέσα σε μικρά σφαιρίδια σε μια μεγάλη κληρωτίδα . Αν τα ανακατέψουμε πολύ καλά και μετά τραβήξουμε η σφαιρίδια από την κληρωτίδα , θα έχουμε ένα τυχαίο δείγμα από η ανθρώπους . Όμως , ακόμη και αυτή η προσέγγιση είναι δύσκολη αν το N είναι πολύ μεγάλο . Καταρχήν , θα χρειαζόμασταν πόλη μεγάλη κληρωτίδα . Ένα άλλο πρόβλημα είναι ότι , αν τα σφαιρίδια είναι πολλά , είναι δύσκολο να τα ανακατέψουμε καλά . Αν τα σφαιρίδια δεν ανακατεντούν καλά , τότε οι άνθρωποι των οποίων τα ονόματα τοποθετήθηκαν τελευταία έχουν μεγαλύτερη πιθανότητα να επιλέγουν , με αποτέλεσμα το δείγμα μας να μην είναι εντελώς τυχαίο .

Κάτι παρόμοιο συνέβη στην πραγματικότητα με την επιλογή κληρωτών (στρατιωτών) στις Η.Π.Α. το 1970. Αριθμοί (κλήροι) ανατέθηκαν

στους ανθρώπους με βάση την ημερομηνία γέννησης τους , όπου ένας μικρός αριθμός δήλωνε ότι η πιθανότητα στρατολόγησης ήταν μεγαλύτερη . Οι 366 ημερομηνίες τοποθετήθηκαν σε σφαιρίδια , ανακατεύτηκαν έγινε κλήρωση , και στις ημερομηνίες αντιστοιχήθηκαν λαχνοί με αριθμούς 1 , 2, κ.λ.π. Προφανώς , τα σφαιρίδια δεν ανακατεύτηκαν καλά , γιατί ο μέσος όρος των λάχνων των ατόμων που είχαν γεννηθεί το Δεκέμβριο ήταν 124,5 , ένας αριθμός που βρίσκεται πολύ μακριά από το μέσο όρο των αριθμών 1 έως 366, ο οποίος είναι ίσος με 183. Στην επιλογή κληρωτών του 1971 λήφθηκαν μέτρα ώστε τα αποτελέσματα να είναι περισσότερο τυχαία : έγινε κλήρωση τόσο των ημερομηνιών όσο και των αριθμών των λάχνων από κληρωτίδες , μετά από πολύ καλύτερο ανακάτεμα .

Μια ευκολότερη μέθοδος είναι να εκχωρήσουμε κάποιον αριθμό σε όλα τα μέλη του πληθυσμού , να επιλέξουμε στη συνέχεια ένα σύνολο από τυχαίους αριθμούς , και να πάρουμε τη γνώμη των ανθρώπων των οποίων τους αριθμούς έχουμε επιλέξει . Πως μπορούμε να επιλέξουμε, όμως , τους τυχαίους αριθμούς ; Δεν είναι τόσο εύκολο όσο φαίνεται . Δεν μπορούμε απλώς να αρχίσουμε να σκεφτόμαστε αριθμούς , καθώς είναι δύσκολο για έναν άνθρωπο να επινοήσει μια μεγάλη ακολουθία αριθμών χωρίς να υποπέσει σε κάποιο είδος επαναλαμβανόμενου "μοτίβου" . (Δοκιμάστε το και εσείς κάποια φορά .)Θα μπορούσαμε να χρησιμοποιήσουμε ζάρι αν χρειαζόμασταν αριθμούς μόνο από το 1 έως το 6, αλλά για να πάρουμε μια μεγαλύτερη ομάδα αριθμών χρειαζόμαστε κάποιο καλύτερο σύστημα .

Τον παλιό καιρό (πριν από την εμφάνιση των υπολογιστών) , ο καλύτερος τρόπος ήταν να χρησιμοποιήσουμε έναν πίνακα τυχαίων ψηφίων . Ο πίνακας τυχαίων ψηφίων είναι ένας πίνακας που έχει δημιουργηθεί από κάποιων που δουλεία του είναι να δημιουργεί τυχαίους αριθμούς . Οι αριθμοί έχουν ελεγχθεί για να διαπιστωθεί αν περνούν κάποιους συγκεκριμένους έλεγχους τυχαιότητας . Σήμερα μπορείτε να ζητήσετε από τον υπολογιστή σας να δημιουργήσει τους τυχαίους αριθμούς για λογαριασμό σας . Τα περισσότερα υπολογιστικά συστήματα διαθέτουν ενσωματωμένες γεννήτριες τυχαίων αριθμών . Οι αριθμοί που δημιουργούνται με αυτόν τον τρόπο δεν είναι γνήσιοι τυχαίοι αριθμοί , γιατί δημιουργούνται σύμφωνα με κάποιο συγκεκριμένο και αμετάβλητο κανόνα . Ωστόσο , επειδή αυτός ο κανόνας είναι αρκετά απρόβλεπτος , σε όλες σχεδόν τις πρακτικές χρήσεις οι αριθμοί φαίνεται να έχουν επιλέγει εντελώς τυχαία .

3)ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

Στα προαναφερόμενα αντιμετωπίσαμε ένα συγκεκριμένο πρόβλημα έλεγχου υποθέσεων : αν στρίψουμε ένα κέρμα πολλές φορές μπορούμε να εξακριβώσουμε αν είναι κάλπικο ; Στην συνέχεια θα εξετάσουμε μια

γενικότερη αντιμετώπιση της μεθοδολογίας που χρησιμοποιούν οι στατιστικοί όταν διατυπώνουν και ελέγχουν υποθέσεις .

Θυμηθείτε ότι η υπόθεση που πρόκειται να ελέγξουμε ονομάζεται **μηδενική υπόθεση** ή **υπόθεση μηδέν** (null hypothesis) - συμβολίζεται H_0 - και η υπόθεση ότι " η μηδενική υπόθεση είναι λανθασμένη " ονομάζεται **εναλλακτική υπόθεση** (alternative hypothesis). Ακολουθούν μερικά παραδείγματα μηδενικών υποθέσεων :

Ένα κέρμα είναι σωστό .

Το μέσο πλήθος σταφίδων σε μια συγκεκριμένη μάρκα κέικ με σταφίδες είναι 70.

Οι διάφορες στα αποτελέσματα τεσσάρων σιροπιών για το βήχα οφείλονται στην τύχη και μόνο .

Η συχνότητα των υποθέσεων που καταλήγουν στον Αρειό Πάγο ακολουθεί την κατανομή Poisson.

Αν αποφασίσουμε να απορρίψουμε τη μηδενική υπόθεση , αυτό σημαίνει πως είμαστε σχεδόν βέβαιοι ότι η υπόθεση δεν είναι αληθής . Πιο συγκεκριμένα , συνήθως σχεδιάζουμε τους έλεγχοι έτσι ώστε να υπάρχει μόνο 5% πιθανότητα να απορρίψουμε την υπόθεση ενώ στην πραγματικότητα είναι αληθής . Πάντως , αν αποφασίσουμε να αποδεχτούμε την υπόθεση , αυτό δεν σημαίνει αναγκαστικά ότι η υπόθεση είναι πράγματι αληθής . Σημαίνει απλώς ότι δεν έχουμε βρει ακόμη επαρκή στατιστικά στοιχεία για να την απορρίψουμε .

Μια σημαντική αρχή του έλεγχου των υποθέσεων είναι ότι πρέπει να διατυπώνουμε τις υποθέσεις μας πριν συλλέξουμε τα δεδομένα . Αν διατυπώσουμε την υπόθεση μετά την συλλογή των δεδομένων , είναι πολύ εύκολο να θεωρήσουμε λανθασμένα ως αποτέλεσμα κάτι που στην πραγματικότητα αποτελεί απλώς σύμπτωση . Για παράδειγμα , ας υποθέσουμε ότι συλλέξαμε κάποια δεδομένα και παρατηρήσαμε ότι το 20% των ανθρώπων του δείγματος μας έχουν το επώνυμο Παπαδόπουλος και προχωρήσαμε στον έλεγχο της υπόθεσης με βάση τα δεδομένα . Είναι προφανές ότι θα δεχτούμε την υπόθεση γιατί τη διατυπώσαμε μετά τη συλλογή δεδομένων .

Κως θα συλλέγουμε όλο και περισσότερα δεδομένα , θα χρειάζεται να τροποποιούμε τις υποθέσεις μας . Το βασικό σημείο , όμως, είναι ότι δεν

μπορούμε να χρησιμοποιήσουμε κάποια δεδομένα για να ελέγξουμε μια υπόθεση όταν η υπόθεση είχε βασιστεί εξαρχής στα ίδια αυτά δεδομένα . Η λύση είναι μια : να συλλέξουμε περισσότερα δεδομένα . Αν η υπόθεση φαίνεται να είναι σωστή ακόμη και για τα δεδομένα που συλλέξαμε μετά τη διατύπωση της , είναι πιθανό να βρισκόμαστε στο σωστό δρόμο .

Μια άλλη σημαντική αρχή είναι ότι πρέπει να ορίζουμε το *επίπεδο σημαντικότητας* (significance level) πριν από τη συλλογή των δεδομένων . Με αυτόν τον τρόπο εξασφαλίζουμε την αντικειμενικότητα μας και δεν καταλήγουμε στο επίπεδο σημαντικότητας αφού δούμε την τιμή του στατιστικού στοιχείου έλεγχου , όποτε και μπορούμε να αλλάξουμε την τιμή του ανάλογα με την αναμενόμενη τιμή του αποτελέσματος.

1.Οι βασικοί όροι του έλεγχου υποθέσεων που πρέπει να γνωρίζουμε είναι οι εξής :

μηδενική υπόθεση : η υπόθεση που θέλουμε να ελέγξουμε (συμβολίζεται με H_0)

εναλλακτική υπόθεση : η υπόθεση ότι " η μηδενική υπόθεση είναι λανθασμένη "

σφάλμα τύπου I: να συμπεράνουμε ότι η μηδενική υπόθεση είναι *ψευδής* ενώ στην πραγματικότητα είναι *αληθής*

σφάλμα τύπου OII: να συμπεράνουμε ότι η μηδενική υπόθεση είναι *αληθής* ενώ στην πραγματικότητα είναι *ψευδής*

2. Η συνηθισμένη διαδικασία έλεγχου μιας υπόθεσης είναι να υπολογίσουμε , βασιζόμενοι στις παρατηρήσεις μας , μια ποσότητα που ονομάζεται *στατιστικό στοιχείο έλεγχου*

3.Αν η μηδενική υπόθεση είναι αληθής, το στοιχείο έλεγχου θα είναι μια *τυχαία μεταβλητή* με γνωστή κατανομή .

4. Αν φαίνεται πιθανό να ακολουθεί η υπολογισμένη τιμή αυτή την κατανομή , αποδεχόμαστε τη μηδενική υπόθεση .

5. Αν η υπολογισμένη τιμή του στοιχείου έλεγχου φαίνεται μάλλον απίθανο να ακολουθεί αυτή την κατανομή , απορρίπτουμε την μηδενική υπόθεση .

4) ΣΥΣΧΕΤΙΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Στη στατιστική, συχνά χρειάζεται να απαντήσουμε σε ερωτήσεις αυτής της μορφής: Υπάρχει σχέση ανάμεσα σε δυο ποσότητες; Προκαλούν οι αλλαγές στη μια ποσότητα αλλαγές στην άλλη ποσότητα; Για παράδειγμα, επηρεάζουν οι αλλαγές των επιτοκίων τη ζήτηση για κατοικίες; Κάποιες άλλες φορές πάλι χρειάζεται να μάθουμε αν μπορούμε να χρησιμοποιήσουμε μια μεταβλητή για να προβλέψουμε την τιμή μιας άλλης μεταβλητής. Για παράδειγμα, ως υποθέσουμε ότι θέλουμε να υπολογίσουμε την τιμή των καταναλωτικών δαπανών της επόμενης χρονιάς. Ίσως είναι δύσκολο να προβλέψουμε τις καταναλωτικές δαπάνες απευθείας, αλλά μπορεί να είναι σχετικά εύκολο να προβλέψουμε το διαθέσιμο (περισσευόμενο) εισόδημα. Αν καταφέρουμε να βρούμε σχέση ανάμεσα στο εισόδημα και την κατανάλωση, θα βρισκόμαστε σε πολύ καλό δρόμο, γιατί θα μπορούσαμε να χρησιμοποιήσουμε αυτή τη σχέση καθώς και τη γνώση του ύψους του διαθέσιμου εισοδήματος για να προβλέψουμε το ύψος των καταναλωτικών δαπανών.

Στα οικονομικά υποθέτουμε ότι το επίπεδο του εισοδήματος επηρεάζει την ποσότητα που θα ζητηθεί από κάποιο συγκεκριμένο αγαθό. Για τα περισσότερα αγαθά, το υψηλότερο εισόδημα οδηγεί σε μεγαλύτερη ζήτηση. Αυτό, όμως, δεν συμβαίνει πάντα. Υπάρχουν κάποια αγαθά - που ονομάζονται **κατώτερα αγαθά** (inferior goods) - τα οποία οι καταναλωτές τα αγοράζουν σπανιότερα όταν αυξάνεται το εισόδημα τους. (Μάλλον αγοράζουν κάτι καλύτερο όταν μπορούν να διαθέσουν τα χρήματα. Για παράδειγμα, τα όσπρια θα μπορούσαν να θεωρηθούν κατώτερα αγαθά αν οι άνθρωποι προτιμούσαν να αγοράζουν μπριζόλες όταν αυξάνεται το εισόδημα τους.) Ο μόνος τρόπος για να διαπιστώσει κανείς αν κάποιο συγκεκριμένο αγαθό είναι κατώτερο αγαθό είναι να συγκεντρώσει στοιχεία.

Ας υποθέσουμε ότι έχουμε μετρήσεις για το μέσο εισόδημα και τις συνολικές πωλήσεις πίτσας για μια περίοδο ενός μηνός σε οκτώ διαφορετικές πόλεις:

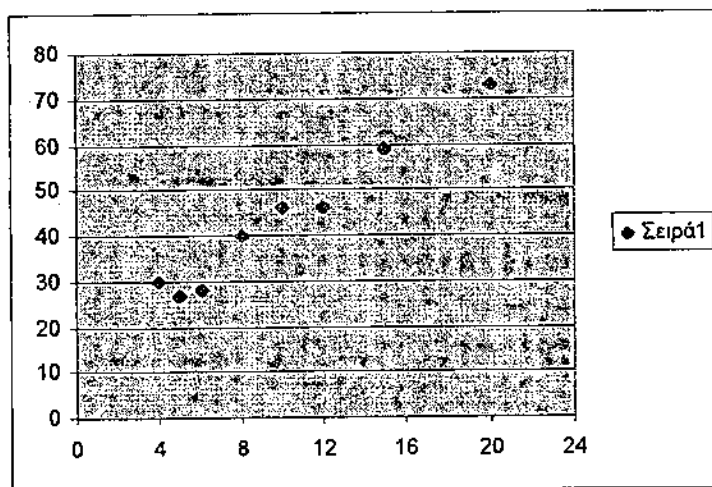
Πωλη	Ημερομίσθιο	Πωλήσεις πίτσας
1	5	27
2	10	46
3	20	73
4	8	40
5	4	30
6	6	28
7	12	46
8	15	59

Όταν πρόκειται να αντιμετωπίσουμε κάποιο πραγματικό πρόβλημα .

Είναι καλύτερα να έχουμε περισσότερες μετρήσεις , αλλά και αυτό το δείγμα των οκτώ μετρήσεων αρκεί για να δούμε τους υπολογισμούς

Ένας καλύτερος τρόπος για να καθορίσουμε αν υπάρχει σχέση ανάμεσα στο εισόδημα και τις πωλήσεις πίτσας είναι να σχεδιάσουμε μια "εικόνα " . Θα σχεδιάσουμε , λοιπόν, ένα γράφημα στο οποίο το εισόδημα θα εμφανίζεται στον οριζόντιο άξονα , οι πωλήσεις πίτσας στο κατακόρυφο , και κάθε μέτρηση θα σημειώνεται με μια κουκίδα . Αυτό το είδος γραφήματος ονομάζεται **διάγραμμα διασποράς** (scatter diagram ή scatter plot)

Εισόδημα (χ)	
€	£
5	27
10	46
20	73
8	40
4	30
6	28
12	46
15	59



Από το διάγραμμα μπορούμε να δούμε καθαρά ότι υπάρχει κάποια σχέση ανάμεσα στο εισόδημα και τις πωλήσεις πίτσας , και μπορούμε επίσης να δούμε ότι μια αύξηση του εισοδήματος οδηγεί σε

αύξηση των πωλήσεων πίτσας . Επόμενος , μπορούμε να συμπεράνουμε ότι η πίτσα δεν είναι κατώτερο αγαθό. όπως λέει και το ρητό , "μια εικόνα αξίζει όσο χίλιες λέξεις " , και θα πρέπει να μην υποτιμάμε την αξία ενός διαγράμματος για την απεικονίσει και την διασάφηση της φύσης μιας σχέσεις . Ωστόσο θα ήταν χρήσιμο να αναπτύσσαμε έναν ακριβέστερο τρόπο χαρακτηρισμού αυτού του είδους της σχέσης . Το εργαλείο που θα χρησιμοποιήσουμε ονομάζεται *ανάλυση παλινδρόμησης* (regression analysis) .Θα εξετάσουμε την *απλή γραμμική παλινδρόμηση* (simple linear regression), η οποία εφαρμόζεται όταν υπάρχει μόνο μια ανεξάρτητη μεταβλητή που να επηρεάζει την τιμή της εξαρτημένης μεταβλητής .Στην πορεία θα ασχοληθούμε με περιπτώσεις που υπάρχουν περισσότερες από μια ανεξάρτητες μεταβλητές , και για τις οποίες θα χρησιμοποιήσουμε την *πολλαπλή γραμμική παλινδρόμηση* (multiple linear regression).

5) ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ

Η διαπίστωση ότι , για τη μελέτη της δομής , της οργάνωσης και της λειτουργίας της κοινωνίας , χρησιμοποιείται μεγάλος αριθμός αλληλεξαρτώμενων μεταβλητών , οδήγησε τους ερευνητές των κοινωνικών επιστημών και της ανθρώπινης συμπεριφοράς να εφαρμόσουν νέες τεχνικές ανάλυσης . Μεταξύ των τεχνικών σε μεγάλο βαθμό χρησιμοποιείται η τεχνική με το ευρύτερο όνομα γνωστή ως *παραγοντική ανάλυση* .

Η ανάγκη για την χρησιμοποίηση τέτοιας τεχνικής προέκυψε από το γεγονός ότι ο ερευνητής των κοινωνικών επιστημών και των επιστημών της ανθρώπινης συμπεριφοράς δεν γνωρίζει ούτε ποιες είναι σπουδαιότερες μεταβλητές ούτε είναι σε θέση να αντιληφθεί τις συσχετίσεις μεταξύ τους . Ακόμη , οι μεταβλητές που μπορεί να παρατήρηση συνδέονται , τις περισσότερες φορές , πολύ ασθενώς με εκείνες τις μεταβλητές που μετρούν τις έννοιες για τις οποίες πραγματικά ενδιαφέρεται . Με αλλά λόγια , κινείται σε μια σφαίρα πολύπλοκων και αλληλεξαρτώμενων , λίγο η πολύ , μεταβλητών και ακαθόριστων εννοιών που περιορίζουν την άμεση μέτρηση . Με αυτήν την άποψη , η παραγοντική ανάλυση υποκαθιστά το πλήθος των αλληλεξαρτώμενων μεταβλητών με ομάδα παραγόντων -υποθετικές ενότητες -που συμπεριφέρονται κατά τον ίδιο τρόπο με τις στατιστικά σημαντικές μεταβλητές που περιέχονται σε αυτούς .

Με την παραγοντική ανάλυση ο ερευνητής είναι σε θέση να γνωρίσει την οργανική συνοχή των μεταβλητών που χρησιμοποιεί , να βρει δηλαδή τις περισσότερο η και λιγότερο αντιπροσωπευτικές

για τη μεγαλύτερη ή μικρότερη σημαντικότητα . Όπως σημειώνει ο Cartel (1977:16) , <<η παραγοντική ανάλυση , λειτουργούσα ως είδος ραντάρ , ειδοποιεί τον ερευνητή να αποφύγει τόσο το ασήμαντο όσο και το μη πραγματικό , γιατί δίνει έστω και χονδρικά στην αρχή , την μορφή των πραγματικών δομών που κρύβονται σε ποικιλώνυμο και αλληλεπιδρούν πλήθος μεταβλητών >>.

Οι παράγοντες είναι λειτουργικά ανεξάρτητη μεταξύ τους και δείχνουν την οργανική συνοχή των μεταβλητών που έχουν χρησιμοποιηθεί . Με άλλα λόγια , η παραγοντική ανάλυση είναι στατιστική τεχνική που δείχνει τον τρόπο ομαδοποίησης διάφορων ανεξάρτητων ποσοτικών μεταβλητών , ατόμων ή ομάδων . Π.χ. , μια κοινότητα μπορεί να περιγράψει με την χρησιμοποίηση μεγάλου αριθμού μεταβλητών , όπως το δείκτη εκβιομηχάνισης , τις εμπορικές δραστηριότητες , τον αριθμό και το μέγεθος των γεωργικών εκμεταλλεύσεων , το μέγεθος του πληθυσμού , την ανεργία , την κινητικότητα (κοινωνική , επαγγελματική , γεωγραφική) , το μέσο οικογενειακό εισόδημα , την κατοικία , το βιοτικό επίπεδο , τους δημογραφικούς δείκτες , τις συνθήκες υγείας , το επίπεδο μόρφωσης , κ.α. Όμως , η περιγραφή μιας κοινότητας θα μπορούσε σε πολύ μεγάλο βαθμό να απλουστεύσει αν υπήρχε δυνατότητα αναγνώρισης κάποιων διαστάσεων (παραγόντων) της κοινότητας αυτής , όπως π.χ. του βαθμού αστικοποίησης , της κοινωνικής και οικονομικής ευημερίας , της επενδυτικής δραστηριότητας και των εισροών , κ.λ.π.

Η παραγοντική ανάλυση πλεονεκτεί της τεχνικής της πολλαπλής παλινδρόμησης , καθόσον αντιμετωπίζεται επιτυχώς το πρόβλημα της πολυσυγγραμμικότητας (Nieuwoudt 1972, Tench 1975) και αποτελεί την καταλληλότερη μέθοδο στην ανάλυση των σχέσεων μεταξύ των κοινωνικοψυχολογικών χαρακτηριστικών των ατόμων . Βέβαια , μπορεί να έχει εφαρμογή σε πλήθος ερευνητικών αναζητήσεων , όπου το πλήθος των μεταβλητών καθιστά δυσχερή οποιαδήποτε άλλη στατιστική ανάλυση , και το σπουδαιότερο , όπου επιζητούνται συμπεράσματα άμεσης πρακτικής σημασίας με βάση τον τρόπο δόμησης των διερευνώμενων χαρακτηριστικών (μεταβλητών) .

Η τεχνική της παραγοντικής ανάλυσης βασίζεται στην αλληλοσυσχέτιση των μεταβλητών . Με την χρησιμοποίηση της μήτρας R των συντελεστών συσχετίσεις αποκτάται η μήτρα F παραγόντων , με τη διάφορα ότι , ενώ η πρώτη έχει τον ίδιο αριθμό σειρών και στηλών με αυτόν των μεταβλητών , αλλά στήλες τόσες όσοι είναι οι παράγοντες . Ο καθένας από τους παράγοντες έχει τη δική του ταυτότητα και περιλαμβάνει ομάδα μεταβλητών με κοινά χαρακτηριστικά (μεταβλητές που συσχετίζονται) . Οι παράγοντες είναι διανύσματα και μάλιστα ψευδοδιανύσματα ή λανθάνοντα διανύσματα , της μορ-

φής $n \times 1$, ορθογώνιοι (ασυσχέτιστοι μεταξύ τους), που αντιπροσωπεύουν (όπως οι κανονικοποιημένες μεταβλητές (ψευδομεταβλητές), συναρτώμενες από στατιστικά σημαντικές μεταβλητές που περιέχονται στους παράγοντες.

Οι συντελεστές συσχέτισης των μεταβλητών με τους αντιστοίχους παράγοντες καλούνται παραγοντικά φορτία, οι τιμές των οποίων μπορεί να είναι στατιστικά σημαντικές ή μη για ορισμένο επίπεδο σημαντικότητας. Κριτήρια για την σημαντικότητα των παραγοντικών φορτίων υπάρχουν διάφορα (Child, Philip κ.α.), ενώ ο Guilford (!) χαρακτηρίζει ένα παραγοντικό φορτίο <<αξιόλογο >> με τιμή μεγαλύτερη του $\pm 0,25-0,30$. Στην πράξη, συνηθίζεται να χρησιμοποιείται ο τύπος που πρωτινέ ο Harman (1976), που δίνει εκτίμηση του τυπικού σφάλματος των φορτίων.

Για την εξαγωγή των παραγόντων χρησιμοποιούνται διάφορες μέθοδοι, οι γνωστότερες από τις οποίες και καθιερωμένες είναι η ανάλυση σε κύριες συνιστώσες (Jolliffe 1986, Dunteman 1989) και η ανάλυση σε κύριους παράγοντες ή αλλιώς ανάλυση σε κοινούς παράγοντες (Mulaik 1972, Gorsuch 1974, Harman 1976, Cattell 1977 και 1978, Rummel 1979).

Για να θεωρούνται αξιόπιστα τα αποτελέσματα οποιασδήποτε από της παρακάτω μεθόδους, πρέπει οι συντελεστές συσχέτισης μεταξύ των μεταβλητών να είναι αξιόπιστοι. Ο Guilford (1975:533) επισημαίνει ότι ένα δείγμα τουλάχιστον 200 παρατηρήσεων θεωρείται ικανοποιητικό για τον περιορισμό των σφαλμάτων στους συντελεστές συσχέτισης και τη σταθερότητα των αποτελεσμάτων της παραγοντικής ανάλυσης.

Ενώ λογαριθμική ανάλυση, είτε ως ανάλυση των λογάριθμων των συχνοτήτων είτε η ως ανάλυση του λόγου πιθανοτήτων, χρησιμοποιείται έναντι των στατιστικών μεθόδων ανάλυσης σε πίνακες διασταυρώσεων λόγω των σημαντικών πλεονεκτημάτων στη χρησιμοποίηση στατιστικών καλής προσαρμογής των υποδειγμάτων στα δεδομένα και της χρησιμοποίησης εκτιμητών παραμέτρων, εμφανίζει σημαντικές αδυναμίες που έχουν να κάνουν με το μέγεθος του δείγματος των παρατηρήσεων. Έτσι, αν το δείγμα είναι πολύ μικρό, το στατιστικό χ^2 στο οποίο βασίζεται το υπόδειγμα της λογαριθμογραμμικής ανάλυσης είναι ύποπτο ως προς την ακρίβεια του και συνεπώς την αξιοπιστία του. Αντιθέτως, εάν το δείγμα είναι πολύ μεγάλο, δυσχεραίνεται ο ερευνητής να κατάληξη σε κάποιο φειδωλό υπόδειγμα και αδυνατεί να διάκρινε μεταξύ υποδειγμάτων που προσαρμόζονται στα δεδομένα. Καθώς ο αριθμός των μεταβλητών και των τιμών τους (κατηγóriων) αυξάνονται, τα υποδείγματα απαιτούν περισσότερες παραμέτρους, έτσι που πρακτικά ο ερευνητής

αντιμετωπίζει δυσκολίες στην ερμηνεία των εκτιμητών των παραμέτρων αυτών .

Τεχνικές εναλλακτικές ή και συμπληρωματικές της λογαριθμο-γραμμικής ανάλυσης είναι ευρύτερα γνωστές με το όνομα τεχνικές άριστης κλιμάκωσης , που οφείλουν την προέλευση τους στην ιδέα της πολυδιαστατικής κλιμάκωσης . Κοινό γνώρισμα των τεχνικών αυτών είναι η χρησιμοποίηση , επάλληλα , της διαδικασίας ελάχιστων τετραγώνων , με την ευρύτερη έννοια ότι κατά την πρώτη φάση εκτίμησης του υποδείγματος αριστοποιείτε κριτήριο ως προς τις παραμέτρους το υποδείγματος , με τις τιμές των δεδομένων όπως αυτές έχουν και κατά τη δεύτερη φάση αριστοποίησης της κλίμακας αριστοποιείτε κριτήριο ως προς τις παραμέτρους των δεδομένων , με τις παραμέτρους του υποδείγματος όπως αυτές έχουν (Κοοϊή και Meulman 1997) . Κατά την άριστη κλιμάκωση , μεταβάλλοντας το δεδομένο επίπεδο μέτρησης των μεταβλητών , μπορούν να αναζητηθούν λύσεις που είναι οι καλύτερα προσαρμοσμένες στα δεδομένα . Παράλληλα , μπορούν να αποκαλυφθούν μη γραμμικές σχέσεις , σε αντίθεση με τους τυπικούς έλεγχοι στατιστικών υποθέσεων στα πλαίσια της ισχύος της κανονικότητας και της γραμμικότητας κατά την ανάλυση παλινδρομήσεις των αρχικών μεταβλητών .

Στις τεχνικές άριστης κλιμάκωσης , υπάγονται η ανάλυση αντιστοιχιών , η οποία χρησιμοποιείται για δίπλευρους πίνακες διασταυρώσεων , η ανάλυση ομοιογενείας , η μη γραμμική ανάλυση σε κύριες συνιστώσες , η κατηγορική παλινδρόμηση και η ανάλυση μη γραμμικής κανονικοποιημένης συσχέτισης (πρόκειται για ανάλυση ομοιογενείας ομάδων μεταβλητών) .

Οι παραπάνω τεχνικές , όπως και η παραγοντική ανάλυση , ανήκουν στην ευρύτερη κατηγορία πολλαπλών στατιστικών αναλύσεων με το όνομα αναλύσεων ανάγωγης , με την ιδιότητα τους όπου σχέσεις μεταξύ μεταβλητών αντιπροσωπεύονται με μικρό αριθμό <<μορφωμάτων >>(διαστάσεων, δομημάτων) , συνήθως δυο ή τριών . Αυτό καθιστά δυνατή και εύκολα ερμηνεύσιμη την περιγραφή των σχέσεων μεταξύ των μεταβλητών .

Σε όλες τις τεχνικές άριστης κλιμάκωσης είναι αναγκαία η κατηγοριοποίηση(κωδικοποίηση) των κατηγοριών των μεταβλητών με τη χρησιμοποίηση ακέραιων αριθμών. Η επιλογή των αριθμών αυτών καθορίζονται απολύτως από τον ερευνητή.

Για τις μεταβλητές που θα χρησιμοποιηθούν ως ονομαστικές ή ως τακτικές το εύρος των κατηγοριών δεν επηρεάζει την ανάλυση, γι' αυτό και η χρησιμοποίηση οποιασδήποτε επινόησης είναι αδιάφορη. Για τις ονομαστικές μεταβλητές σημασία έχει μόνο το όνομα της

καθεμίας από τις κατηγορίες και όχι η τιμή που συνδέεται με το όνομα αυτό. Για τις τακτικές μεταβλητές, επειδή με τις ποσοτικοποιήσεις των μεταβλητών διατηρείται η σειρά των κατηγοριών, οι τιμές των κατηγοριών αυτές καθ'αυτές δεν έχουν και τόση σημασία. Όλες οι κωδικοποιήσεις, αφού δεν αλλάζουν τη σειρά των κατηγοριών, δίνουν τα ίδια αποτελέσματα. Βέβαια, μολονότι οι κωδικοποιήσεις των ονομαστικών και τακτικών μεταβλητών οδηγούν στα ίδια αποτελέσματα, χρησιμοποιούνται κατά κανόνα κωδικοποιήσεις με μικρές διαφορές μεταξύ των κωδικών αριθμών, για την αποφυγή της επιδράσεως τους στο μέγεθος του αποτελέσματος που συχνά αποβαίνει σύνθετο και δυσερμήνευτο. Έτσι, συνήθως χρησιμοποιούνται για τις παραπάνω μεταβλητές διαδοχικοί ακέραιοι αριθμοί (π.χ. οι αριθμοί 1,2,3,4 κ.λ.π.)

Αντιθέτως, κωδικοποιήσεις που αφορούν αριθμητικές μεταβλητές υπόκεινται σε περισσότερους περιορισμούς. Έτσι, για τις μεταβλητές αυτές οι διαφορές μεταξύ διαδοχικών κατηγοριών είναι μεγαλύτερης σημασίας, οποιαδήποτε δε κωδικοποίηση πρέπει να διατηρεί τις διαφορές μεταξύ των κατηγοριών μετά την ποσοτικοποίηση τους. Ένας τρόπος για αυτό θα μπορούσε να είναι η χρησιμοποίηση ως κωδικών των ίδιων αρχικών τιμών της μεταβλητής. Αυτό όμως θα είχε ως αποτέλεσμα την διατήρηση ενδεχομένως πολλών κενών κατηγοριών, που ίσως να οδηγούσε σε δυσχέρεια στην ερμηνεία των αποτελεσμάτων, ιδιαίτερα όταν η μέγιστη κατηγορία απείχε πολύ της μονάδας και υπήρχαν κενές κατηγορίες μεταξύ της μονάδας και της μέγιστης τιμής.

Τρόπος αντιμετώπισης της κατάστασης είναι η επανακωδικοποίηση των κατηγοριών, όχι όμως με την χρησιμοποίηση διαδοχικών αριθμών που μπορεί να οδηγήσει στη διακινδύνευση των μετρικών γνωρισμάτων της αριθμητικής μεταβλητής. Η επανακωδικοποίηση αυτή, με την οποία εξασφαλίζεται η διατήρηση των διαφορών μεταξύ των κατηγοριών, γίνεται με την αφαίρεση της μικρότερης τιμής (κατηγορίας) από όλες τις τιμές και άθροιση, στη συνέχεια, της μονάδας στις διαφορές αυτές. Έτσι, οι νέοι κωδικοί των κατηγοριών έχουν ελάχιστοι τιμή την μονάδα και όλες οι διαφορές είναι ταυτόσημες με τα αρχικά δεδομένα.

Οι τεχνικές άριστης κλιμάκωσης χειρίζονται δεδομένα εκφρασμένα μόνο σε ακέραιους και θετικούς αριθμούς, περικόπτει τα δεκαδικά στοιχεία στους δεκαδικούς και αγνοεί τις αρνητικές παρατηρήσεις, πράγμα που συχνά δεν είναι επιθυμητό για τον ερευνητή. Ακόμη, εάν μια μεταβλητή έχει περισσότερες κατηγορίες (τιμές) απ' ό,τι πρακτικά μπορούν να ερμηνευθούν, χρειάζεται να γίνει κάποια τροποποίηση πριν από την ανάλυση, προκειμένου να περιοριστεί σε περισσότερο εύχρηστο ο αριθμός των κατηγοριών.

Για παράδειγμα, όταν μια μεταβλητή λαμβάνει πολύ μεγάλο αριθμό κατηγοριών(τιμών),αρκετά σύνθετο και πρακτικά μη ερμηνεύσιμο , είναι σκόπιμο για την απλοποίηση της ανάλυσης να διαιρεθούν οι τιμές της μεταβλητής δια του 10 ή δια του 100,να γίνει στρογγυλοποίηση του αποτελέσματος στην πλησιέστερη μονάδα και αφαίρεση στην συνέχεια της μικρότερης από τις λοιπές τιμές και άθροιση της μονάδας στις προκύπτουσες διάφορες .Στην περίπτωση αρνητικών τιμών, αθροίζεται σε όλες τις παρατηρήσεις κάποιος σταθερός αριθμός(συνήθως ο αμέσως μεγαλύτερος του μικρότερου με απόλυτη τιμή αρνητικού)και στη συνέχεια ακολουθεί ή όχι η διαίρεση του αποτελέσματος δια του 10(με άθροιση της μονάδας στη συνέχεια)ή δια του 100 και η στρογγυλοποίηση του στον πλησιέστερο ακέραιο.

6)ΑΝΑΛΥΣΗ ΟΜΟΙΟΓΕΝΕΙΑΣ

Η ανάλυση ομοιογενείας, σε αντίθεση με την ανάλυση αντιστοιχιών, χρησιμοποιείται για την ανάλυση ανά δυο ονοματικών μεταβλητών σε πολυδιάστατους πίνακες, αγνοώντας τις τυχόν αλληλεπιδράσεις μεταξύ των μεταβλητών, αποσκοπώντας να μεγιστοποιήσει την ομοιογένεια αριθμού μεταβλητών(Burg κ.α. 1988).Είναι ανάλυση πολλαπλών αντιστοιχιών, όπου η μορφή της μήτρας των δεδομένων για ανάλυση παριστάνεται με σειρές που αντιπροσωπεύουν τα υποκείμενα(παρατηρήσεις)και στήλες που αντιπροσωπεύουν τις μεταβλητές.

Η ανάλυση ομοιογενείας μέσω της διαδικασίας της τεχνικής των ελάχιστων τετράγωνων ποσοτικοποιεί με άριστες τιμές κάθε κατηγορία μεταβλητής, έτσι ώστε οι κατηγορίες τις καθεμίας μεταβλητής να έχουν το μέγιστο εύρος διασποράς. Προκείμενου για λύσεις δυο διαστάσεων η ανάλυση ομοιογενείας δίνει μια δεύτερη ομάδα βαθμών στις κατηγορίες της καθεμίας μεταβλητής ,επιδιώκοντας και πάλι στη μεγιστοποίηση του εύρους διασποράς των κατηγοριών αυτών, κ.ο.κ Εξαιτίας του γεγονότος ότι κατά την ανάλυση ομοιογενείας οι κατηγορίες καθεμίας μεταβλητής λαμβάνουν τόσους βαθμούς όσες και οι διαστάσεις της ανάλυσης, οι μεταβλητές θεωρούνται ,από την άποψη της μέτρησης ,πολλαπλές ονοματικές.

Παράλληλα με την ποσοτικοποίηση των κατηγοριών των μεταβλητών, η ανάλυση ομοιογενείας ποσοτικοποιεί και τα υποκείμενα της ανάλυσης ως προς δεδομένη διάσταση, έτσι που η βαθμολογία ως προς καθεμία κατηγορία εκφράζεται με τον αριθμητικό μέσο όρο ή, με αλλά λόγια, με το κεντροειδές της βαθμολογίας των υποκείμενων(παρατηρήσεων)της συγκεκριμένης κατηγορίας. Οι βαθμοί των υποκείμενων μπορούν να χρησιμοποιηθούν ,όπως στην παραγοντική

ανάλυση, με την εφαρμογή λοιπών μεθόδων στατιστικής ανάλυσης(π.χ. ανάλυσης παλινδρόμησης)

Η ανάλυση ομοιογενείας ελαχιστοποιεί την αποτυχία προσαρμογής εκφρασμένης ως προς τα τετράγωνα των αποκλίσεων των τυποποιημένων ποσοτικοποιημένων τιμών των υποκείμενων, εάν επιχειρείται να βρεθεί λύση όπου τα υποκείμενα εντός της ίδιας κατηγορίας μεταβλητής αποτυπώνονται γραφικά το ένα κοντά στο άλλο, ενώ υποκείμενα διαφορετικών κατηγοριών αποτυπώνονται απομακρυσμένα μεταξύ τους. Με άλλα λόγια, οι κατηγορίες των μεταβλητών επιμερίζουν τα υποκείμενα σε ομοιογενείς υποομάδες ,ενώ βέβαια και οι μεταβλητές θεωρούνται ομοιογενείς αφού ταξινομούν τα υποκείμενα εντός των ίδιων υποομάδων(στοιχείο της επιτυχημένης ανάλυσης ομοιογενείας).

Η ανάλυση ομοιογένειας επαναλαμβάνεται στους υπολογισμούς της, μέχρι το σημείο όπου δυο διαδοχικές επαναλήψεις των αναλύσεων δώσουν διάφορα ως προς την συνολική προσαρμογή μικρότερης της τιμής σύγκλισης. Η τιμή σύγκλισης είναι αριθμός κυμαινόμενος μεταξύ 0 και 1, όπου μηδενική τιμή δηλώνει τέλεια προσαρμογή του υποδείγματος της ανάλυσης, ενώ τιμή μοναδιαία δηλώνει πλήρη έλλειψη προσαρμογής.

Η ανάλυση ομοιογενείας μπορεί να χρησιμοποιηθεί, αντί της ανάλυσης αντιστοιχιών, και για δίπλευρους πίνακες διασταυρώσεων ,μολονότι τα αποτελέσματα των δυο αναλύσεων δεν ταυτίζονται μεταξύ τους(η ανάλυση αντιστοιχιών στην περίπτωση αυτή είναι συνήθως προτιμότερη της ανάλυσης της ομοιογενείας, για το λόγο ότι εμφανίζει μοναδικές λύσεις καθώς και ποιότητα ως προς την αντιπροσωπευτικότητα αυτών αλλά και σταθερότητα στην παρεχόμενη πληροφορεία).

Η ανάλυση ομοιογενείας(Leew και Rijckevorsel 1980, Nishisato 1980:ch.5, Young 1981, Meulman 1982, Greenacre 1984:ch.5, Lebart κ.α.1984:ch.6, Tenenhaus και Young 1985, Gifi 1990 και SPSS 1984:ch.8)είναι μορφή ανάλυσης σε κύριες συνιστώσες ονοματικών δεδομένων. Κατά την ανάλυση αυτή οι αρχικές κατηγορικές τιμές ποσοτικοποιούνται (αντικαθίστανται με άριστους βαθμούς)και στην συνέχεια γίνεται εφαρμογή του προγράμματος της ανάλυσης σε κύριες συνιστώσες. Η ανάλυση σε κύριες συνιστώσες δίνει την πρώτη κύρια συνιστώσα ισοδύναμη με την πρώτη διάσταση της ανάλυσης ομοιογενείας. Οι παραγοντικοί βαθμοί της πρώτης κύριας συνιστώσας είναι ίσοι με τους βαθμούς των υποκείμενων και τα παραγοντικά φορτία είναι ίσα με τα τετράγωνα των διακριτών μέτρων. Η δεύτερη διάσταση της ανάλυσης της ομοιογενείας, όμως, δεν είναι ίση με την δεύτερη κύρια συνιστώσα, αφού στην ανάλυση ομοιογενείας οι ονοματολογίες των κατηγοριών των μεταβλητών αντικαθιστού-

νται με αριθμητικές τιμές και συνεπώς μπορούν να εφαρμοστούν ,μετά την ανάλυση ομοιογενείας ,πολλοί διαφορετικοί τρόποι που χρησιμοποιούν αριθμητικά δεδομένα. Το ίδιο ισχύει, όπως θα δούμε στη συνέχεια(μέρος δευτερο,κεφ.τεταρτο και εικοστό, αντίστοιχα),στη μη γραμμική ανάλυση σε κύριες συνιστώσες και στην ανάλυση της μη γραμμικής κανονικοποιημένης συσχέτισης.

Η ανάλυση ομοιογενείας μπορεί να αφορά πολλές διαστάσεις, ο μέγιστος αριθμός των οποίων είναι είτε ο αριθμός των κατηγοριών μείον τον αριθμό των μεταβλητών(αυτών χωρίς ελλείπουσες τιμές),είτε ο αριθμός των παρατηρήσεων μείον ένα, λαμβανόμενου του μικρότερου από τους δυο .Ωστόσο, σπάνια χρησιμοποιείται ο μέγιστος αριθμός των επιτρεπόμενων διαστάσεων ,αφού αφενός μικρότερος αριθμός αυτών είναι ευκολότερα ερμηνεύσιμος και αφετέρου πέρα από κάποιο αριθμό διαστάσεων το εξηγούμενο ποσοστό συνάφειας είναι αμελητέο. Γενικότερα, σε ένα πρόγραμμα ανάλυσης ομοιογενείας δηλώνεται ο αριθμός των διαστάσεων που αναγκαίου για να εξηγήσουν το μεγαλύτερο ποσοστό διακύμανσης .Πρόκειται για αναλύσεις ομοιογενείας μιας ή δυο ή τριών το πολύ διαστάσεων.

Όπως και στην παραγοντική ανάλυση, οι χαρακτηρίσθηκες ρίζες μετρούν την εξηγούμενη διακύμανση, δηλαδή το ποσοστό της κατηγοριοποιημένης πληροφορίας που εξηγείται για καθεμία διάσταση .Η μέγιστη τιμή χαρακτηριστικής ρίζας για κάθε μια διάσταση είναι ίση με τη μονάδα ,διαστάσεις δε με υψηλότερες τιμές της χαρακτηριστικής ρίζας είναι μεγαλύτερης σημασίας στην ολική λύση του προβλήματος.

Οι διαστάσεις συνδυάσθηκα διευκολύνουν στην ερμηνεία ως προς τις θέσεις των υποκείμενων, των μεταβλητών και των κατηγοριών τους. Έτσι π.χ. εάν μια μεταβλητή διαφοροποιείται των άλλων σημαντικά ,τα υποκείμενα θα βρίσκονται πολύ κοντά στις κατηγορίες στις οποίες ανήκουν .Υποκείμενα της ίδιας κατηγορίας θα προσεγγίζονται μεταξύ τους(θα έχουν δηλαδή ίδιους βαθμούς)και οι κατηγορίες των διαφορετικών μεταβλητών θα προσεγγίζονται εφόσον ανήκουν στα ίδια υποκείμενα(δηλαδή δυο υποκείμενα που θα έχουν ίδιους βαθμούς ως προς κάποια μεταβλητή θα έχουν,περιπου,τους ίδιους βαθμούς ως προς τις άλλες μεταβλητές).

Εάν επιθυμούμε να μετρήσουμε χαρακτηριστικά όπως την ικανότητα οδήγησης ,την επιδεξιότητα στη χρήση εκπαιδευτικών μεθόδων , την ικανότητα αυτενέργειας των ατόμων , κ.α. μπορούμε να δημιουργήσουμε για το σκοπό αυτόν κάποιων ειδικό τύπο οργάνου μέτρησης . Ο τύπος αυτός αφορά την κατασκευή κλίμακας (εφαρμογή

έλεγχου) η οποία αποτελείται από ομάδα ειδικών <<θεμάτων >> (ερωτήσεων , προτάσεων , αντικειμένων , κ.λ.π.) , οι απαντήσεις σε καθένα από τα οποία βαθμολογούνται , λαμβάνεται δε στη συνέχεια το συνολικό άθροισμα για το καθένα άτομο . Ανάλογη είναι και η αντιμετώπιση μιας μεταβλητής στάσης , όπου τα επιμέρους θέματα έχουν θέση ομάδας τακτικών μεταβλητών .

Η συνήθης στατιστική μεταχείριση σε τέτοιες πολυθεματικές μεταβλητές είναι ο έλεγχος της ενδοσυνέπειας των θεμάτων ο οποίος πραγματοποιείται είτε με τη χρήση εσωτερικού κριτηρίου (συσχετίσεις ή συνδυακιμάνσεις των θεμάτων ανά δυο μεταξύ τους , ή με τη συσχέτιση καθενός θέματος με το σύνολο της βαθμολογίας των λοιπών θεμάτων) , είτε με τη χρήση εξωτερικού κριτηρίου (Guilford 1975). Ο έλεγχος οδηγεί στον υπολογισμό του συντελεστή αξιοπιστίας α -Cronbach (ή KR-20 των Kuder-Richardson) ή και λοιπών μέτρων αξιοπιστίας .

Με τον όρο αξιοπιστία κλίμακας εννοούμε την ακρίβεια μέτρησης δηλαδή την σχετική ικανότητα στον προσδιορισμό της <<πραγματικής >> αξίας μιας μεταβλητής (Black και Champion , 1976) . Η εκτίμηση της αξιοπιστίας οποιασδήποτε διαδικασίας μέτρησης συνίσταται στον προσδιορισμό του βαθμού διακύμανσης της βαθμολογίας των ατόμων , βαθμού που οφείλεται σε πραγματικές διαφορές (και σταθερά σφάλματα) και του βαθμού διακύμανσης που οφείλεται σε ασυνέπειες της μέτρησης . Η αξιοπιστία αναφέρεται στη συνέπεια απόκτησης ίδιων αποτελεσμάτων (δηλαδή στον ίδιο βαθμό διασποράς των τιμών) μιας μεταβλητής σε επαναλαμβανόμενες μετρήσεις .

Η αξιοπιστία κλίμακας στάσης μπορεί να υπολογιστεί με την τεχνική έλεγχου-επανελέγχου , κατά την οποία συσχετίζονται οι βαθμολογίες των απαντήσεων σε δυο διαφορετικές χρονικές στιγμές και μάλιστα από διαφορετικούς ερευνητές και ελέγχεται στατιστικά η σημαντικότητα του βαθμού συσχετίσεις . Συνηθέστατα όμως , και για την αποφυγή επανάληψης της έρευνας , χρησιμοποιούνται τεχνικές εκτίμησης των συντελεστών αξιοπιστίας βασισμένες σε μετρήσεις που έγιναν σε μια μόνο χρονική στιγμή . Οι περισσότερο γνωστές μέθοδοι είναι : α) η μέθοδος των ισοδύναμων τύπων ,β) η μέθοδος του ημίκλαστου με την εφαρμογή του τύπου των Spearman-Brown (Φιλίας κ.α. 1977) και Guttman και γ) ο συντελεστής α -Cronbach (Bohmstedt 1977, Cronbach 1951) . Ο συντελεστής ισοδυναμίας των Kuder -Richardson αποτελεί μερική περίπτωση του γενικού συντελεστή α -Cronbach.

Η αξιοπιστία μιας κλίμακας στάσης αυξάνεται σημαντικά με την αύξηση του αριθμού των θεμάτων , αλλά μέχρι ενός ορίου (τιμής) , πέραν του οποίου η αύξηση είναι αμελητέα με την αύξηση του αριθμού των θεμάτων . Εάν από μια ομάδα θεμάτων που έχουν υπο-

βληθεί σε θεματική ανάλυση διατηρηθούν μόνο τα σημαντικά θέματα (αυτά με υψηλούς συντελεστές συσχέτισης), ο συντελεστής αξιοπιστίας που προκύπτει έχει τέτοια τιμή , ώστε η προσθήκη των μη σημαντικών θεμάτων μεταβάλει ελάχιστα την τιμή του .

Συντελεστής αξιοπιστίας μεγαλύτερος του 0,80 θεωρείται πολύ ικανοποιητικός , πολλές φορές μάλιστα στην πράξη γίνονται δεκτοί και συντελεστές αξιοπιστίας μικρότεροι , με τιμές μέχρι και 0,60. Τιμές μικρότερες του 0,60 επιβάλλουν την επινόηση νέων θεμάτων και την επανάληψη της ερευνάς με νέα επαναβαθμολόγηση των απαιτήσεων και υπολογισμό νέου συντελεστή αξιοπιστίας . Ο ερευνητής πολλές φορές είναι σε θέση να αποκτήσει αξιόπιστες μετρήσεις ακόμη και με 8-10 θέματα αρκεί αυτά να έχουν εξασφάλιση την εγκυρότητα της μεταβλητής στάσης .

Θα πρέπει να σημειωθεί ότι ένας έλεγχος πρέπει να είναι αξιόπιστος προκειμένου να είναι χρήσιμος . Όμως , δεν είναι αρκετό να είναι αξιόπιστος . Χρειάζεται να είναι έγκυρος , δηλαδή να μέτρα αυτό που επιδιώκει ο ερευνητής να μετρήσει , συχνά όχι με επιτυχία , ιδιαίτερα στις μεταβλητές στάσης. Ο έλεγχος της εγκυρότητας (περιεχομένου και εγκυρότητας δομής των θεμάτων) γίνεται με την εφαρμογή της παραγοντικής ανάλυσης (Bohmstedt 1977 , Kerlinger και Kaya 1977) . Στο παρόν κεφάλαιο ο εστιασμός θα αφορά την ανάλυση μόνο της αξιοπιστίας πολυθεματικών μεταβλητών .

Πριν γίνει αναφορά για την αξιοπιστία και τη μέτρηση της , είναι σκόπιμη η αναφορά σε ορισμένα βασικά μέτρα της περιγραφικής στατιστικής , αναγκαία για την αξιολόγηση κάποιας κλίμακας .

Καταρχήν , σε ότι αφορά τα μεμονωμένα θέματα , μέτρα είναι οι αριθμητικοί μεσοί όροι και οι τυπικές απόκλιση των τιμών καθενός από τα θέματα της πολυθεματικής μεταβλητής , καθώς και οι συντελεστές συσχέτισης των θεμάτων ανά δυο μεταξύ τους .

Σε ότι αφορά την κλίμακα (στο σύνολο των θεμάτων), μέτρα είναι ο αριθμητικός μέσος όρος , η διακύμανση και η τυπική απόκλιση των τιμών της συνολικής βαθμολογίας των θεμάτων, καθώς και ο μέσος συντελεστής συσχέτισης, η μέση διακύμανση των τιμών του, το μέσο εύρος και οι μέσες ακραίες τιμές των συντελεστών συσχέτισης των θεμάτων ανά δύο μεταξύ τους.

Τέλος, σε ότι αφορά τη σχέση της βαθμολογίας καθενός θέματος με την αθροιστική βαθμολογία του συνόλου των υπολοίπων θεμάτων, μέτρα που θα μπορούσαν να αναφερθούν είναι ο αριθμητικός μέσος και η διακύμανση της βαθμολογίας των θεμάτων πλην ενός, ο διορθωμένος συντελεστής συσχέτισης καθενός θέματος με το σύνολο της βαθμολογίας των λοιπών θεμάτων, ο συντελεστής πολλαπλού προσδιορισμού (το τετράγωνο του συντελεστή πολλαπλής συσχετι-

σης) και τέλος οι τιμές των συντελεστών αξιοπιστίας α -Cronbach, με την απάλειψη διαδοχικά από την κλίμακα καθενός των θεμάτων.

Πολύ συχνά ο ερευνητής επιθυμεί να γνωρίσει τα στοιχεία εκείνα που καθορίζουν τις αποφάσεις των ατόμων (π.χ. τα χαρακτηριστικά αυτοκινήτου για την αγορά του, τα γνωρίσματα ενός υποψηφίου για την εκλογή του, τα προσόντα ενός άτυπης μορφής ηγέτη στην αγροτική κοινότητα για την προσέγγισή του, κ.τ.λ.). Σε όλες τις περιπτώσεις, όπως αντιλαμβάνεται κανείς, όλες οι μεταβλητές που καθορίζουν τις αποφάσεις των ατόμων είναι υποκειμενικές και συνεπώς υποκειμενικές είναι και οι μονάδες μέτρησης των μεταβλητών αυτών.

Δύο είναι οι μέθοδοι κατασκευής αντικειμενικής κλίμακας μέτρησης των μεταβλητών που θα μπορούσε λογικά να ανταποκρίνεται προς την "εσωτερική" κλίμακα - κριτήριο - που χρησιμοποιείται από τα άτομα. Η μια αφορά την απόκτηση πολλαπλών δεδομένων και την εφαρμογή, στη συνέχεια, της παραγοντικής ανάλυσης και η άλλη την απόκτηση "ανόμοιων" δεδομένων - δεδομένων ανομοιοτήτας - και την ανάλυσή τους, στη συνέχεια, με τη χρησιμοποίηση πολυδιαστατικής κλιμάκωσης.

7) ΠΟΛΥΔΙΑΣΤΑΤΗ ΚΛΙΜΑΚΩΣΗ

Πολυδιαστατική κλιμάκωση χαρακτηρίζεται η κλιμάκωση που αποσκοπεί να αναλύσει δεδομένα των οποίων η μέτρηση στηρίζεται στις μεταξύ τους αποστάσεις, δεδομένα που χαρακτηρίζονται για την ανομοιοότητά τους, ή που υποδηλώνουν το βαθμό της ανομοιοτήτας (ή ομοιοτήτας) των υποκειμένων ανά δύο ως προς συγκεκριμένο γνώρισμα. Με άλλα λόγια, η πολυδιαστατική κλιμάκωση επιδιώκει τη δημιουργία χάρτη θέσεων σημείων σχετικών μεταξύ τους από δεδομένα που καθορίζουν την διαφοροποίηση αντικειμένων (ατόμων, γεγονότων, κ.τ.λ.). Για παράδειγμα, η πολυδιαστατική κλιμάκωση χρησιμοποιείται για την ανάλυση δεδομένων που υποδηλώνουν την (αν)ομοιοότητα ζευγών αυτοκινήτων ως προς κάποιο γνώρισμα τους, την (αν)ομοιοότητά ζευγών ατόμων ως προς κάποιο ψυχολογικό γνώρισμα τους, κ.λ.π.

Η πολυδιαστατική κλιμάκωση έχει τις ρίζες της στην ψυχομετρία και χρησιμοποιείται προκειμένου να βοηθήσει τους ειδικούς στην κατανόηση της κριτικής σκέψης των πελατών τους ως προς κάποια ομάδα θεμάτων - ζητημάτων. Σήμερα, η τεχνική της πολυδιαστατι-

κής κλιμάκωσης έχει γενικευθεί στην ανάλυση δεδομένων σε πολλούς και ποικίλους τομείς έρευνας.

Τα δεδομένα μιας πολυδιαστατικής κλιμάκωσης γνωστά ως "ομοιότητες" ή "ανομοιοτήτες" ή "αποστάσεις" ή "εγγύτητες", αντανακλούν το βαθμό της (αν)ομοιότητας μεταξύ ζευγών υποκειμένων.

Κατά βάση, ανομοιοτήτες είναι υποκειμενικά δεδομένα που βασίζονται στο κριτήριο των ερωτώμενων (π.χ. στην κρίση τους ως προς την (αν)ομοιότητα μεταξύ δύο εκπαιδευτικών μεθόδων). Ωστόσο, μπορεί να προκύπτουν και με αντικειμενική μέτρησή τους (π.χ. η συχνότητα επικοινωνίας ατόμων ανά δύο μεταξύ τους). Τέλος, σπανιότερα, ανομοιοτήτες μπορούν να προκύψουν από πολλαπλές παρατηρήσεις (χρησιμοποίηση βαθμολογιών ομάδας βαθμολογητών ανά δύο λαμβανομένων).

Μήτρα ανομοιοτήτων μπορεί να είναι μία απλή μόνο, η οποία παριστάνει συχνότητες (π.χ. συχνότητες στην επικοινωνία μεταξύ ζευγών ατόμων) ή πολλές (ισάριθμες με τον αριθμό των κριτών - ερωτώμενων), καθεμία από τις οποίες χαρακτηρίζει τις (αν)ομοιότητες μεταξύ ζευγών αντικειμένων (π.χ. εκπαιδευτικών μεθόδων), όπως αυτά αξιολογούνται από τα ερωτώμενα άτομα.

Με βάση τα παραπάνω, θα λέγαμε ότι ένα υπόδειγμα χαρακτηρίζεται ως πολυδιαστατικής κλιμάκωσης όταν διερευνά υποκείμενα ή γεγονότα που παριστάνονται με σημεία σε πολυδιάστατο χώρο. Τα σημεία είναι έτσι διευθετημένα, ώστε οι αποστάσεις μεταξύ ζευγών υποκειμένων συνδέονται με τη μεγαλύτερη δυνατή σχέση κατά το βαθμό της ομοιότητας μεταξύ των ζευγών των υποκειμένων. Με άλλα λόγια, δύο όμοια υποκείμενα παριστάνονται με δύο σημεία που προσεγγίζονται μεταξύ τους, ενώ δύο ανόμοια υποκείμενα από δύο σημεία απομακρυσμένα μεταξύ τους. Συνήθως, ο χώρος είναι δις-διαστατικός ή τρις-διαστατικός ευκλείδειος χώρος, οπότε το υπόδειγμα χαρακτηρίζεται ευκλείδειο, αλλά μπορεί να έχει και περισσότερες διαστάσεις. Μερικές φορές χρησιμοποιείται υπόδειγμα που αποσκοπεί να απεικονίσει ατομικές διαφορές ανομοιοτήτων, καλούμενο υπόδειγμα της "κλιμάκωσης ατομικών διαφορών" και το οποίο σταθμίζει τα ερωτώμενα άτομα ως προς τις διαφορές διαστάσεις μέτρησης.

Ο τύπος ανάλυσης της πολυδιαστατικής κλιμάκωσης εξαρτάται από τον αριθμό των μητρών ανομοιοτήτων, το επίπεδο μέτρησης των δεδομένων και το υπόδειγμα που θα χρησιμοποιηθεί για την ανάλυσή τους. Έτσι, η ταξινόμηση των τύπων των αναλύσεων πολυδιαστατικής κλιμάκωσης είναι ανάλογη με τον αριθμό των μητρών και το είδος του χρησιμοποιούμενου υποδείγματος. Διακρίνουμε τέσσερα είδη αναλύσεων: α) την κλασική ανάλυση πολυδιαστατικής κλιμάκωσης, όπου χρησιμοποιείται μία μόνο μήτρα και το ευ-

κλείδειο υπόδειγμα, β) την επαναλαμβανόμενη ανάλυση, με πολλές μήτρες και το ευκλείδειο υπόδειγμα, γ) τη σταθμισμένη ανάλυση με πολλές μήτρες και το σταθμισμένο ευκλείδειο υπόδειγμα (καλούμενο και υπόδειγμα κλιμάκωσης ατομικών διαφορών) και δ) τη γενικευμένη ανάλυση πολυδιαστατικής κλιμάκωσης με πολλές μήτρες και γενικευμένο το ευκλείδειο υπόδειγμα.

Ταξινομήση των τύπων πολυδιαστατικής κλιμάκωσης γίνεται ακόμη ανάλογα με την κλίμακα μέτρησης των ανομοιοτήτων (αποστάσεων). Εάν τα δεδομένα μετρούνται σε τακτική κλίμακα, η πολυδιαστατική κλιμάκωση χαρακτηρίζεται μη μετρική, ενώ εάν μετρούνται σε κλίμακα ισοδιαστημάτων ή αναλογική ή πολυδιαστατική κλιμάκωση χαρακτηρίζεται μετρική. Μετρικές και μη μετρικές μετρήσεις μπορούν να συνδυαστούν με την κλασική, την επαναλαμβανόμενη και τη σταθμισμένη πολυδιαστατική κλιμάκωση, παρέχοντας έξι διαφορετικούς τύπους αυτής.

8) ΟΙΚΟΝΟΜΕΤΡΙΚΕΣ ΜΕΘΟΔΟΙ

Ο συστηματικός έλεγχος της θεωρίας αποτελεί αναπόσπαστο μέρος κάθε επιστήμης. Μια θεωρία, ανεξάρτητη από τη λογική της συνέπεια, δεν μπορεί να σταθεί χωρίς αναφορά στα πραγματικά δεδομένα, χωρίς με άλλα λόγια, κάποιο εμπειρικό έλεγχο. Η Οικονομική επιστήμη, αντίθετα προς τα Μαθηματικά ή τη Λογική, που είναι *a priori* επιστήμες. Επομένως, οι οικονομικές θεωρίες ή προτάσεις μπορούν και πρέπει να ελέγχονται και αξιολογούνται με βάση τα πραγματικά δεδομένα της οικονομικής ζωής.

Ο Διευθυντής της Οικονομικής Σχολής του Λονδίνου Willoam H. Beveridge, στον αποχαιρετιστήριο λόγο του το 1937, επέκρινε τους οικονομολόγους, γιατί μέσα σε μια περίοδο εκατό χρόνων Πολιτικής Οικονομίας, χρησιμοποίησαν τα γεγονότα όχι ως έλεγχο της θεωρίας αλλά ως επεξηγήσεις. Έκτοτε όμως η έμφαση και η εφαρμογή των ποσοτικών μεθόδων στην ανάλυση και ερευνά των οικονομικών φαινομένων αποτελεί το σπουδαιότερο χαρακτηριστικό των εξελίξεων στην Οικονομική επιστήμη.

Η Οικονομετρία αναφέρεται κύριος στην ποσοτική πλευρά της Οικονομικής επιστήμης και προσπαθεί να δώσει εμπειρικό περιεχόμενο στις <<αφηρημένες>> σχέσεις της οικονομικής θεωρίας. Στην Οικονομετρία, η μαθηματικό-οικονομική και η στατιστική ανάλυση και ερευνά χρησιμοποιούνται συνδυασμένα, με κύριο αντικειμενικό σκοπό την εμπειρική εκτίμηση των σχέσεων αυτών αλλά και την επαλήθευση (έλεγχο) της οικονομικής θεωρίας. Ο συνδυασμός αυτός της Μαθηματικής Οικονομικής και της Στατιστικής έχει γίνει ίσως η αφορμή να συγχέεται ή και να ταυτίζεται πολλές φορές το

περιεχόμενο της Οικονομετρίας με αυτό της Μαθηματικής Οικονομικής ή της Στατιστικής, ενώ πραγματικά είναι διαφορετικό. Η οικονομική ανάλυση χρησιμοποιεί τις συναρτησιακές σχέσεις της οικονομικής θεωρίας και αφού τις μετατρέψει σε μαθηματικές, δηλαδή αφού κατασκευάσει ένα *υπόδειγμα* (model), προσπαθεί να τις εκτιμήσει εμπειρικά. Γι' αυτή όμως την εκτίμηση χρησιμοποιεί στατιστικές μεθόδους προσαρμοσμένες στα ιδιαίτερα χαρακτηριστικά των οικονομικών φαινομένων.

Όπως γνωρίζει ο σπουδαστής της Οικονομικής επιστήμης, βασικό χαρακτηριστικό των οικονομικών θεωριών είναι ότι αποτελούν αφαιρέσεις (ή απλουστεύσεις) της πραγματικότητας και πως συνήθως εκφράζονται ως *συναρτησιακές σχέσεις* μεταξύ διάφορων μεταβλητών. Οι συναρτησιακές σχέσεις όμως της οικονομικής θεωρίας είναι σχέσεις *ακριβείς* (exact) ή *προσδιοριστικές* (deterministic). Για παράδειγμα η γνωστή Κευνσιανή συνάρτηση καταναλώσεως:

$$C = \alpha + \beta Y$$

όπου C = δαπάνες καταναλώσεως

Y = διαθέσιμο εισόδημα

είναι σχέση προσδιοριστική, πράγμα που σημαίνει ότι σε κάθε επίπεδο διαθέσιμου εισοδήματος αντιστοιχεί μόνο ένα επίπεδο καταναλώσεως. Είναι φανερό όμως ότι τέτοια συμπεριφορά δεν ανταποκρίνεται στα πραγματικά δεδομένα. Τα σημεία που προσδιορίζονται από τις ετήσιες παρατηρήσεις για την κατανάλωση και το διαθέσιμο εισόδημα για την ελληνική οικονομία στα τελευταία π.χ. τριάντα χρόνια δεν θα βρίσκονται όλα πάνω σε μια ευθεία γραμμή· θα υπάρχουν, με αλλά λόγια, απόκλιση ή διαφορές. Οι απόκλιση από την προσδιοριστική σχέση της οικονομικής θεωρίας μπορούν να ληφθούν υπόψη με την προσθήκη μιας *τυχαίας μεταβλητής*, όποτε η *προσδιοριστική* σχέση της Μαθηματικής Οικονομικής μετατρέπεται σε *στοχαστική*. Στο παράδειγμα μας, η συνάρτηση καταναλώσεως γίνεται:

$$C = \alpha + \beta Y + u$$

όπου u είναι η τυχαία μεταβλητή που επηρεάζει την συμπεριφορά της καταναλώσεως.

Με τη φύση και τα ιδιαίτερα χαρακτηριστικά του τυχαίου όρου u θα ασχοληθούμε αργότερα. Προς το παρόν θέλουμε να τονίσουμε ότι τα οικονομικά φαινόμενα είναι βασικά, φαινόμενα στοχαστικά,

δηλαδή οι οικονομικές σχέσεις περιλαμβάνουν και τυχαίους παράγοντες που αγνοούνται όμως από τη Μαθηματική Οικονομική . Το κενό ανάμεσα στις προσδιοριστικές σχέσεις της οικονομικής θεωρίας και την οικονομική πραγματικότητα καλύπτει η Οικονομετρία με την στοχαστική θεώρηση των οικονομικών σχέσεων . Οι οικονομετρικές σχέσεις είναι σχέσεις στοχαστικές , αντίθετα προς τις σχέσεις της οικονομικής θεωρίας που είναι προσδιοριστικές ή ακριβείς . Επιπλέον , ενώ η οικονομετρική ανάλυση αίνος φαινομένου είναι ποσοτικής φύσεως , η ανάλυση της Μαθηματικής Οικονομικής είναι περισσότερο ποιοτικής φύσεως .

Για παράδειγμα , ας πάρουμε το απλό Κευνσιανο υπόδειγμα προσδιορισμού του εισοδήματος :

$$C = \alpha + \beta Y$$

$$Y = C + I$$

όπου C = κατανάλωση

Y = εισόδημα

I = επένδυση , που λαμβάνεται ως σταθερή (ανεξάρτητα από το εισόδημα) .

Η Μαθηματική Οικονομική , αφού εκφράσει την οικονομική θεωρία , μαθηματικά, δηλαδή αφού κατασκευάσει ένα υπόδειγμα όπως το παραπάνω , ασχολείται με τη συμπεριφορά των ενδογενών ή εξαρτημένων μεταβλητών (στο παράδειγμα μας , η κατανάλωση (C) και το εισόδημα (Y)) , όταν δίνονται οι εξωγενείς ή ανεξάρτητες μεταβλητές (στο παράδειγμα μας η επένδυση (I)) . Με άλλα λόγια , η Μαθηματική Οικονομική , αφού καθορίσει τα πρόσημα των παραμέτρων του υποδείματος (στο παράδειγμα μας α και β) , εξετάζει τις συνέπειες που θα έχουν στις ενδογενείς μεταβλητές οι μεταβολές στις εξωγενείς μεταβλητές . Τι θα συμβεί στο εισόδημα και στην κατανάλωση όταν μεταβληθούν οι επενδύσεις . Όπως ξέρουμε , το εισόδημα θα αυξηθεί ή θα μειωθεί , ανάλογα με το αν οι επενδύσεις αυξήθηκαν ή μειώθηκαν , αλλά η Μαθηματική Οικονομική δε μπορεί να δώσει απάντηση στο ερώτημα πόσο θα αυξηθεί το εισόδημα ή η κατανάλωση αν οι επενδύσεις αυξηθούν κατά ένα συγκεκριμένο πόσο . Για να δοθεί απάντηση στο ερώτημα , θα πρέπει να είναι γνωστές οι τιμές των παραμέτρων του υποδείματος . Στο παράδειγμα μας , θα πρέπει να είναι γνωστή η οριακή ροπή για κατανάλωση , δηλαδή ο συντελεστής β .

Ας σημειωθεί ότι η Μαθηματική Οικονομική μπορεί να καθορίσει , εκτός από τα πρόσημα , και τα όρια για τις τιμές των συντελεστών . Η οριακή ροπή για κατανάλωση π.χ. , είναι μικρότερη από τη μονάδα και μεγαλύτερη από το μηδέν . Πάλι όμως η απάντηση στο προηγμένο ερώτημα δεν θα είναι συγκεκριμένη . Αν π.χ. $\beta = 0,5$, ο

πολλαπλασιαστής του εισοδήματος είναι 2, ενώ αν $\beta=0,9$, ο πολλαπλασιαστής είναι 10. Στην πρώτη περίπτωση, η μεταβολή στο επίπεδο ισορροπίας του εισοδήματος θα είναι το διπλάσιο της μεταβολής της επενδύσεως, ενώ στη δεύτερη θα είναι το δεκαπλάσιο. Για να δοθεί απάντηση στο ερώτημα, απαιτείται η αριθμητική τιμή της οριακής ροπής για κατανάλωση.

Η Οικονομετρία, με την χρησιμοποίηση κατάλληλων μεθόδων, που ονομάζονται *οικονομετρικές*, προβαίνει σε εκτίμηση των παραμέτρων του υποδείγματος. Οι κλασικές στατιστικές μέθοδοι δεν είναι πάντα κατάλληλες για την ανάλυση μετρήσεων που δεν προέρχονται από ένα ελεγχόμενο πείραμα· ένα πείραμα δηλαδή, όπως το εργαστηριακό, όπου ο ερευνητής, μπορεί να ελέγχει όλες τις μεταβλητές ενός φαινομένου και να αφήνει ελεύθερη εκείνη της οποίας την επίδραση θέλει να μελετήσει. Οικονομική επιστήμη στερείται αυτής της << πολυτελείας >>, του ελεγχόμενου πειράματος και των παρατηρήσεων που προέρχονται από αυτό. Οι οικονομικές μετρήσεις προέρχονται από την πραγματική, καθημερινή ζωή και δεν είναι το αποτέλεσμα ελεγχόμενου πειράματος. Οι ετήσιες παρατηρήσεις, π.χ., για τις εισαγωγές και το εισόδημα δεν προέρχονται από κανένα πείραμα όπου παραμένουν σταθεροί όλοι οι άλλοι παράγοντες που επηρεάζουν την συμπεριφορά των εισαγωγών, και μεταβάλλεται μόνο το εισόδημα. Στην πραγματικότητα, όλες οι οικονομικές μεταβλητές μεταβάλλονται ταυτόχρονα και συνεχώς. Για τον οικονομολόγο, το πείραμα είναι δεδομένο και το εργαστήριο είναι η κοινωνία.

Για τους παραπάνω λόγους, για την εμπειρική διερεύνηση των οικονομικών φαινομένων, γίνεται προσαρμογή, όπου απαιτείται των κλασικών στατιστικών μεθόδων στα ιδιαίτερα χαρακτηριστικά των οικονομικών σχέσεων, αφού λάβουμε υπόψη ότι οι οικονομικές μετρήσεις δεν είναι πειραματικές. Η ανάπτυξη και η εφαρμογή κατάλληλων μεθόδων για την μέτρηση των οικονομικών σχέσεων είναι το περιεχόμενο της Οικονομετρίας.

Η οικονομική ανάλυση ενός οικονομικού φαινομένου και η εμπειρική εκτίμηση των παραμέτρων μιας οικονομικής σχέσεως επιχειρείται για έναν ή περισσότερους από τους τρεις παρακάτω σκοπούς :

α) την εμπειρική επαλήθευση ή των έλεγχο της θεωρίας. Ο έλεγχος μιας θεωρίας ή μιας υποθέσεως είναι έλεγχος των προβλέψεων της θεωρίας ή της υποθέσεως. Ο όρος *πρόβλεψη* δεν αναφέρεται μόνο στο μέλλον, αλλά μπορεί να αναφέρεται και στο παρελθόν. Η ορθότητα και η ακρίβεια των προβλέψεων της θεωρίας, η ικανότητα

δηλαδή της θεωρίας να εξηγεί το φαινόμενο , ελέγχεται μόνο με αναφορά στα πραγματικά δεδομένα . Η θεωρία γίνεται δεκτή , ή λέμε ότι η θεωρία επαληθεύεται , αν επανειλημμένα αποτυγχάνουμε να αποδείξουμε ότι δεν είναι ορθή .

Το πρώτο βήμα για τον οικονομετρικό έλεγχο μιας θεωρίας είναι να εκφράσουμε την θεωρία μαθηματικά , δηλαδή να διατυπώσουμε το υπόδειγμα ή τη διατηρούμενη υπόθεση . Αν απ την αντιπαράθεση του υποδείγματος με τα πραγματικά δεδομένα προκύψει ότι το υπόδειγμα εξηγεί την πραγματική συμπεριφορά των οικονομικών μονάδων , καταναλωτών ή παραγωγών , αν με άλλα λόγια η θεωρία συμβιβάζεται με τα πραγματικά δεδομένα , τότε η θεωρία ή δε γίνεται δεκτή ή τροποποιείται με βάση τα καινούρια δεδομένα .

β) Την άσκηση οικονομικής πολιτικής . Οι αριθμητικές τιμές των παραμέτρων των διάφορων οικονομικών σχέσεων είναι απαραίτητες για την άσκηση οικονομικής πολιτικής . Αν και κατά πόσο θα ελαττωθεί η κατανάλωση , π.χ. , ενός αγαθού ή ποσά θα είναι τα φορολογικά έσοδα από την επιβολή ενός έμμεσου φόρου , θα εξαρτηθεί κατά κύριο λόγο από την ελαστικότητα της ζήτησής του . Επομένως , για να δοθούν απαντήσεις στα προηγούμενα ερωτήματα είναι απαραίτητη η γνώση της αριθμητικής τιμής της ελαστικότητας . Η χρησιμότητα της Οικονομετρίας για την άσκηση οικονομικής πολιτικής είναι συνεπώς και φανερή και σπουδαία .

γ) Την πρόβλεψη των μελλοντικών τιμών των οικονομικών μεταβλητών . Η σημασία που έχουν οι προβλέψεις για τη μελλοντική πορεία των διάφορων οικονομικών μεγεθών είναι φανερή . Αποτελούν τη βάση για ένα ορθολογικότερο προγραμματισμό και μια ορθολογικότερη λήψη αποφάσεων , είτε σε επίπεδο μικροοικονομικό (επιχείρηση) είτε σε επίπεδο μακροοικονομικό (κράτος) . Με άλλα λόγια , οι προβλέψεις είναι απαραίτητες για τον έλεγχο και την λήψη των αναγκαίων μέτρων από τους φορείς λήψης αποφάσεων που θα επηρεάσουν τις τιμές των διάφορων οικονομετρικών μεταβλητών . Για τους παραπάνω σκοπούς , με τη βοήθεια των διάφορων οικονομετρικών υποδειγμάτων , μπορούμε να έχουμε τις απαραίτητες εκτιμήσεις των μελλοντικών τιμών των διάφορων οικονομικών μεταβλητών , όπως το εισόδημα , η απασχόληση , η κατανάλωση , οι εισαγωγές κ.λπ.

Η οικονομετρική ερευνά ενός οικονομικού φαινομένου περιλαμβάνει ορισμένα βασικά στάδια , που θα περιγράψουμε με λίγα λόγια . Με βάση την οικονομική θεωρία , το πρώτο βήμα είναι η εξειδίκευση του υποδείγματος ή της διατηρούμενης υποθέσεως . Η εφαρμογή

των κατάλληλων οικονομετρικών μεθόδων για την εμπειρική εκτίμηση του υποδείγματος είναι το δεύτερο στάδιο . Ο έλεγχος του υποδείγματος και η αξιολόγηση της ικανότητας του για προβλέψεις είναι οι τελευταίες φάσεις στην οικονομετρική ανάλυση ενός φαινομένου .

α) *Εξειδίκευση του υποδείγματος* . Η πιο σημαντική αλλά και η πιο δύσκολη φάση στην οικονομετρική ανάλυση ενός οικονομικού φαινομένου είναι η εξειδίκευση του υποδείγματος . Η εξειδίκευση του υποδείγματος αναφέρεται στον καθορισμό των διάφορων μεταβλητών που θα περιληφθούν στο υπόδειγμα , στο διαχωρισμό τους σε ενδογενείς και εξωγενείς καθώς και στη μαθηματική διατύπωση του υποδείγματος . Το πρόβλημα με την εξειδίκευση του υποδείγματος είναι ότι δεν υπάρχουν αναμφίβολα κριτήρια ή κανόνες για την επιλογή του πιο κατάλληλου υποδείγματος για τη μελέτη ενός φαινομένου .

Η Οικονομετρία προϋποθέτει την ύπαρξη οικονομικής θεωρίας , αλλά η εξειδίκευση του υποδείγματος που θα χρησιμοποιηθεί για την μελέτη ενός συγκεκριμένου προβλήματος δεν βασίζεται αποκλειστικά και μόνο στην οικονομική θεωρία . Σύμφωνα με τη << θεωρητική >> άποψη (theory only approach) , η επιλογή του υποδείγματος πρέπει να βασίζεται αποκλειστικά στην οικονομική θεωρία , ενώ κατά την << εμπειρική >> άποψη (facts only approach) δεν είναι ανάγκη να προϋπάρχει θεωρία . Από τα πραγματικά δεδομένα θα προκύψει το υπόδειγμα . Και οι δυο όμως απόψεις είναι ακραίες . Η οικονομική θεωρία μπορεί να υποδείξει ποιες μεταβλητές οπωσδήποτε σημαντικές και πρέπει να συμπεριληφθούν στο υπόδειγμα , ποιες άλλες είναι ίσως σχετικές , αλλά συνήθως δεν καθορίζει τη μαθηματική μορφή που συνδέει τις μεταβλητές . Επιπλέον , πολλές μεταβλητές της οικονομικής θεωρίας μπορεί να μη μπορούν να μετρηθούν ή να παρατηρηθούν . Στην πράξη , ο ερευνητής συνήθως δοκιμάζει διάφορες μορφές και με βάση ορισμένα κριτήρια , που θα εξετάσουμε αργότερα , επιλέγει την πιο κατάλληλη . Με αλλά λόγια , η επιλογή των μεταβλητών και η επιλογή της μαθηματικής μορφής της συναρτησιακής σχέσεως που συνδέει τις μεταβλητές είναι συνδυασμός των πληροφοριών από την οικονομική θεωρία και από τα πραγματικά δεδομένα .

Ας υποθέσουμε ότι θέλουμε να μελετήσουμε τη ζήτηση ενός συγκεκριμένου αγαθού . Από την οικονομική θεωρία γνωρίζουμε ότι οι προτιμήσεις των καταναλωτών (T) , η τιμή του αγαθού (P) , οι τιμές των υποκατάστατων αγαθών (Ps) , οι τιμές των συμπληρωματικών αγαθών (Pc) και το εισόδημα (Y) είναι οι πιο σημαντικοί παράγοντες που προσδιορίζουν τη ζητούμενη ποσότητα (Q) ενός αγαθού . Στη

γενική της μορφή , κατά τα γνωστά , η συνάρτηση της ζήτησεως μπορεί να γράφει ως εξής :

$$Q = f(T,P,Ps,Pc,Y)$$

Φυσικά μπορεί να υπάρχουν και άλλοι παράγοντες οι οποίοι επηρεάζουν τη ζήτηση ενός αγαθού , όπως το εισόδημα της προηγούμενης περιόδου , η διανομή του εισοδήματος κ.λπ. Τα αποτελέσματα από προηγούμενες μελέτες καθώς και προσθετές πληροφορίες για τα ιδιαίτερα χαρακτηριστικά του εξεταζόμενου αγαθού πρέπει και συνήθως λαμβάνονται υπόψη από τον ερευνητή για τον καθορισμό του είδους και του αριθμού των ανεξάρτητων μεταβλητών .

Γενικά , όπως αναφέραμε παραπάνω , η οικονομική θεωρία δεν καθορίζει την ακριβή μαθηματική μορφή των οικονομικών σχέσεων . Αν υποθέσουμε ότι η συνάρτηση της ζήτησεως είναι γραμμική , μπορούμε να γράψουμε :

$$Q = \beta_0 + \beta_1 P + \beta_2 P_s + \beta_3 P_c + \beta_4 Y + u$$

Η παραπάνω εξίσωση αποτελεί το υπόδειγμα που θα ερευνηθεί οικονομικά . Για την πλήρη εξειδίκευση του απαιτείται επιπλέον ο καθορισμός των προσδοκώμενων πρόσχημων των συντελεστών και ο καθορισμός του τυχαίου όρου .

Ως προς τα πρόσχημα των συντελεστών , μπορούμε να παρατηρήσουμε ότι, σύμφωνα με την οικονομική θεωρία , θα πρέπει να περιμένουμε αρνητική τιμή για το συντελεστή β_1 καθώς και για το συντελεστή β_3 αφού αναφέρεται σε συμπληρωματικό αγαθό . Η τιμή του συντελεστή β_4 , δηλαδή του εισοδήματος , θα πρέπει να είναι θετική , εκτός αν το αγαθό είναι *κατώτερο* (inferior) αγαθό , όποτε μπορεί να είναι αρνητική .

β) *Εκτίμηση του υποδείγματος* . Η εκτίμηση του υποδείγματος αναφέρεται στην εφαρμογή των κατάλληλων οικονομετρικών μεθόδων για την εκτίμηση των παραμέτρων (συντελεστών) του υποδείγματος . Τα ειδικά προβλήματα που αντιμετωπίζουμε για την εκτίμηση των παραμέτρων θα εξετασθούν στα κεφάλαια που ακολουθούν .

Ως προς το είδος των στατιστικών παρατηρήσεων για τις μεταβλητές του υποδείγματος , παρατηρούμε πως βασικά έχουμε δυο κατηγορίες . Τις *χρονολογικές σειρές* και τα *διαστρωματικά στοιχεία* . Οι χρονολογικές σειρές αναφέρονται σε διαχρονικές παρατηρήσεις ,

δηλαδή παρατηρήσεις για μια σειρά ετών , μηνών κ.λπ. Τα διαστρωματικά στοιχεία αναφέρονται σε παρατηρήσεις για ένα στρώμα οικονομικών μονάδων σε μια δεδομένη στιγμή (χρόνο) , όπως π.χ., το εισόδημα ή οι δαπάνες καταναλώσεως για ένα δείγμα οικογενειών σε ένα ορισμένο μήνα .

γ) Έλεγχος του υποδείγματος . Μετά την εκτιμήσει του υποδείγματος , το επόμενο στάδιο στη διαδικασία της οικονομετρικής αναλύσεως αναφέρεται στην αξιολόγηση και τον έλεγχο των αποτελεσμάτων της εκτιμήσεως . Για το σκοπό αυτό , χρησιμοποιούνται κριτήρια που μπορούν να διακριθούν σε οικονομικά , στατιστικά και οικονομετρικά .

Τα οικονομικά κριτήρια είναι α ρισιό κριτήρια από την οικονομική θεωρία και αναφέρονται κυρίως στα πρόσημα και τα όρια των συντελεστών .

Όπως αναφέραμε προηγουμένως , η οικονομική θεωρία καθορίζει τα πρόσημα και πολλές φορές τα όρια των τιμών των συντελεστών , που δεν είναι τίποτε άλλο παρά ελαστικότητας , πολλαπλασιαστές , οριακές ροπές κ.λπ. Ο συντελεστής π.χ., της τιμής στη συνάρτηση ζήτησεως ενός αγαθού θα πρέπει να είναι μικρότερη από την μονάδα και με μεγαλύτερη από το μηδέν . Αν τα πρόσημα ή και τα μεγέθη των συντελεστών που προκύπτουν από την εκτιμήσει δεν είναι σύμφωνα με την οικονομική θεωρία , δεν θα πρέπει να γίνουν δεκτές οι εκτιμήσεις , εκτός αν υπάρχουν βάσιμοι λόγοι να αμφισβητηθεί η ορθότητα της οικονομικής θεωρίας στο συγκεκριμένο πρόβλημα .

Ανεξάρτητα από το στατιστικό και οικονομικό έλεγχο , ένα οικονομετρικό υπόδειγμα συχνά αξιολογείται και ως προς την ικανότητα του για προβλέψεις των μελλοντικών τιμών των ενδογενών μεταβλητών . Με αλλά λόγια , ελέγχεται η <<επίδοση>> (performance) του υποδείγματος σε περίοδο έξω από το δείγμα . Δυστυχώς ένα υπόδειγμα που κρίνεται καλό σύμφωνα με τα προηγούμενα κριτήρια , δεν είναι απαραίτητο να δίνει ακριβείς προβλέψεις . Ενώ , δηλαδή , η << επίδοση >> του υποδείγματος για την περίοδο του δείγματος μπορεί να είναι ικανοποιητικοί , οι προβλέψεις μπορεί να διαφέρουν σημαντικά από τις πραγματικές τιμές .

Τα ανωτέρω συνοψίζονται , παραστατικά, χωρίς περαιτέρω σχόλια .

Η διαδικασία της Οικονομετρικής Αναλύσεως

Οικονομική Θεωρία



Οικονομικό Υπόδειγμα



Οικονομετρικό Υπόδειγμα



Δεδομένα → Εκτίμηση του υποδείματος ← Οικονομετρικές μέθοδοι



Έλεγχος Υποδείματος



Χρησιμοποίηση του Υποδείματος



**Έλεγχος
Θεωρίας
(Υποθέσεων)**



**Αξιολόγηση
Οικονομικής
Πολιτικής**

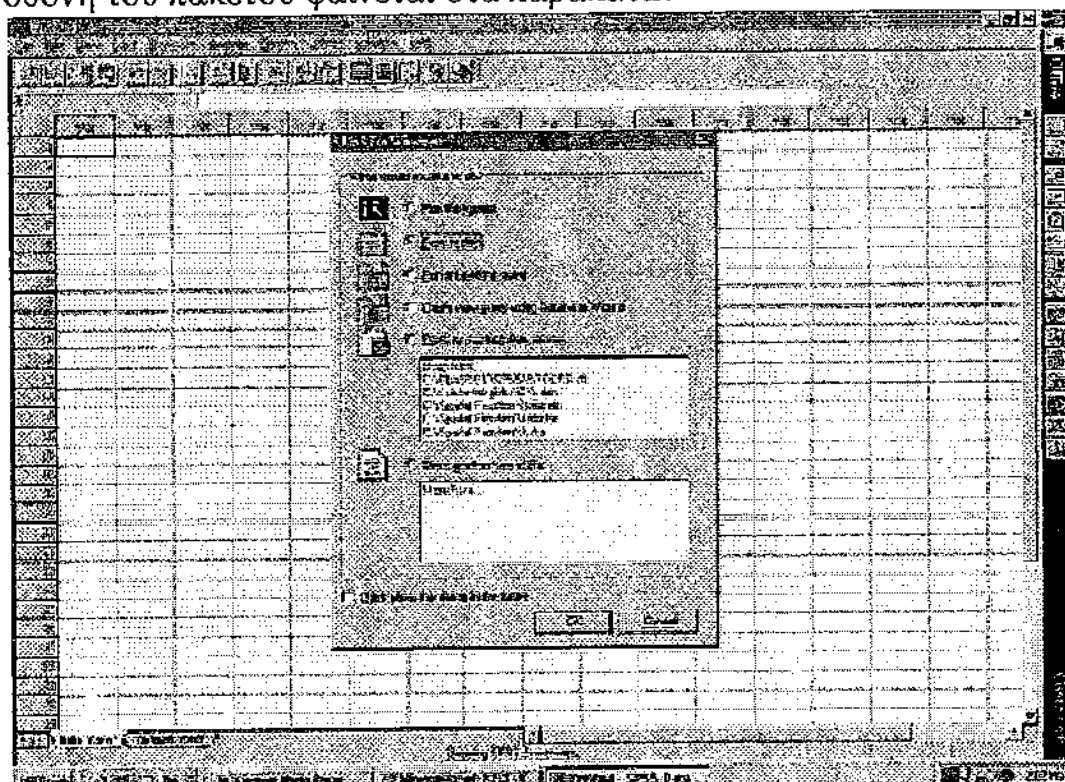


Προβλέψεις

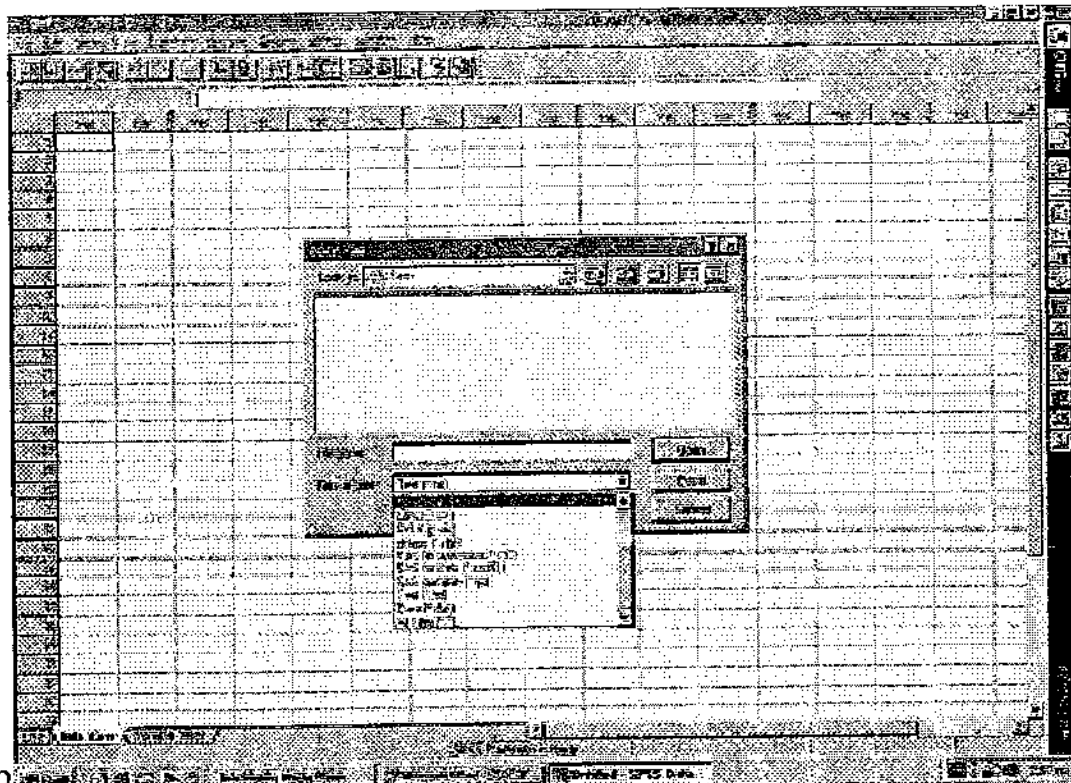
ΜΕΡΟΣ ΤΡΙΤΟ

ΕΙΣΑΓΩΓΗ ΣΤΟ SPSS

Το SPSS είναι ένα πακέτο για γενικές στατιστικές αναλύσεις το οποίο από την σκοπιά του οικονομολόγου προσφέρει δυνατότητες που είναι ενδιάμεσα στο Excel και στο Eviews. Θα αρχίσουμε διαβάζοντας το αρχείο stocks.xls με τα χρηματιστηριακά στοιχεία και στην συνέχεια θα κάνουμε γραφικά και στατιστικές αναλύσεις. Η αρχική οθόνη του πακέτου φαίνεται στα παρακάτω.

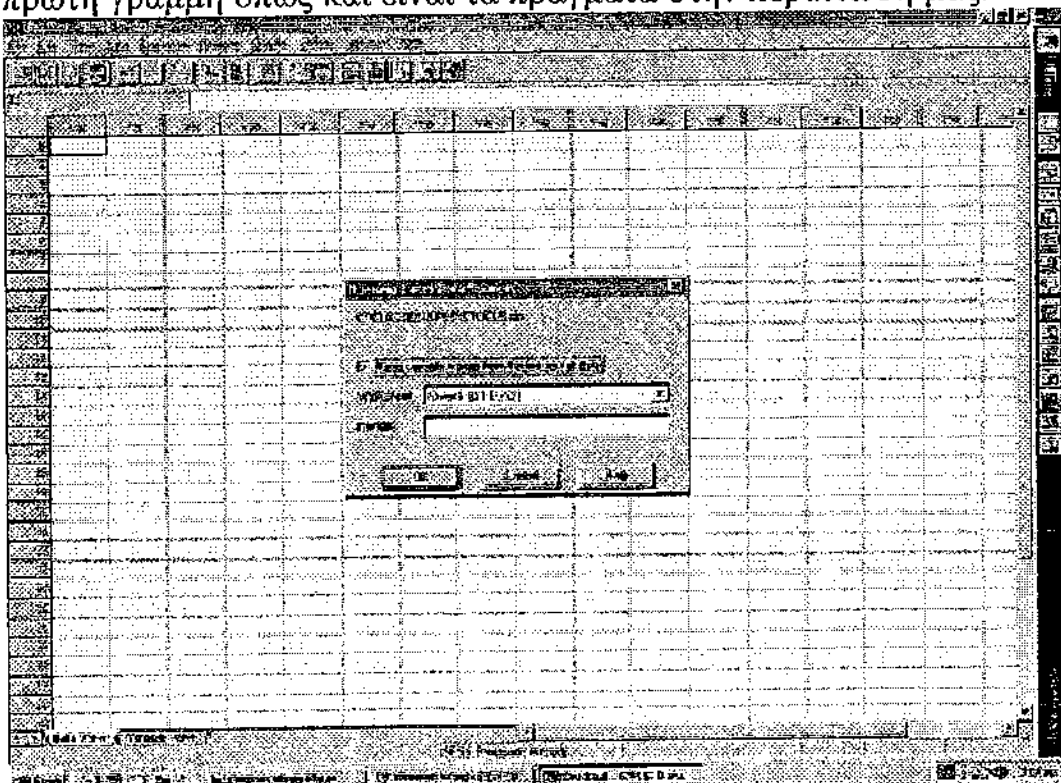


- Θα επιλέξουμε “Type in data” ώστε να εισάγουμε νέα στοιχεία και θα επιλέξουμε OK.
Στην συνέχεια επιλέγουμε File/Read Text Data ώστε να διαβάσουμε τα στοιχεία και έχουμε την εμφάνιση της ακόλουθης οθόνης.



2

Στην οθόνη αυτή επιλέγουμε να διαβάσουμε αρχεία XLS, δηλαδή αρχεία του Excel, βρίσκουμε το αρχείο που θέλουμε να διαβάσουμε και εμφανίζεται η επόμενη οθόνη στην οποία μας ζητείται να προσδιορίσουμε τα πεδία στα οποία είναι οι μεταβλητές και αν τα ονόματά τους είναι στην πρώτη γραμμή όπως και είναι τα πράγματα στην περίπτωση μας.

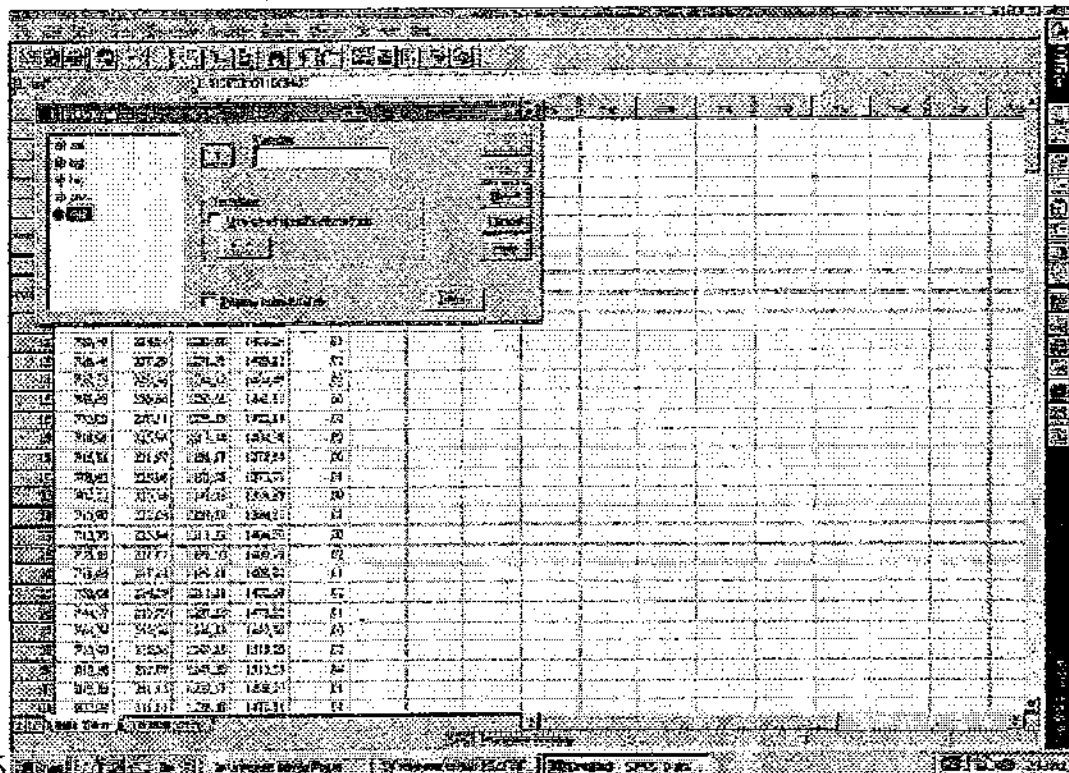


Με την επιλογή του OK τα στοιχεία φαίνονται στην οθόνη μας.

στον χώρο που διατίθεται για τον σκοπό αυτό. Στην συνέχεια η νέα σειρά εμφανίζεται στο φύλλο εργασίας.

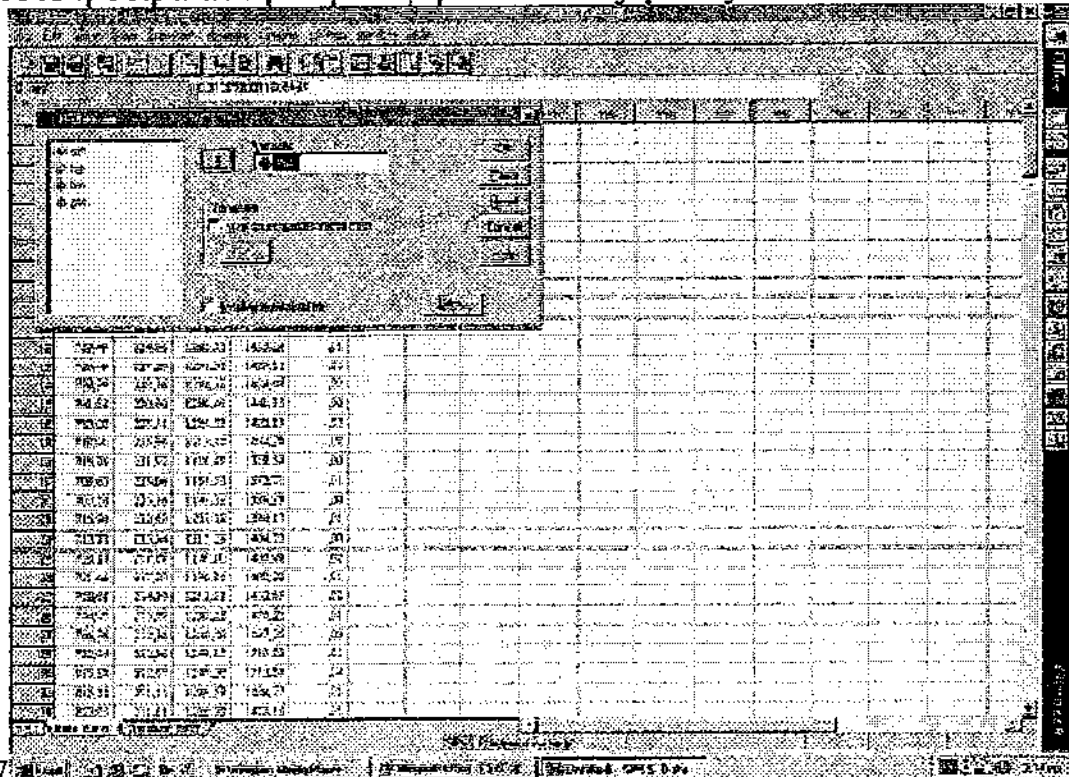
	1	2	3	4	5	6	7	8	9	10
1	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
2	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
3	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
4	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
5	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
6	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
7	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
8	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
9	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
10	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
11	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
12	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
13	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
14	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
15	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
16	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
17	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
18	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
19	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
20	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Μπορούμε να κάνουμε το ίδιο και για τις υπόλοιπες σειρές αν το επιθυμούμε αλλά στις σημειώσεις αυτές δεν θα ασχοληθούμε με άλλες σειρές εκτός της GEN. Για να κάνουμε ένα ιστόγραμμα της σειράς επιλέγουμε Graph/Histogram και εμφανίζεται η ακόλουθη οθόνη



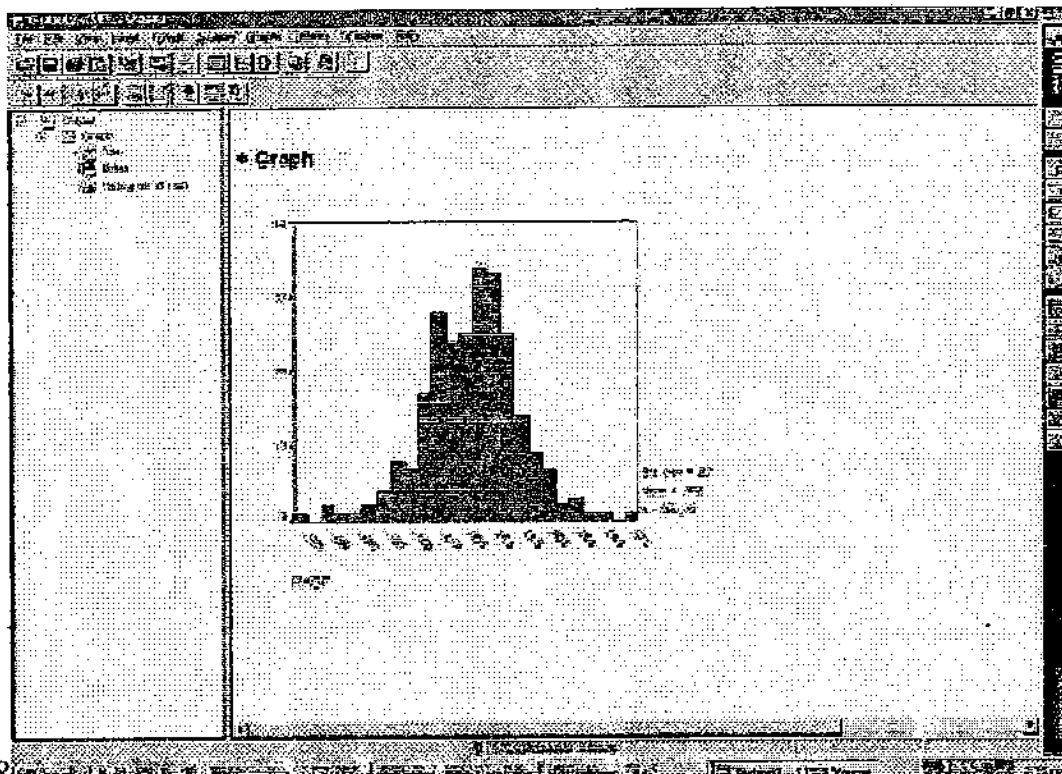
6

Στην οθόνη αυτή πρέπει να ορίσουμε την σειρά πράγμα που μπορούμε να κάνουμε αν επιλέξουμε την σειρά που θέλουμε και πατήσουμε το βελάκι ώστε η σειρά αυτή να μεταφερθεί στο δεξιό μέλος.



7

Το αποτέλεσμα φαίνεται στην παρακάτω οθόνη.

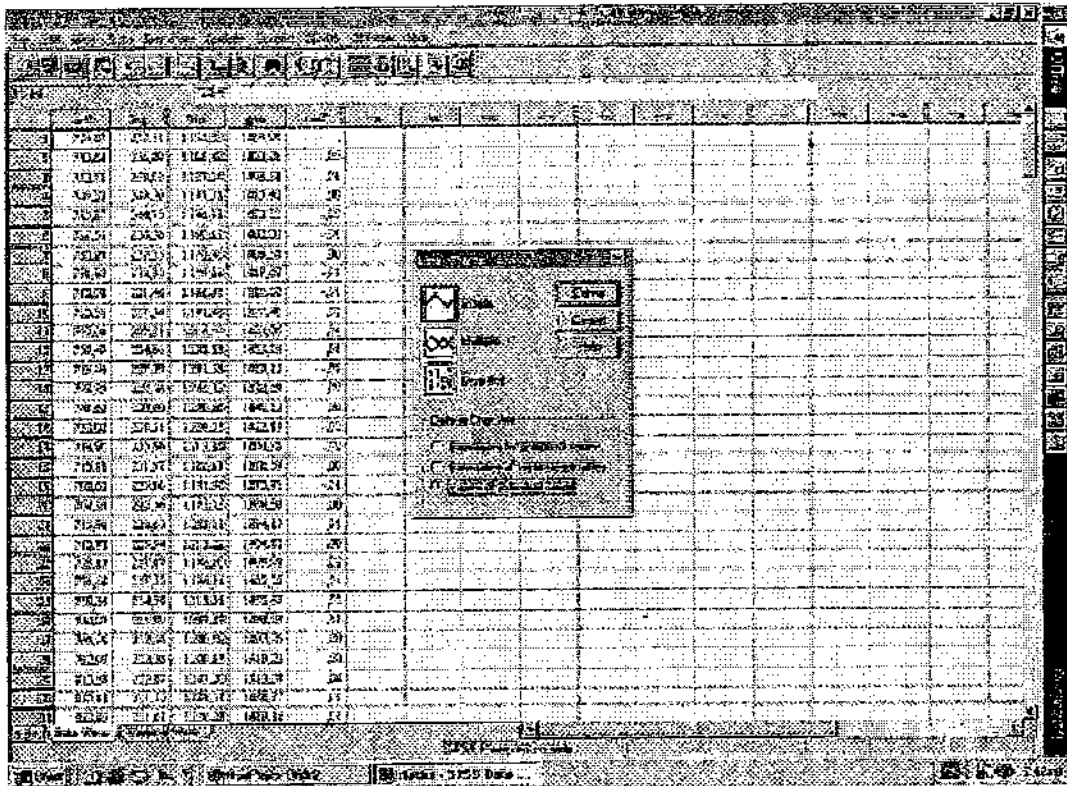


8

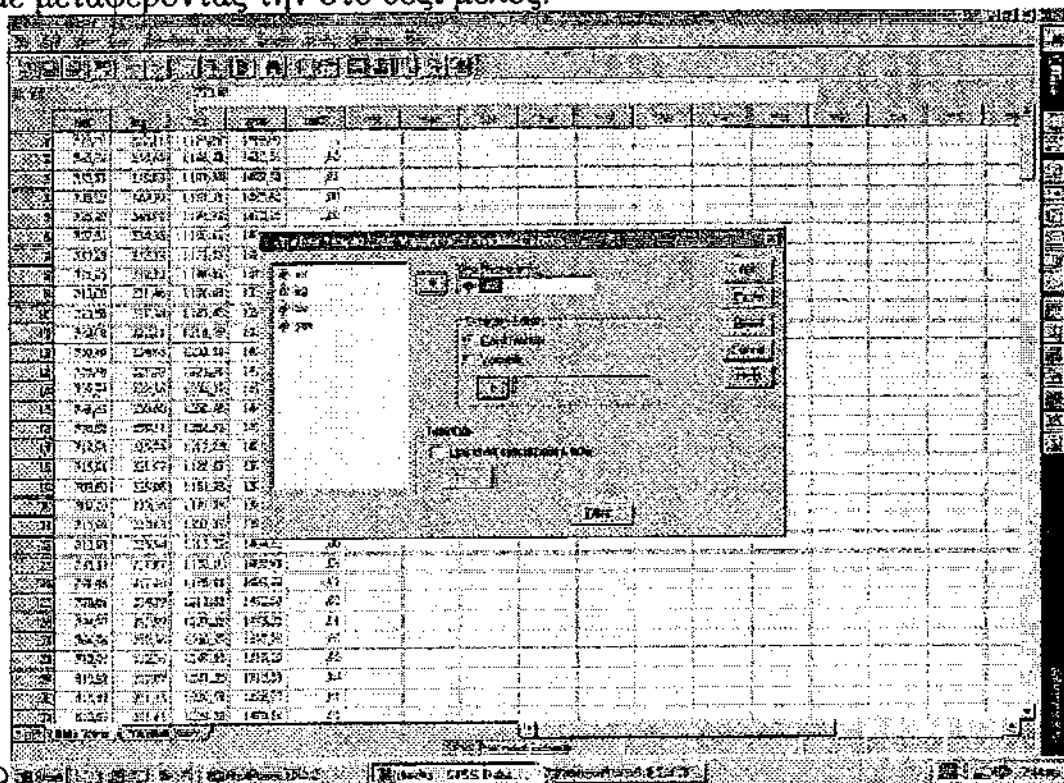
Για να διώξουμε την οθόνη αυτή απλά την ακυρώνουμε και επιλέγουμε να μην αποθηκευθεί το αποτέλεσμα.

Για να αποθηκεύσουμε το αρχείο με τα στοιχεία μας επιλέγουμε File/Save, στην συνέχεια διαλέγουμε σε ποιον φάκελο θέλουμε να γίνει η αποθήκευση, συμφωνούμε η αποθήκευση να γίνει με την μορφή αρχείου του SPSS και επιλέγουμε το όνομα stocks.

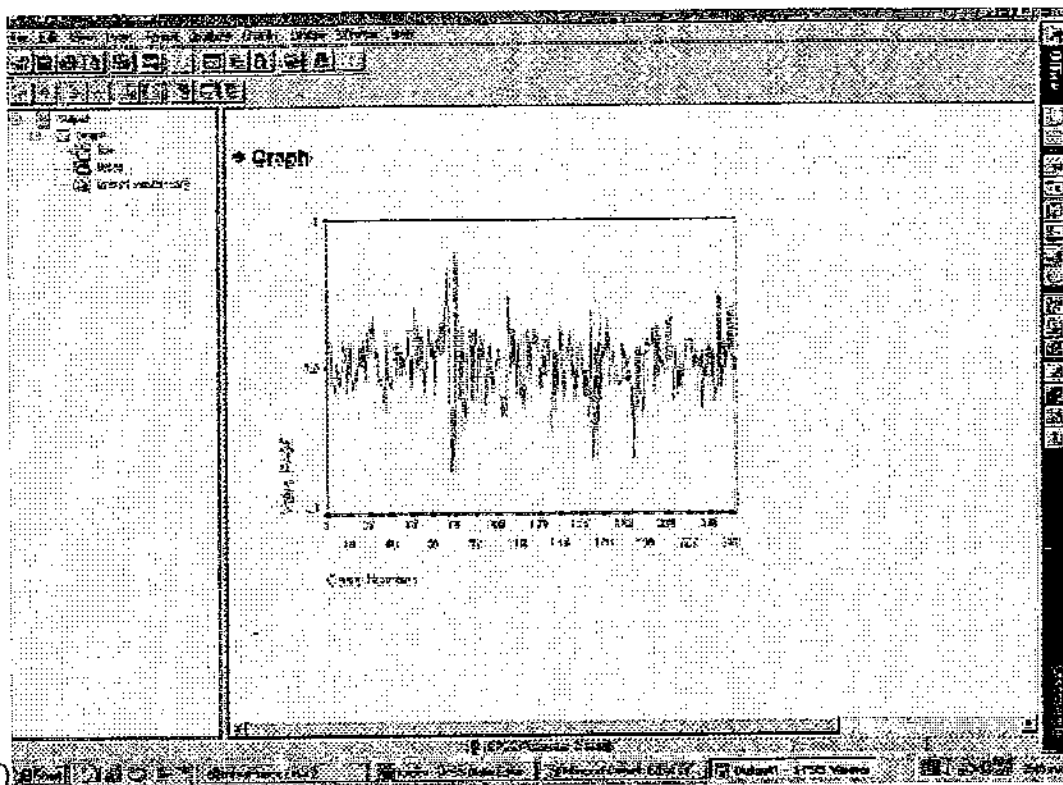
Για να κάνουμε ένα απλό διάγραμμα των αποδόσεων επιλέγουμε Graph/Line και στην οθόνη που εμφανίζεται τσεκάρουμε την επιλογή Values of individual series.



Στην συνέχεια επιλέγουμε Define και μαρκάρουμε την σειρά που θέλουμε μεταφέροντάς την στο δεξί μέλος.



Το αποτέλεσμα φαίνεται στην επόμενη οθόνη.

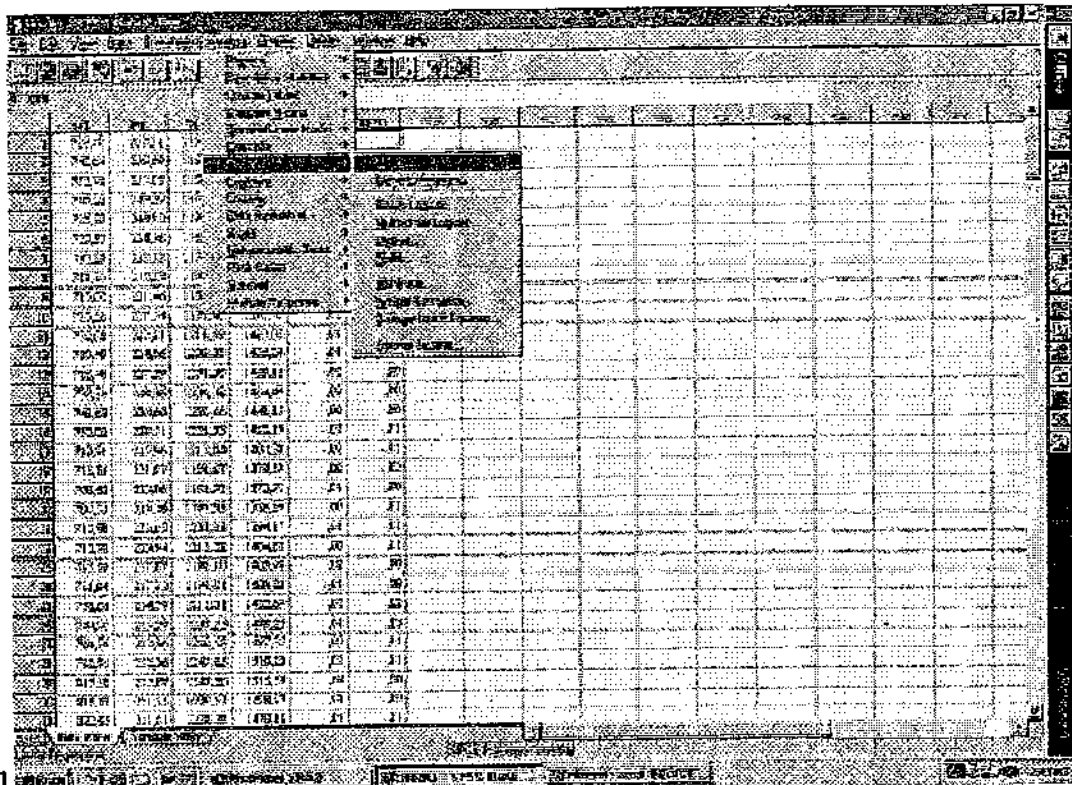


10 Το διάγραμμα αυτό μπορούμε να μεταφέρουμε σε οποιαδήποτε εφαρμογή των Windows όπως πχ το Word. Στην συνέχεια κατασκευάζουμε τις αποδόσεις του γενικού δείκτη GEN που βάζουμε στην νέα μεταβλητή RGEN.

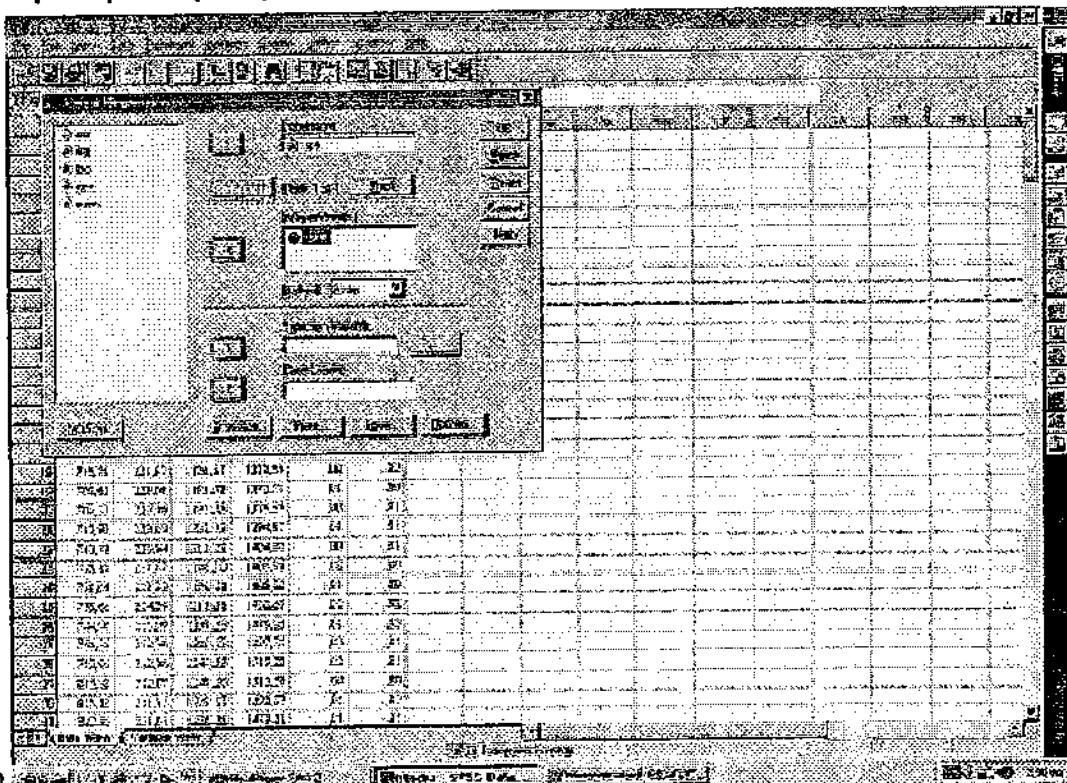
The screenshot shows a software window with a menu bar and a toolbar. A dialog box is open in the foreground, partially obscuring a data table. The table has several columns of numerical data. The dialog box contains various options and buttons, including a "OK" button at the bottom. The table data is as follows:

Case	GEN	RGEN
1	1000	1000
2	1000	1000
3	1000	1000
4	1000	1000
5	1000	1000
6	1000	1000
7	1000	1000
8	1000	1000
9	1000	1000
10	1000	1000
11	1000	1000
12	1000	1000
13	1000	1000
14	1000	1000
15	1000	1000
16	1000	1000
17	1000	1000
18	1000	1000
19	1000	1000
20	1000	1000
21	1000	1000
22	1000	1000
23	1000	1000
24	1000	1000
25	1000	1000
26	1000	1000
27	1000	1000
28	1000	1000
29	1000	1000
30	1000	1000
31	1000	1000
32	1000	1000
33	1000	1000
34	1000	1000
35	1000	1000
36	1000	1000
37	1000	1000
38	1000	1000
39	1000	1000
40	1000	1000
41	1000	1000
42	1000	1000
43	1000	1000
44	1000	1000
45	1000	1000
46	1000	1000
47	1000	1000
48	1000	1000
49	1000	1000
50	1000	1000

Για να εκτιμήσουμε το υπόδειγμα της αγοράς με παλινδρόμηση επιλέγουμε Analyze/Regression/Linear



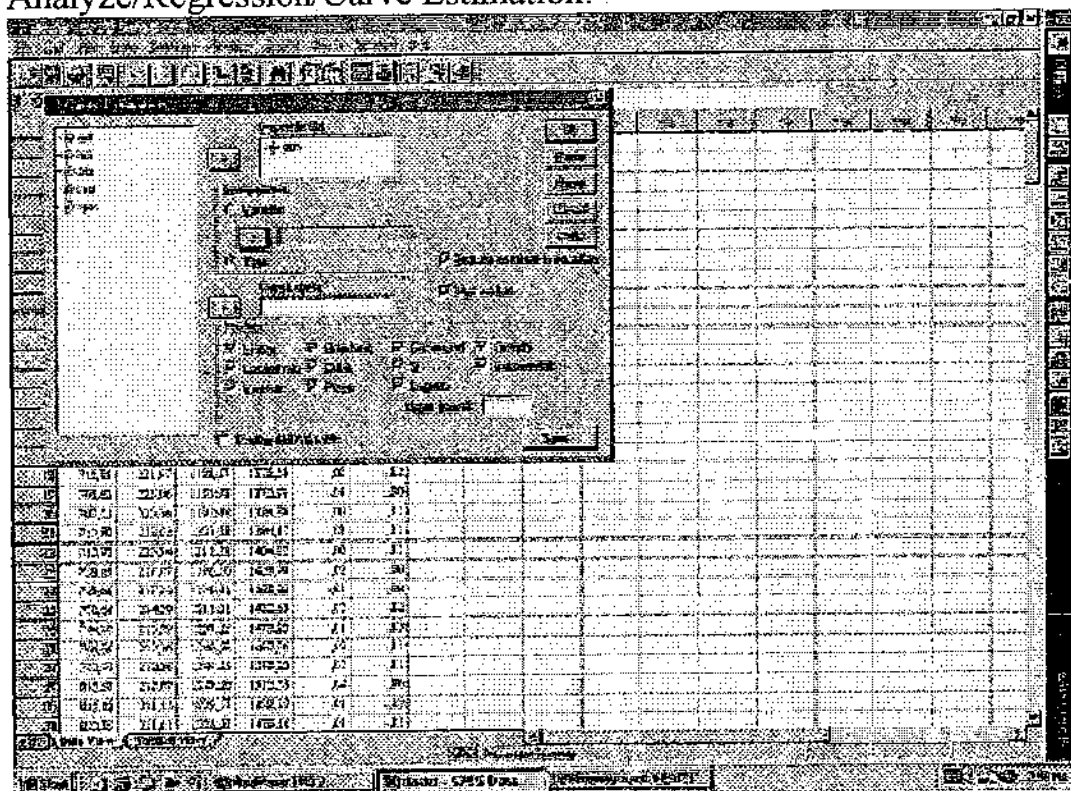
11 Στην συνέχεια επιλέγουμε την εξαρτημένη και ερμηνευτικές μεταβλητές με την γνωστή διαδικασία μεταφοράς στο δεξί μέλος με το βελάκι και έχουμε την επόμενη οθόνη.



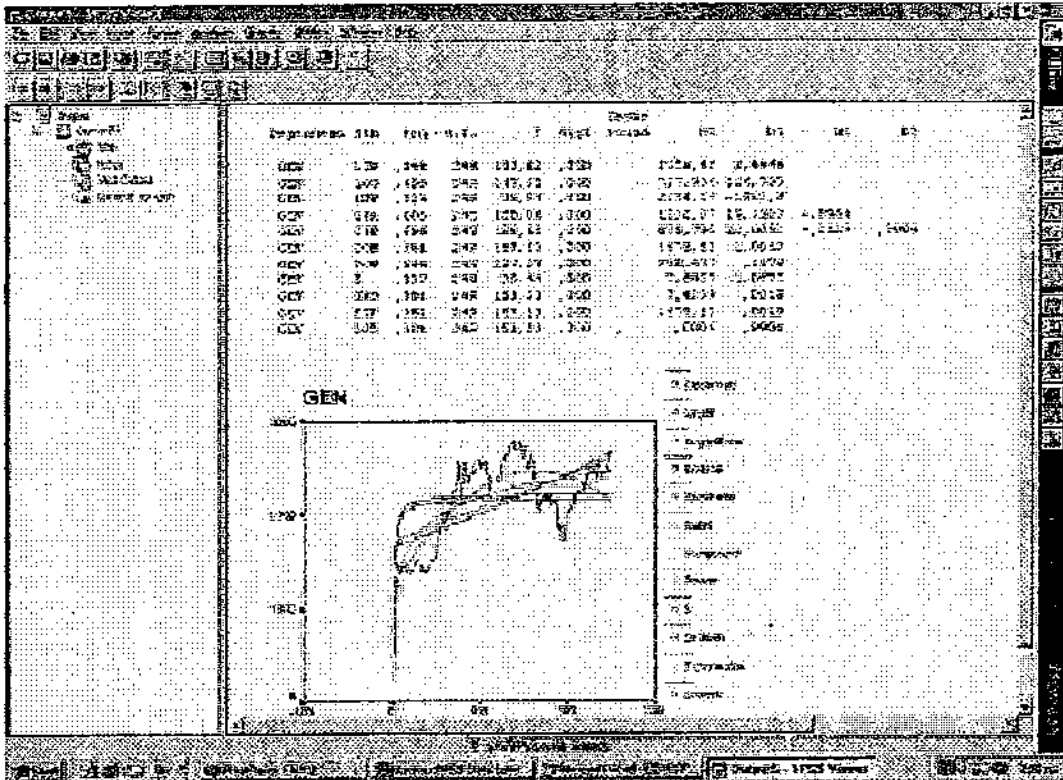
12 Τα αποτελέσματα που δίνει το πακέτο φαίνονται στην επόμενη οθόνη.

$$y_t = f(t) + u_t$$

για διάφορες εξειδικεύσεις της συνάρτησης f , όπου t είναι μια τάση (αλλά μπορεί να είναι και οποιαδήποτε άλλη μεταβλητή). Ο τρόπος με τον οποίο επιτυγχάνουμε την εκτίμηση είναι με τις εντολές Analyze/Regression/Curve Estimation.



1 Στην οθόνη που εμφανίζεται επιλέγουμε όλα τα υποδείγματα που θέλουμε να εκτιμήσουμε και τσεκάρουμε την επιλογή Time για να δηλώσουμε ότι έχουμε εκτίμηση τάσης. Τα αποτελέσματα είναι στην επόμενη οθόνη.

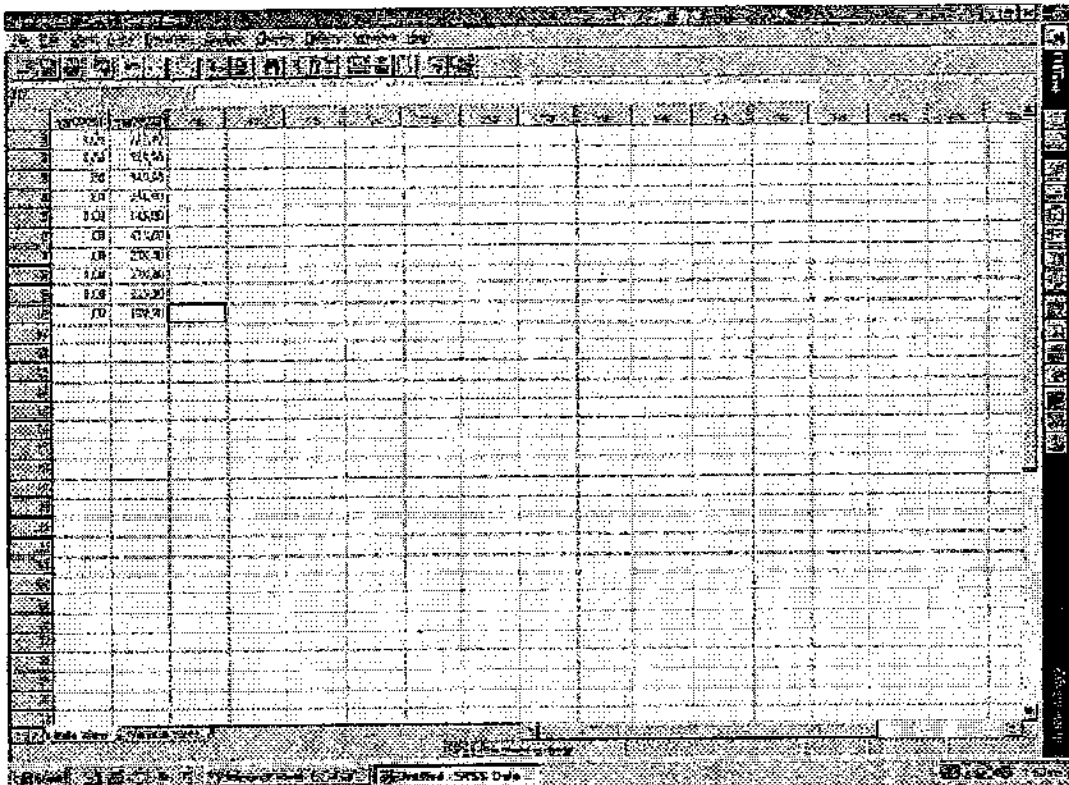


15

Το πακέτο μας δίνει εκτιμήσεις όλων των παραμέτρων κάθε τάσης και παριστάνει γραφικά τις πραγματικές τιμές της σειράς μαζί με τις προσαρμοσμένες τιμές που προκύπτουν από το κάθε υπόδειγμα. Φυσικά δεν θα είναι όλες κατάλληλες ή ενδεχομένως καμιά να μην είναι κατάλληλη αφού απλές τάσεις δεν μπορούν να περιγράψουν σειρές που προκύπτουν σαν το αποτέλεσμα πολύπλοκων οικονομικών διαδικασιών.

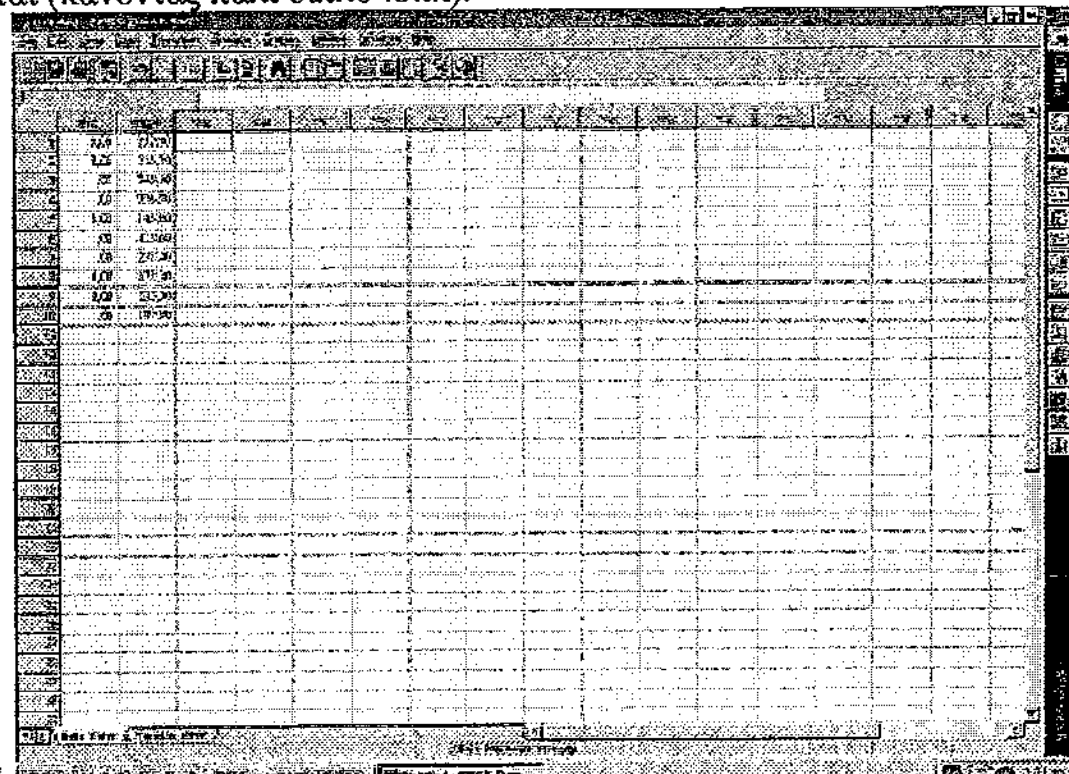
Απλοί έλεγχοι στατιστικών υποθέσεων

Με το SPSS μπορούμε να διεξάγουμε τους γνωστούς στατιστικούς ελέγχους. Ας υποθέσουμε ότι έχουμε τους μισθούς ενός δείγματος 10 ατόμων και το φύλο τους (1 αν είναι γυναίκα και 0 αν είναι άνδρας).



16

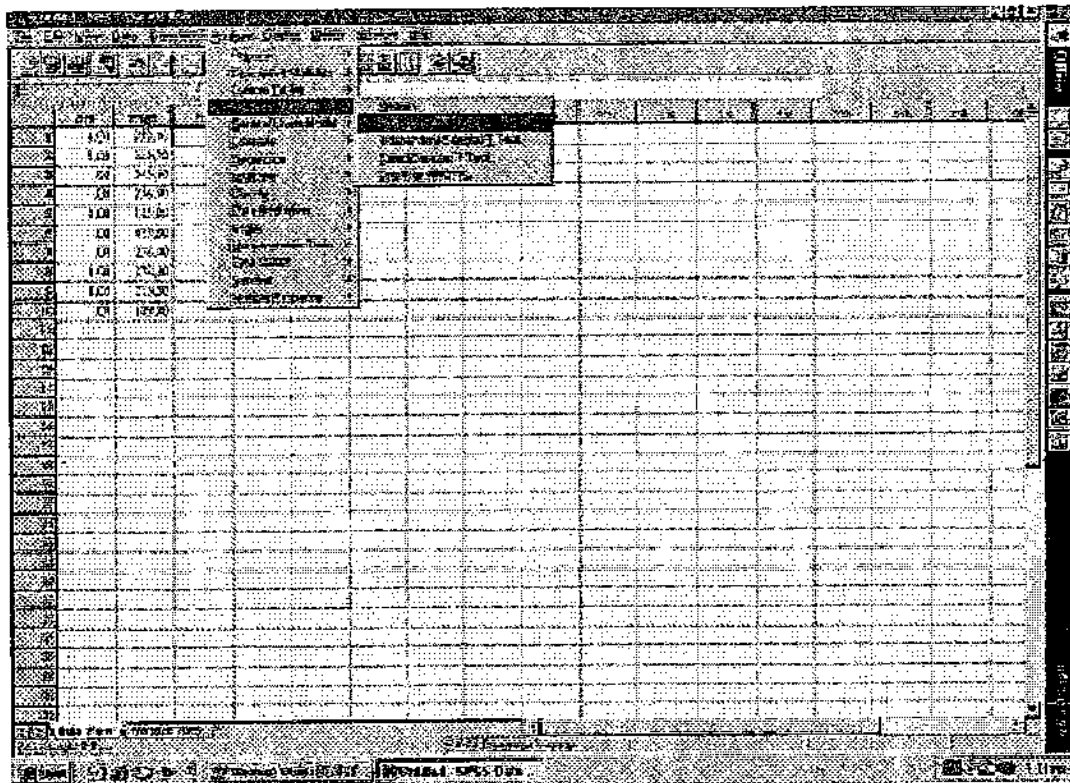
Σε πρώτη φάση θα θέλαμε να δώσουμε ονόματα στις μεταβλητές μας πράγμα που μπορούμε να επιτύχουμε αν κάνουμε διπλό κλικ στην σειρά και στην συνέχεια αλλάξουμε το όνομα στην επιλογή Name που εμφανίζεται (κάνοντας πάλι διπλό κλικ).



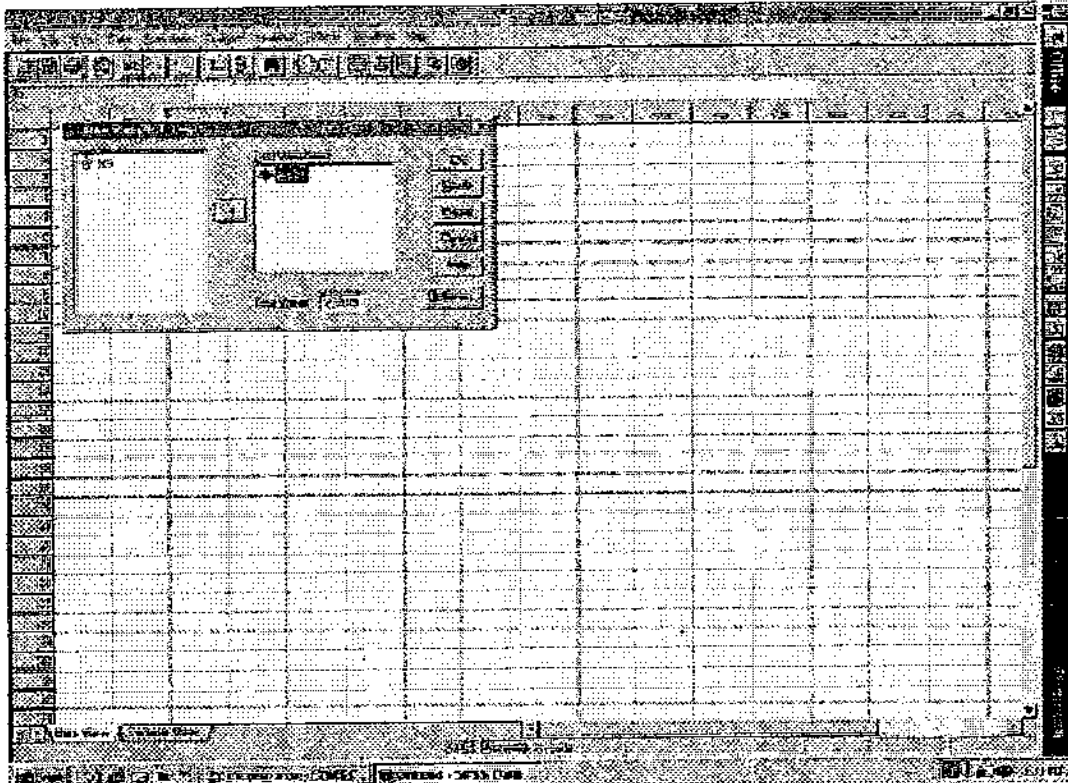
17

Με την επιλογή View/Data μπορούμε να επανέλθουμε στην αρχική οθόνη με τα στοιχεία.

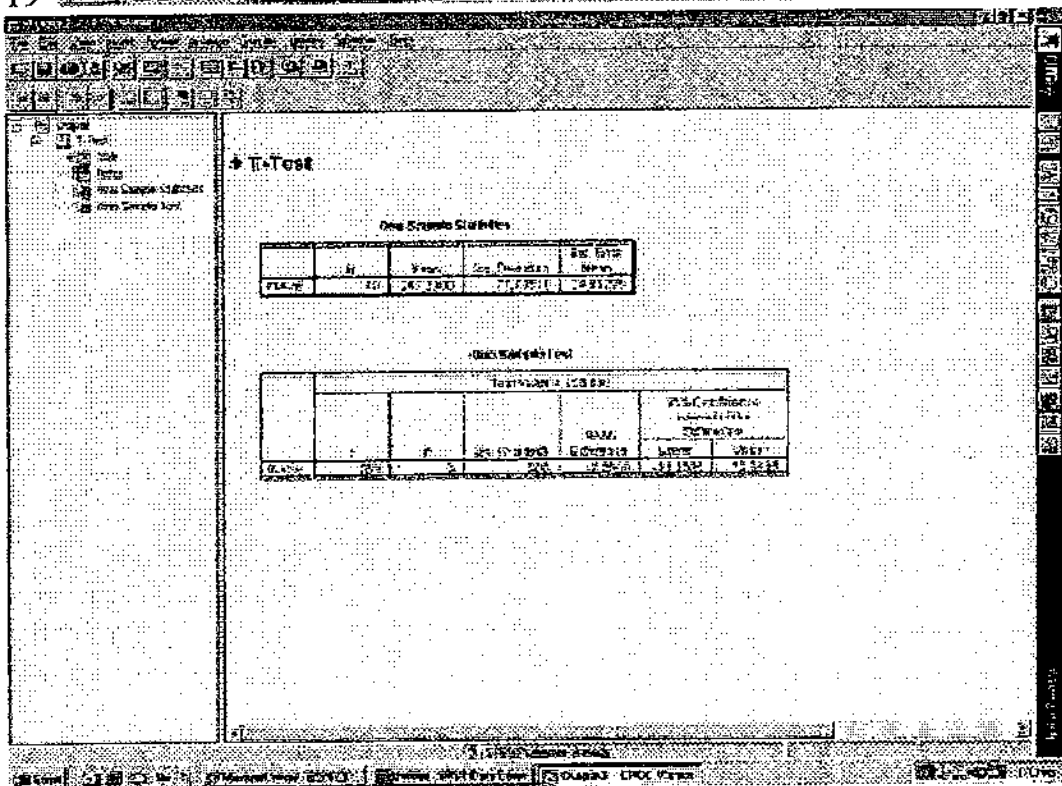
Για να ελέγξουμε ότι ο μέσος του πληθυσμού (με την υπόθεση ότι έχουμε τυχαίο δείγμα) είναι 250,000 επιλέγουμε Analyze/Compare means/One sample t-test και έχουμε την επόμενη οθόνη.



18 Στην συνέχεια επιλέγουμε την μεταβλητή που περιέχει το τυχαίο δείγμα και την τιμή του ελέγχου.



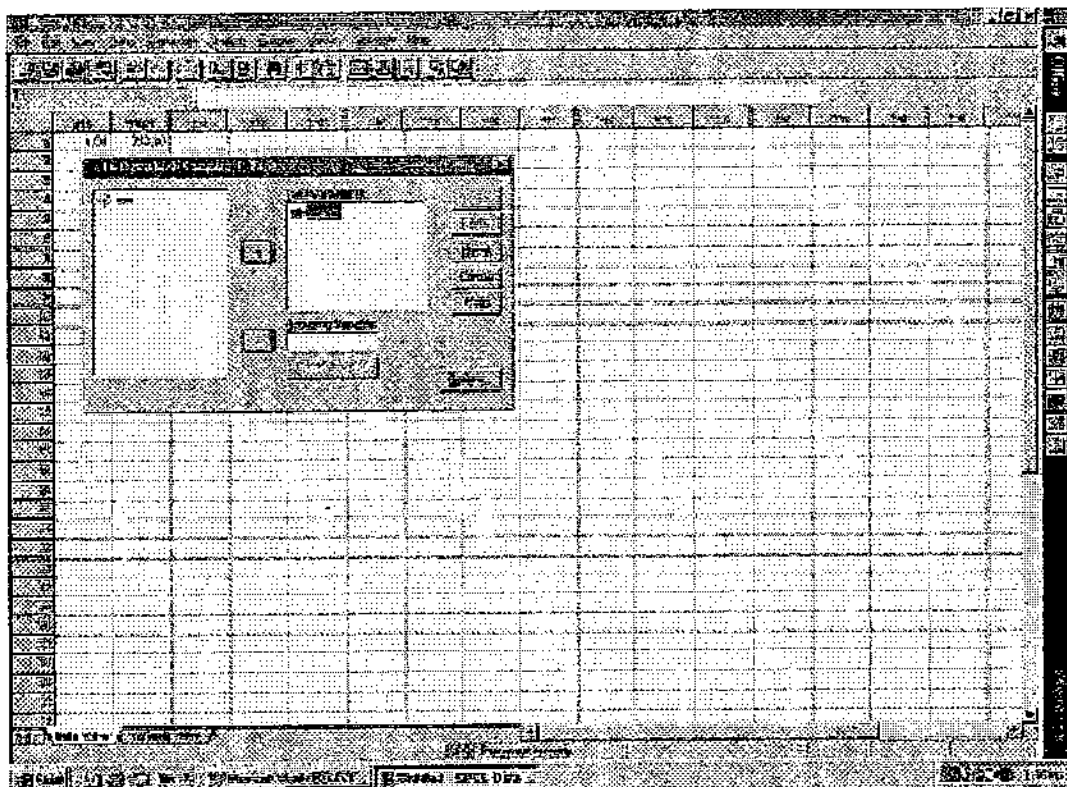
19



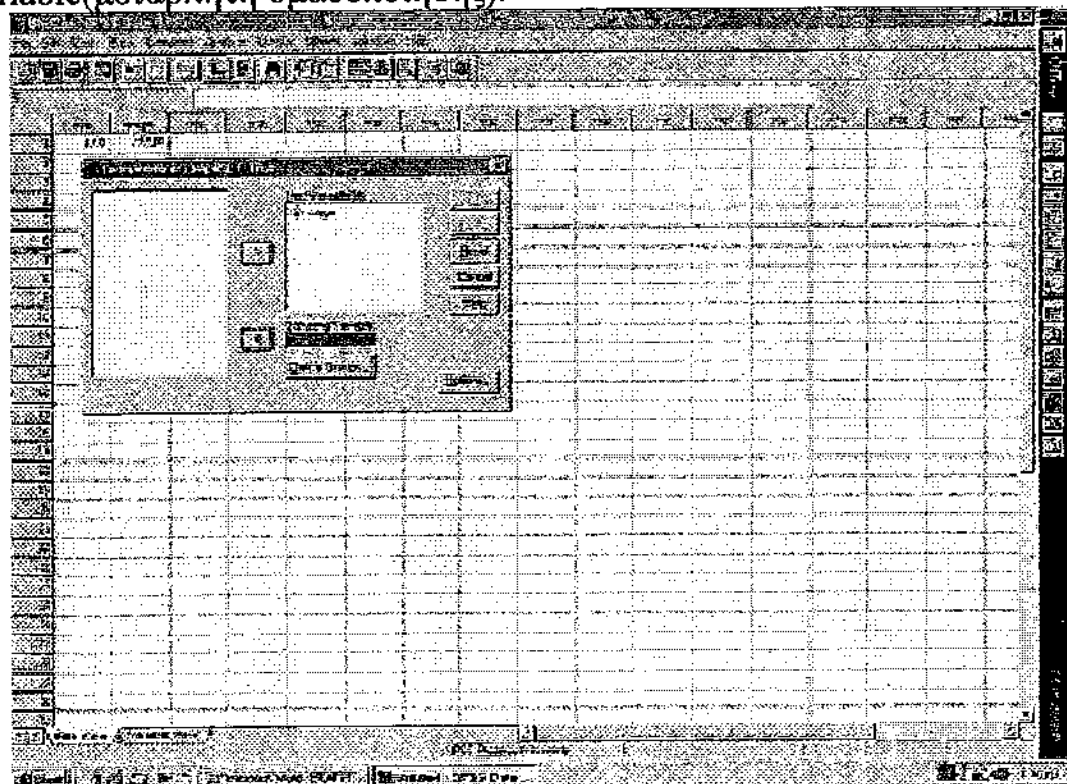
Όπως φαίνεται η $\bar{\eta}^2$ στατιστική έχει 9 βαθμούς ελευθερίας (df) και η τιμή της είναι 0,703 με πιθανότητα 0,500 οπότε δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση.

Για να ελέγξουμε την υπόθεση της ισότητας των μισθών ανδρών και γυναικών χρησιμοποιούμε την επιλογή

Analyze/Comparemeans/Independent samples t-test.
Καταρχήν σαν test variable επιλέγουμε την μεταβλητή wage.

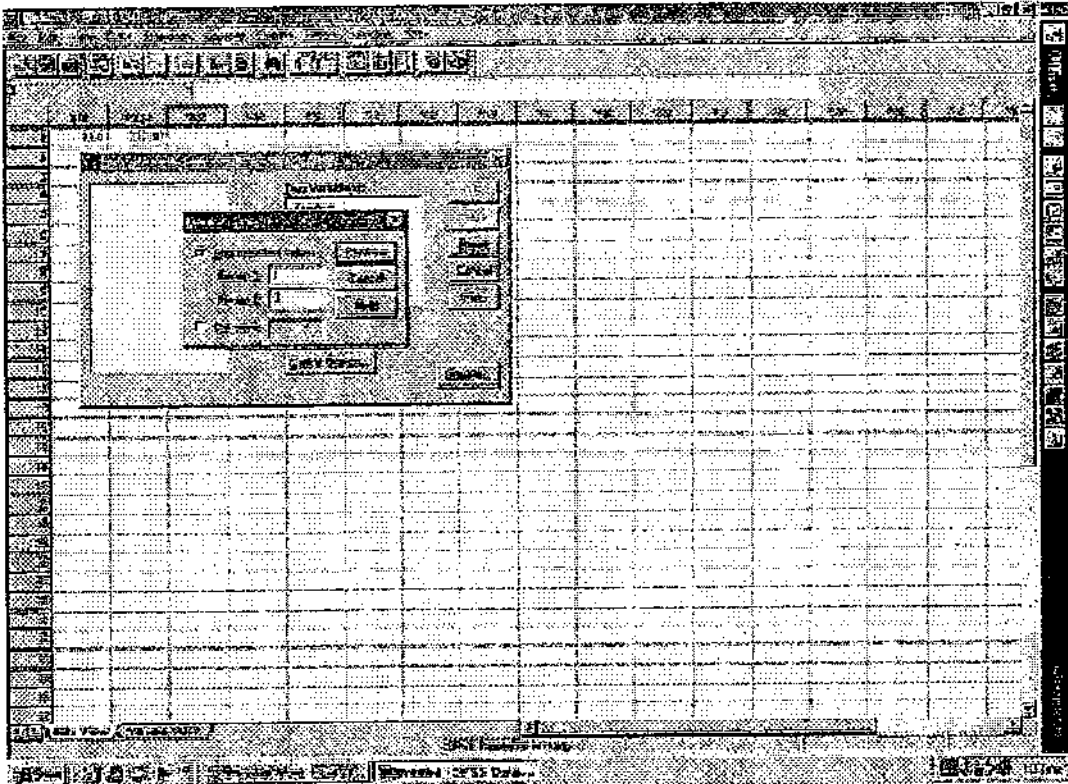


20 Στην συνέχεια επιλέγουμε την μεταβλητή sex σαν grouping variable(μεταβλητή ομαδοποίησης).

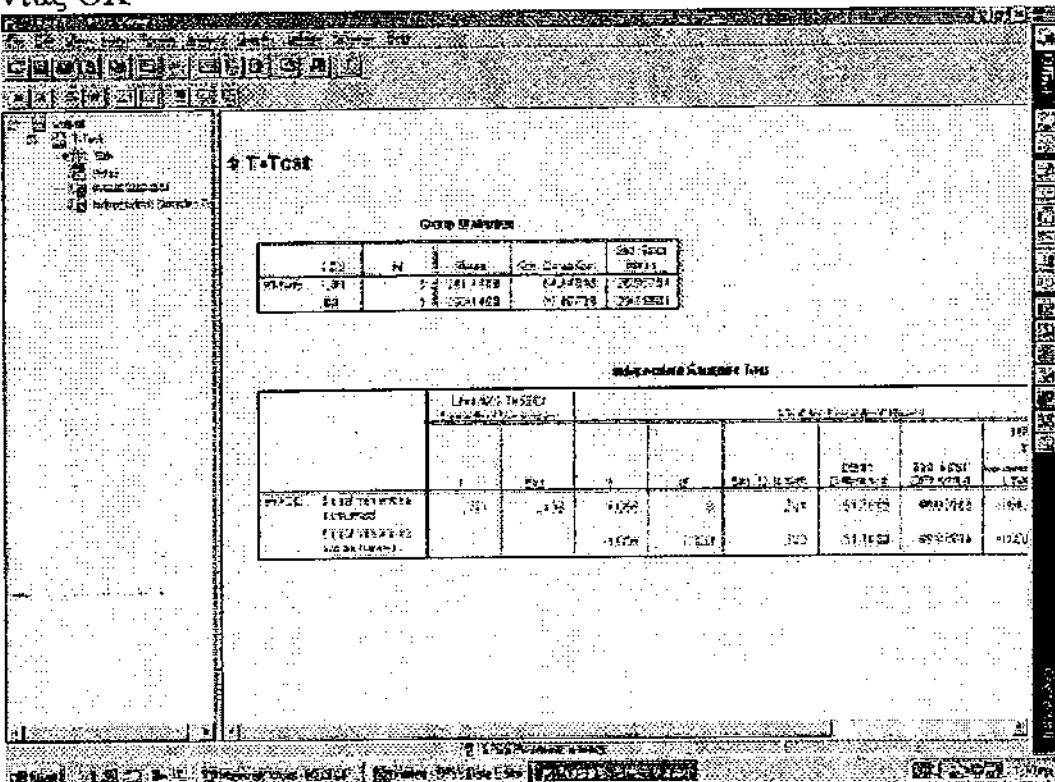


21 Στην συνέχεια πρέπει να ορίσουμε τις κατηγορίες με την επιλογή define

groups.



22 Με την επιλογή Continue έχουμε τα αποτελέσματα του ελέγχου επιλέγοντας OK



Στα αποτελέσματα αυτά η t στατιστική είναι $-1,059$ με την υπόθεση ανίσησων διακυμάνσεων και έχει πιθανότητα $0,323$ οπότε δεν μπορούμε να

απορρίψουμε την υπόθεση της ισότητας των μισθών. Επίσης έχουμε απευθείας και τον έλεγχο F ισότητας των διακυμάνσεων που είναι 0,771 με πιθανότητα 0,406 οπότε δεν μπορούμε να απορρίψουμε την υπόθεση της ισότητας των διακυμάνσεων. Θα παρατηρήσατε ότι οι βαθμοί ελευθερίας για τον t έλεγχο με άνισες διακυμάνσεις δεν είναι ακέραιος αριθμός (7,36). Αυτό οφείλεται στο ότι ο έλεγχος δυο μέσων με άνισες διακυμάνσεις είναι ένα δύσκολο στατιστικό πρόβλημα (γνωστό σαν πρόβλημα των Behrens-Fisher) που δεν έχει ακριβή λύση, δηλαδή η κατανομή της t στατιστικής δεν είναι γνωστή σε μικρά δείγματα οπότε χρησιμοποιείται σαν προσέγγιση η κατανομή Student- t με βαθμούς ελευθερίας που υπολογίζονται για να είναι καλή η προσέγγιση.

Τυχαίοι αριθμοί

Το SPSS διαθέτει μια μεγάλη επιλογή κατανομών από τις οποίες μπορούμε να έχουμε τυχαίους αριθμούς. Ας υποθέσουμε ότι έχουμε ανοίξει το αρχείο stocks.xls και θέλουμε να δημιουργήσουμε μια σειρά U τυχαίων αριθμών από την κανονική κατανομή με μέσο $\mu = 1$ και τυπική απόκλιση 2. Για να παράγουμε μια ακολουθία τυχαίων αριθμών την οποία μπορούμε να έχουμε και κάποια επόμενη φορά πρέπει να σταθεροποιήσουμε τις παραμέτρους της γεννήτριας τυχαίων αριθμών με την εντολή Transform/Random number seed και στην συνέχεια δίνουμε έναν ακέραιο, πχ 12.

Για να παράγουμε τους τυχαίους αριθμούς, επιλέγουμε transform/Compute και στην συνέχεια επιλέγουμε την συνάρτηση RV.NORMAL. Αρκετές άλλες συναρτήσεις είναι διαθέσιμες στην κατηγορία RV, όπως πχ κατανομή γάμμα, βήτα, Cauchy κλπ. Στην συνέχεια επιλέγουμε τις παραμέτρους της κανονικής κατανομής και έχουμε την ακόλουθη οθόνη.

The screenshot shows the SPSS 'Analyze' menu path: Analyze > Nonparametric Tests > 1-Sample K-S... The dialog box is open, showing the 'Test Distribution' list with 'Normal' selected. The 'Data File' is 'SPSS.DAT'. The main window displays a data table with the following columns: 'Case #', 'X', 'Y', 'Z', 'W', 'V', 'U', 'T', 'S', 'R', 'Q', 'P', 'O', 'N', 'M', 'L', 'K', 'J', 'I', 'H', 'G', 'F', 'E', 'D', 'C', 'B', 'A'. The data rows contain numerical values for each variable.

Ένας έλεγχος που μπορεί να χρησιμοποιηθεί για να ελέγξουμε την υπόθεση ότι ένα δείγμα προέρχεται από μια ορισμένη κατανομή είναι ο έλεγχος Kolmogorov-Smirnov που είναι διαθέσιμος στο μενού Analyze/Nonparametric tests/1-sample K-S.

The screenshot shows the SPSS 'Analyze' menu path: Analyze > Nonparametric Tests > 1-Sample K-S... The dialog box is open, showing the 'Test Distribution' list with 'Normal' selected. The 'Data File' is 'SPSS.DAT'. The main window displays a data table with the following columns: 'Case #', 'X', 'Y', 'Z', 'W', 'V', 'U', 'T', 'S', 'R', 'Q', 'P', 'O', 'N', 'M', 'L', 'K', 'J', 'I', 'H', 'G', 'F', 'E', 'D', 'C', 'B', 'A'. The data rows contain numerical values for each variable.

24 Στην συνέχεια επιλέγουμε τις κατανομές που θέλουμε και οι οποίες μπορεί να είναι η κανονική, εκθετική, Poisson και ομοιόμορφη και έχουμε τα

αποτελέσματα του ελέγχου.

The screenshot shows a statistical software window with a data table. The table has several columns, likely representing different variables or parameters. The data is organized in rows, with some cells containing numerical values and others possibly representing categorical or missing data. The interface includes a menu bar at the top and a toolbar with various icons.

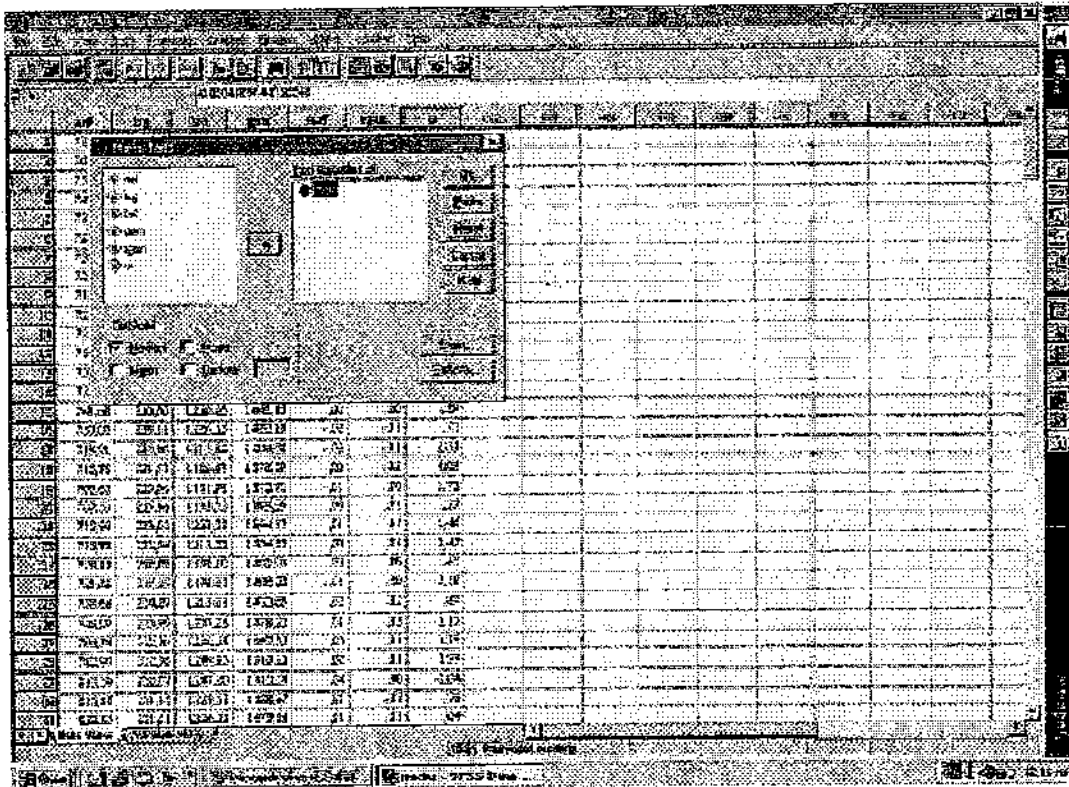
The screenshot shows a window titled "NPar Tests" with a table of test results. The table has two columns: the test name and the corresponding value. Below the table, there are some additional notes or instructions.

Test Name	Value
U	251
U-Mann-Whitney	2813
U-Mann-Whitney	134024
U-Mann-Whitney	812
U-Mann-Whitney	852
U-Mann-Whitney	1012
U-Mann-Whitney	251
U-Mann-Whitney	812

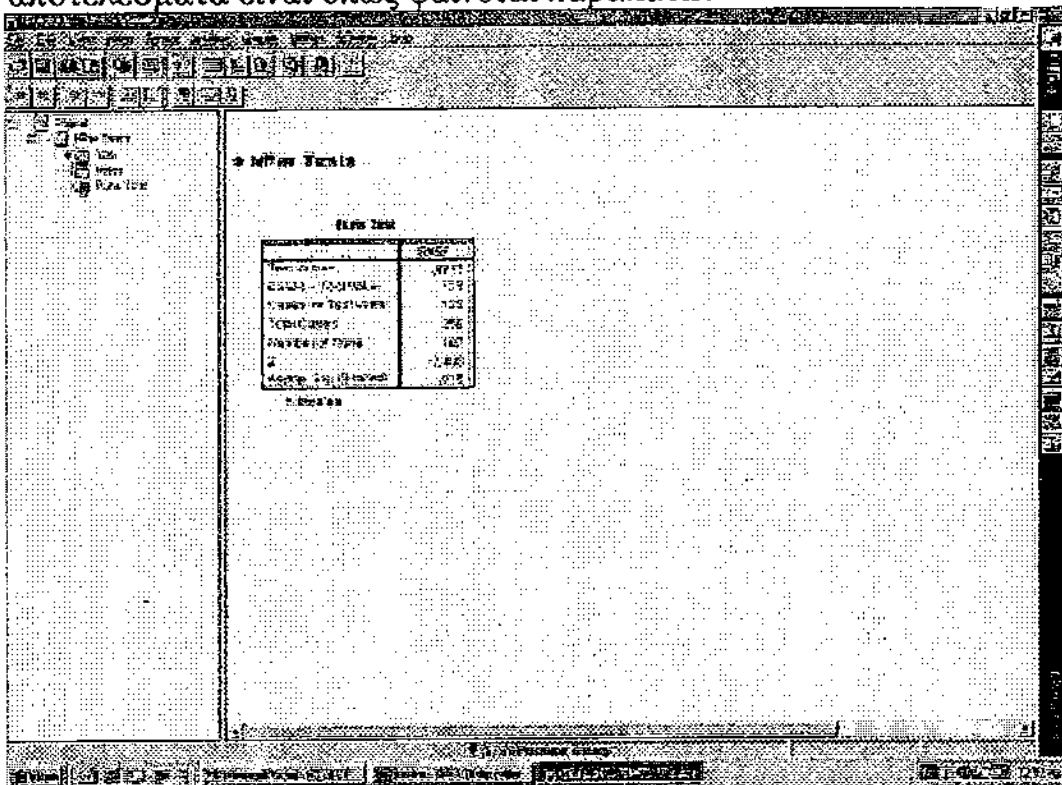
25 Από την πιθανότητα του ελέγχου (0,961) είναι σαφές ότι η υπόθεση της κανονικότητας δεν μπορεί να απορριφθεί.

Ένας άλλος χρήσιμος έλεγχος είναι ο έλεγχος ανεξαρτησίας σε μια χρονολογική σειρά που βασίζεται στα λεγόμενα runs, δηλαδή τις μεταβολές προσήμου της σειράς.

Ο έλεγχος είναι διαθέσιμος στο μενού Analyze/Nonparametric tests/Runs. Για να ελέγξουμε τις αποδόσεις των ασφαλειών για τυχαιότητα χρησιμοποιούμε τις επιλογές αυτές και έχουμε την ακόλουθη οθόνη.



Τα αποτελέσματα είναι όπως φαίνεται παρακάτω.



27 Ο έλεγχος αυτός είναι έλεγχος Z . Η τιμή της στατιστικής είναι $-2,408$ και έχει πιθανότητα $0,016$. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% μπορούμε να απορρίψουμε την υπόθεση ότι η σειρά είναι τυχαία, αλλά σε επίπεδο 1% δεν μπορούμε να την απορρίψουμε. Δεν έχουμε λοιπόν επαρκή στοιχεία για να λάβουμε μια οριστική απόφαση. Αυτό είναι ένα συμπέρασμα που βρίσκεται σε συμφωνία με όσα ξέρουμε για τις χρηματοοικονομικές χρονολογικές σειρές στις οποίες δεν είναι σαφές αν υπάρχει δομή με βάση την οποία μπορούν να προβλεφθούν.

Προβλέψεις με την μέθοδο Box-Jenkins

Το SPSS είναι ιδιαίτερα καλό στην χρησιμοποίηση των υποδειγμάτων ARIMA για την διεξαγωγή προβλέψεων σε χρονολογικές σειρές. Τα υποδείγματα Box-Jenkins ή ARIMA (autoregressive integrated moving average) είναι υποδείγματα που χρησιμοποιούνται εκτεταμένα για τις προβλέψεις χρονολογικών σειρών. Ο σκοπός των υποδειγμάτων είναι να πραγματοποιήσουν προβλέψεις για μια χρονολογική σειρά Y_t με βάση μόνον τις παρελθούσες τιμές της σειράς και χωρίς άλλη πληροφόρηση διαρθρωτικής μορφής. Πχ δεν χρειάζεται να έχουμε πληροφόρηση σχετικά με το ποιες ερμηνευτικές μεταβλητές επηρεάζουν την Y_t . Η βασική αρχή των υποδειγμάτων είναι ότι η σειρά Y_t πρέπει να είναι στάσιμη (stationary) δηλαδή να μην υπάρχει προφανής τάση στα επίπεδα ή στην διακύμανση της μεταβλητής. Αν η

σειρά δεν είναι στάσιμη στα επίπεδα θα πρέπει να πάρουμε πρώτες διαφορές, δηλαδή να σχηματίσουμε την σειρά

$$DY_t = Y_t - Y_{t-1}$$

και να εφαρμόσουμε τα υποδείγματα ARIMA στην νέα σειρά. Αν ούτε και η

είναι στάσιμη παίρνουμε δεύτερες διαφορές

$$D^2Y_t = DY_t - DY_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}$$

κλπ μέχρι να επιτύχουμε στασιμότητα την οποία κρίνουμε διαγραμματικά. Γενικά την τάξη διαφορών την οποία πρέπει να επιβάλλουμε για να έχουμε στασιμότητα την συμβολίζουμε με d .

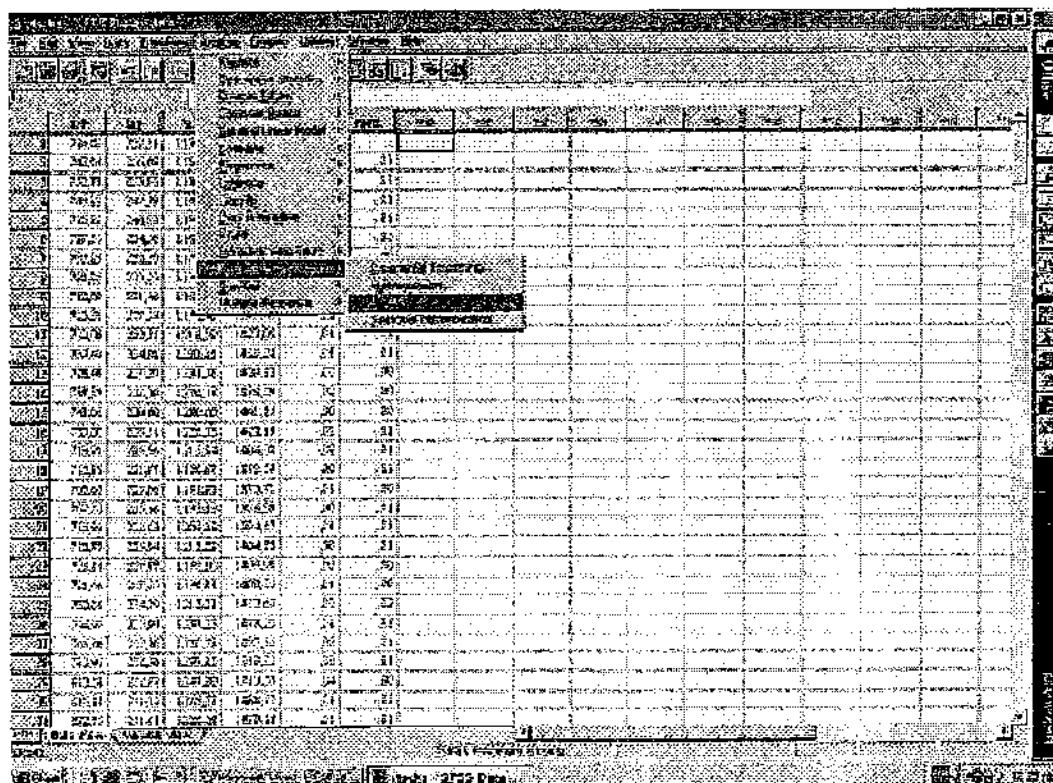
Το υπόδειγμα ARIMA(p,d,q) έχει την μορφή

$$z_t = \beta_0 + \beta_1 z_{t-1} + \beta_2 z_{t-2} + \dots + \beta_p z_{t-p} + e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2} + \dots + \alpha_q e_{t-q}$$

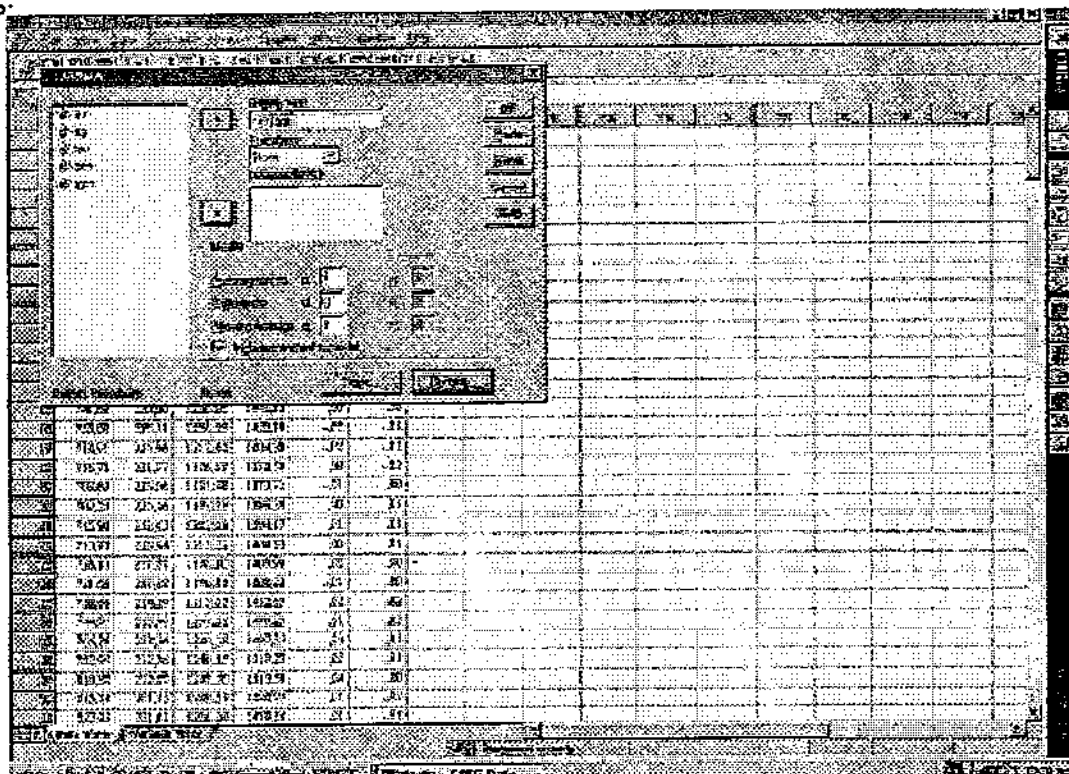
$$z_t = D^d Y_t$$

όπου β_i γνωστοί σαν αυτοπαλίνδρομοι (autoregressive) συντελεστές και α_i είναι γνωστοί σαν συντελεστές κινητού μέσου (moving average).

Για να εκτιμήσουμε ένα υπόδειγμα ARIMA επιλέγουμε τα μενού Analyze/Time Series/ARIMA.

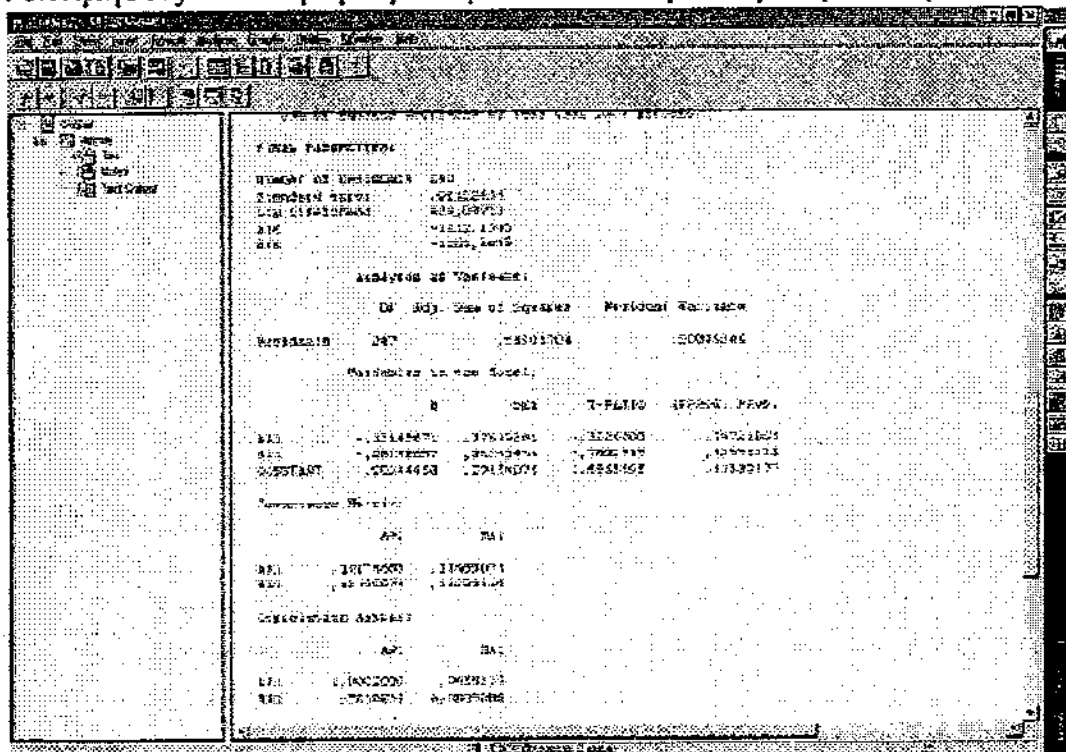


Για να εκτιμήσουμε το υπόδειγμα ARIMA(1,0,1) δίνουμε τις εξής επιλογές.



29

Οι εκτιμήσεις των παραμέτρων φαίνονται στην επόμενη οθόνη.



3

Οι εκτιμήσεις αυτές δεν έχουν κάποια διαθροωτική ερμηνεία αλλά είναι σαφές ότι τόσο ο όρος AR όσο και ο όρος MA δεν είναι στατιστικά σημαντικοί αφού οι t στατιστικές είναι πολύ μικρές. Αν επιστρέψουμε στην κύρια οθόνη θα δούμε ότι

έχουν δημιουργηθεί ορισμένες νέες σειρές που έχουν ως εξής
 fit_{-1} είναι οι προσαρμοσμένες τιμές της ARIMA παλινδρόμησης ή προβλέψεις

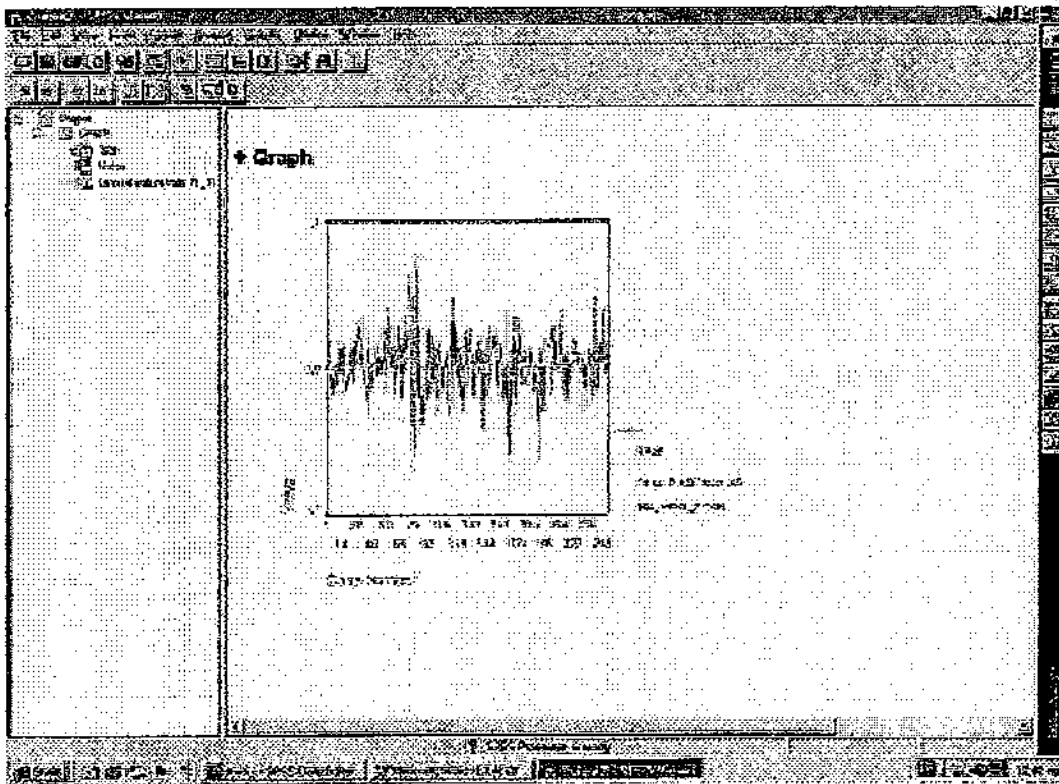
err_{-1} τα σφάλματα (οι πραγματικές τιμές μείον τις προσαρμοσμένες)

lcl_{-1} κατώτερο 95% όριο των προβλέψεων

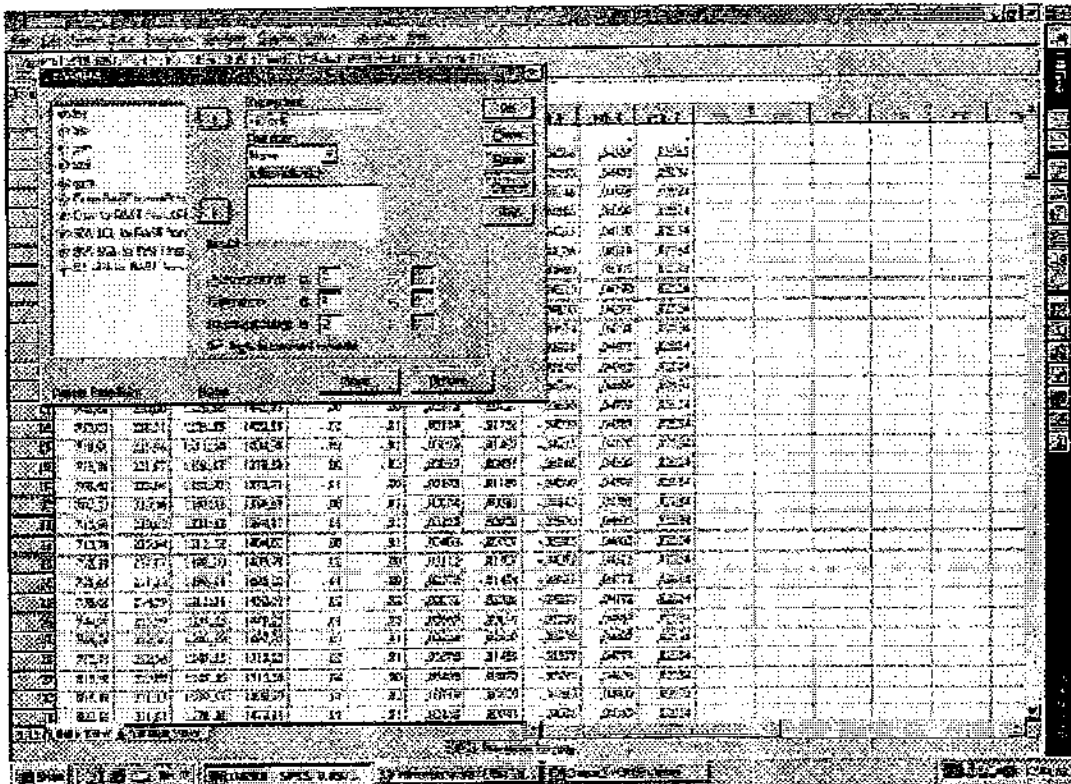
ucl_{-1} 95% όριο των προβλέψεων

sep_{-1} είναι το τυπικό σφάλμα των προβλέψεων.

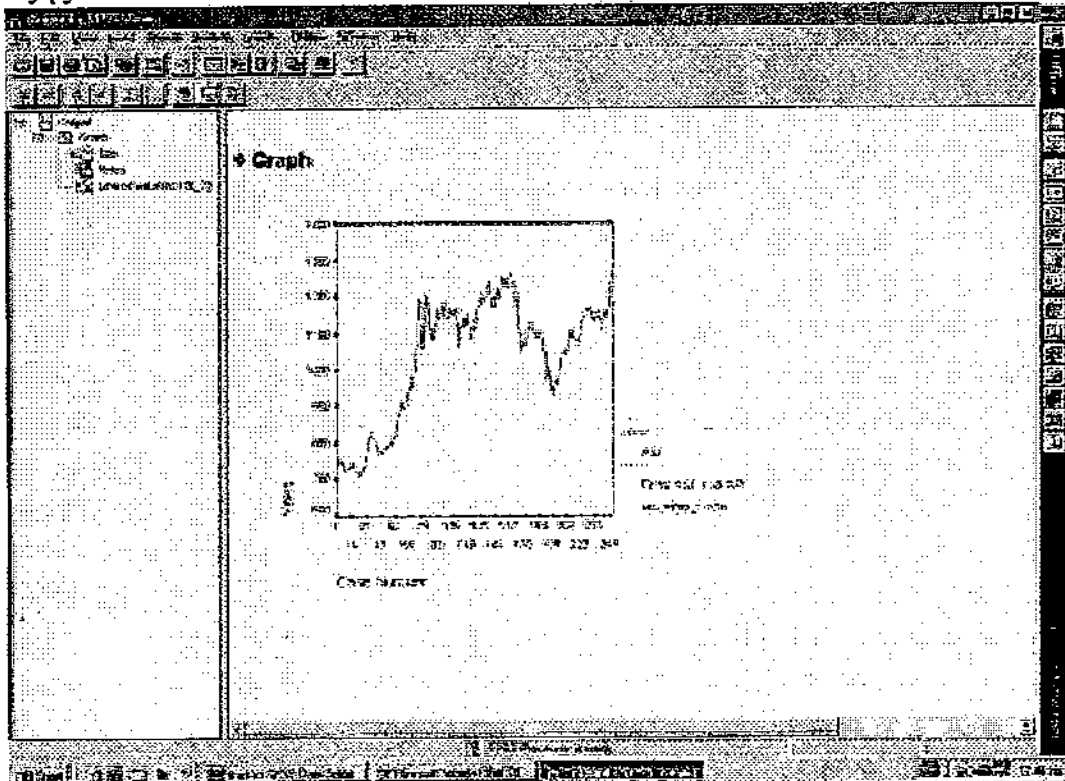
Για να κάνουμε ένα διάγραμμα των θεωρητικών και πραγματικών τιμών μπορούμε να επιλέξουμε το μενού Graph/Line ως εξής.



Ένα άλλο υπόδειγμα που θα μπορούσαμε να χρησιμοποιήσουμε είναι το $ARIMA(2,1,2)$ για την αρχική σειρά των ασφαλειών. Στην περίπτωση αυτή θα δώσουμε απευθείας την σειρά των τιμών και θα ορίσουμε $d = 1$ αντί να δώσουμε την σειρά σε πρώτες διαφορές.



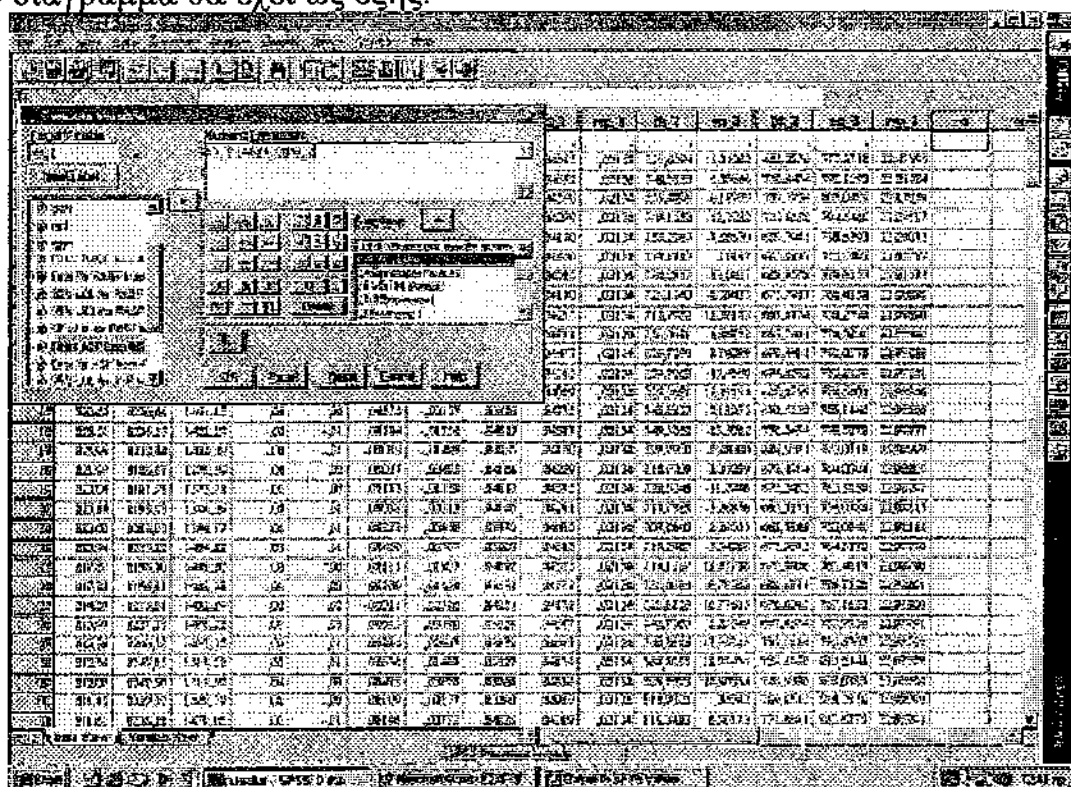
31 Όπως θα δείτε οι όροι 2 AR και 2 MA είναι στατιστικά σημαντικοί και οι προβλέψεις της αρχικής σειράς σε σχέση με τις πραγματοποιήσεις έχουν ως εξής.



32 Η καλή προσαρμογή είναι παραπλανητική και θα πρέπει να δούμε πως το υπόδειγμα συμπεριφέρεται σε όρους αποδόσεων. Για τον σκοπό αυτό πρέπει να δημιουργήσουμε την σειρά αποδόσεων που αντιστοιχεί στην

σειρά fit_2 ως εξής.

Το διάγραμμα θα έχει ως εξής.



33

Παραγοντική ανάλυση

Πολλές φορές έχουμε έναν αριθμό μεταβλητών X_1, X_2, \dots, X_m , πχ τιμές διαφόρων αγαθών και θέλουμε να διευκολύνουμε την ανάλυση με το να αντικαταστήσουμε αυτές τις σειρές με λιγότερες, πχ μία ή δυο. Σε περίπτωση που έχουμε πχ τιμές είναι λογικό να προσπαθήσουμε να τις αντικαταστήσουμε με μια σειρά που να παριστάνει κάποιον «δείκτη τιμών». Η προφανής λύση στο πρόβλημα είναι να θεωρήσουμε μια σειρά που είναι ο μέσος όρος, δηλαδή

$$Y_t = \frac{1}{m} \sum_{i=1}^m X_{it}, \quad t = 1, 2, \dots, T$$

Το θέμα είναι ότι η σειρά αυτή μπορεί να μην είναι πολύ αντιπροσωπευτική και να πρέπει να θεωρήσουμε έναν διαφορετικό γραμμικό συνδυασμό των σειρών ή ίσως να μην είναι δυνατόν να περιγράψουμε ικανοποιητικά όλες τις σειρές με μια αλλά να χρειάζονται δυο ή ίσως περισσότερες. Στην παραγοντική ανάλυση (factor analysis) το πρόβλημα είναι να βρούμε έναν γραμμικό συνδυασμό των σειρών, έστω

$$Y_{jt} = \sum_{i=1}^m a_{ij} X_{it}$$

ο οποίος να εξηγεί όσο το δυνατόν περισσότερο την διακύμανση των

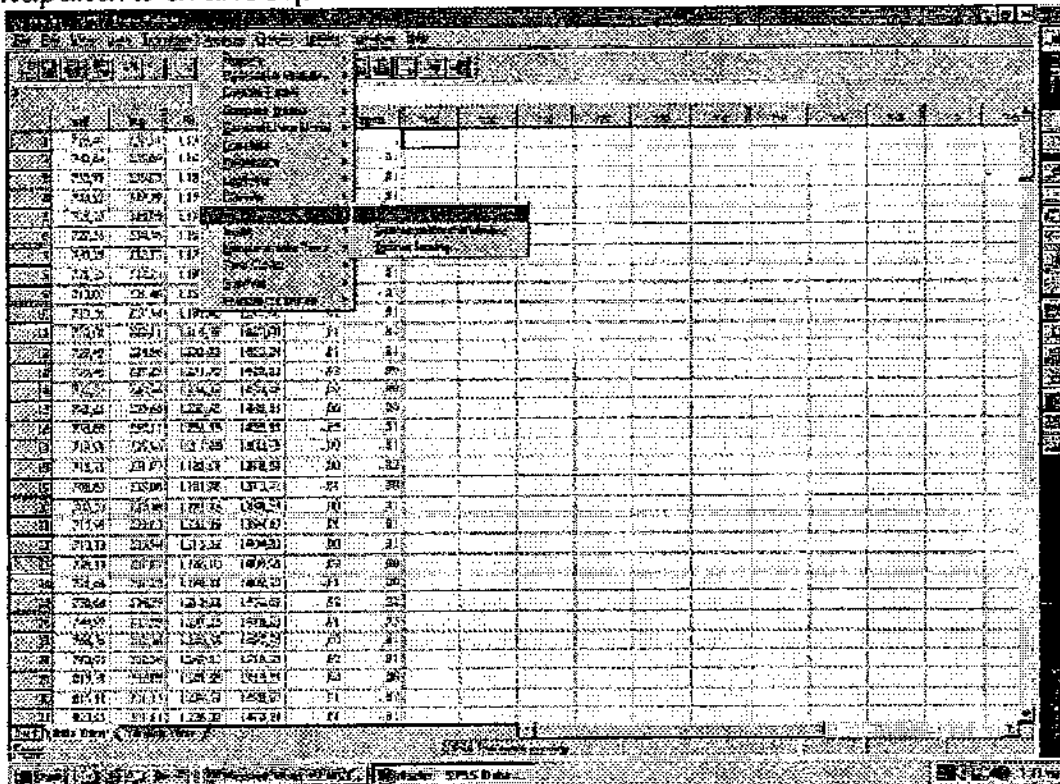
σειρών. Στην συνέχεια μπορούμε να βρούμε έναν άλλο γραμμικό συνδυασμό

$$Y_2 = \sum_{i=1}^m b_i X_i$$

που εξηγεί όσο το δυνατόν περισσότερο από την διακύμανση που απομένει με τον περιορισμό ότι η $2^{η}$ Y είναι ασυσχέτιστη με την $1^{η}$ Y . Παρόμοια μπορούμε να

κατασκευάσουμε m γραμμικούς συνδυασμούς Y_1, \dots, Y_m . Σαν παράδειγμα ας θεωρήσουμε τις σειρές τιμών *asf*, *bio* και *lsg*.

Με το μενού *Analyze/Data reduction/Factor* μπορούμε να διεξάγουμε την παραπάνω ανάλυση.



Στην συνέχεια επιλέγουμε τις σειρές που θέλουμε να αναλύσουμε, δηλαδή τις X μεταβλητές.

Case	1	2	3	4	5	6
1	100	100	100	100	100	100
2	100	100	100	100	100	100
3	100	100	100	100	100	100
4	100	100	100	100	100	100
5	100	100	100	100	100	100
6	100	100	100	100	100	100
7	100	100	100	100	100	100
8	100	100	100	100	100	100
9	100	100	100	100	100	100
10	100	100	100	100	100	100
11	100	100	100	100	100	100
12	100	100	100	100	100	100
13	100	100	100	100	100	100
14	100	100	100	100	100	100
15	100	100	100	100	100	100
16	100	100	100	100	100	100
17	100	100	100	100	100	100
18	100	100	100	100	100	100
19	100	100	100	100	100	100
20	100	100	100	100	100	100
21	100	100	100	100	100	100
22	100	100	100	100	100	100
23	100	100	100	100	100	100
24	100	100	100	100	100	100
25	100	100	100	100	100	100
26	100	100	100	100	100	100
27	100	100	100	100	100	100
28	100	100	100	100	100	100
29	100	100	100	100	100	100
30	100	100	100	100	100	100
31	100	100	100	100	100	100
32	100	100	100	100	100	100
33	100	100	100	100	100	100
34	100	100	100	100	100	100
35	100	100	100	100	100	100

35 Τα αποτελέσματα της ανάλυσης φαίνονται στην επόμενη οθόνη.

Factor Analysis

Component	Initial Eigenvalue	Extraction
1	1.923	92,073
2	.077	3,699
3	.002	0,028

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Total Variance	% of Variance	Cumulative %	Sum of Squared Multiple R	% of Squared Multiple R
1	1,923	92,073	92,073	1,910	91,971
2	.077	3,699	95,772	.009	0,428
3	.002	0,028	95,800	.000	0,000

Extraction Method: Principal Component Analysis.

Component Matrix

Component	1
AS	.921
AK	.060
LS	.002

Extraction Method: Principal Component Analysis.

36 Τα αποτελέσματα που μας ενδιαφέρουν βρίσκονται στον πίνακα "Total variance explained". Η στήλη με τίτλο "% variance" μας δίνει το ποσοστό της διακύμανσης που ερμηνεύει η κάθε συνιστώσα Y. Πχ η πρώτη συνιστώσα 1 Y ερμηνεύει το 92,073% της διακύμανσης των τριών σειρών.

Χρησιμοποιώντας την δεύτερη συνιστώσα αυξάνεται κατά 4,5%. Αυτή η μικρή αύξηση μας λέει ότι μπορούμε να σταματήσουμε στην μια συνιστώσα. Ο πίνακας με τίτλο "Component matrix" μας δίνει τους συντελεστές i a της πρώτης συνιστώσας, δηλαδή αυτή είναι

$$F_1 = 0,962X_1 + 0,953X_2 + 0,963X_3$$

Οι συντελεστές αυτοί είναι όλοι κοντά στην μονάδα οπότε η πρώτη συνιστώσα δεν είναι παρά το άθροισμα των σειρών. Οι συντελεστές i a μπορούν να κανονικοποιηθούν με διάφορους τρόπους, οπότε μπορούμε να πούμε ότι η πρώτη συνιστώσα δεν είναι παρά ο απλός μέσος των σειρών. Γενικότερα αυτό δεν θα ισχύει και οι συνιστώσες δεν μπορούν να ερμηνευθούν εύκολα σε οικονομικούς όρους. Για τον λόγο αυτό εφαρμόζονται διάφορες περιστροφές (rotations) του διανύσματος των συντελεστών με σκοπό την ευκολότερη ερμηνεία τους. Τέτοιες περιστροφές όπως πχ η λεγόμενη varimax είναι κατευθείαν διαθέσιμες στις επιλογές του SPSS.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- 1) Στατιστική Επιχειρήσεων
Douglas + Clark
- 2) Μέθοδοι και Μοντέλα Ανάλυσης Δεδομένων
Γ. Μαυρομάτης
- 3) Στατιστική για Οικονομολόγους
Δ. Χατζηνικολάου
- 4) Στατιστική με το SPSS
Denis Howitt and Duncan Crammer
- 5) Εφαρμογές της Στατιστικής
Δ. Αθανασίου και Β. Μπένος

