

Η ΜΕΘΟΔΟΛΟΓΙΑ ΤΗΣ «ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ» ΚΑΙ Η ΑΞΙΑ ΤΗΣ ΓΙΑ ΤΙΣ ΕΠΙΣΤΗΜΕΣ ΤΟΥ ΑΝΘΡΩΠΟΥ ΚΑΙ ΤΗΣ ΚΟΙΝΩΝΙΑΣ

ΘΕΟΔΩΡΟΣ ΜΠΕΧΡΑΚΗΣ *

Πολλοί υποστηρίζουν ότι τα μαθηματικά είναι η απαραίτητη γλώσσα κάθε επιστήμης που σέβεται τον εαυτό της. Είναι όμως γεγονός ότι κάθε αλήθεια δεν κερδίζει όταν λεχθεί με μαθηματικούς όρους: εξάλλου είναι εύκολο να εκφράσουμε με τους όρους αυτούς μια σκέψη άνευ περιεχομένου ή και λανθασμένη.

Έτσι σήμερα πολλές σύγχρονες επιστήμες παρουσιάζουν πληθωρισμό μαθηματικών τύπων, που συχνά δεν είναι παρά παραφράσεις λαθών ή κοινοτυπιών.

Μαθηματικά μοντέλα

Είναι πολύ διαδεδομένη η χρήση μοντέλων για την επίλυση προβλημάτων των διαφόρων επιστημονικών κλάδων. Τι είναι ένα μοντέλο; Ένα μοντέλο είναι η παράσταση οντοτήτων μιας κάποιας τάξης με οντότητες άλλης τάξης, έτσι ώστε με πράξεις που είναι δυνατόν να γίνουν στην τελευταία να μπορούμε να εξαγάγουμε συμπεράσματα για την πρώτη.

Επειδή τα μαθηματικά προσφέρονται για μια απειρία κατασκευών, είναι μια ανεξάντλητη πηγή μοντέλων για τις κοινωνικές επιστήμες. Η λογική της δημιουργίας των μοντέλων είναι η προσπάθεια μετατροπής της κοινωνικής πραγματικότητας σε μια αλληλουχία μαθηματικών τύπων που ακολουθούν τους κανόνες της μαθηματικής γλώσσας.

Δύο αντιλήψεις στατιστικής ανάλυσης

Η Στατιστική μπροστά σε ένα σύνολο πληροφοριών, που αφορούν ένα συγκεκριμένο φαινόμενο ενδέχεται να έχει δύο διαφορετικούς τρόπους αντιμετώπισης: ή έχοντας δεδομένο ένα μοντέλο – δηλαδή μια συγκεκριμένη *a priori* αντίληψη για το φαινόμενο – να προσπαθεί να το επαληθεύσει, δηλαδή τελικά να μετρήσει πόσο το μοντέλο απέχει απ' την πραγματικότητα ή χωρίς να θεωρεί δεδομένο κανένα *a priori* μοντέλο, και με μόνη προϋπόθεση ένα ικανό πλήθος στοιχείων να πετυχαίνει μια περιγραφή του φαινομένου. Έτσι διευκολύνει τον ειδικό επιστήμονα πρώτα στην παρατήρηση του υλικού του και εν συνεχεία στην εξαγωγή των συμπερασμάτων του, πράγμα το οποίο νομίζουμε ότι είναι σωστότερο.

Η ενδεχόμενη ύπαρξη κάποιου γνωστού μοντέλου, πρέπει να είναι συνέπεια μιας επεξεργασίας και όχι να προηγείται της οποιας επεξεργασίας.

Η Ανάλυση Δεδομένων

Για την «Ανάλυση Δεδομένων»¹ η αξία της στατιστικής βρίσκεται στην δυνατότητα μετασχηματισμού, του πλέον κοινού στατιστικού αντικειμένου: ενός ορθογώνιου πίνακα αριθμών.

Ο μετασχηματισμός αυτός γίνεται μετά από αριθμητικούς υπολογισμούς, για τους οποίους ο ηλεκτρονικός υπολογιστής είναι το απαραίτητο εργαλείο.

Με τον μετασχηματισμό αυτό έχουμε μια περιγραφή του αρχικού πίνακα με την αποκάλυψη των ομοιοτήτων και των πιο «υποβλητικών» σχέσεων μεταξύ των γραμμών και των στηλών του. Σε μια κοινωνιολογική έρευνα για παράδειγμα εμφανίζονται ομάδες ατόμων που επικαλύπτονται από ομάδες ποιοτικών χαρακτηριστικών. Έτσι μετά την επεξεργασία των δεδομένων μπορούμε να διαβάσουμε στην έξοδο (output) αυτό που στην είσοδο (input) ήταν δυσανάγνωστο.

Όλη η τέχνη έγκειται στην κατασκευή πινάκων και στον τρόπο μετάφρασης της ποιότητας σε ποσότητα και αντίστροφα. Θα δώσουμε μερικά παραδείγματα δημιουργίας πινάκων σε διάφορους τομείς.

Στην ανάλυση των παρατηρήσεων των βιβλίων Επιδημιών Ι και ΙΙΙ του Ιπποκράτη, η επεξεργασία γίνεται σε ένα πίνακα του οποίου οι γραμμές έχουν για όνομα τις λέξεις του κειμένου, ενώ οι στήλες αντιπροσωπεύουν τις παρατηρήσεις: σημειώνεται για παράδειγμα στη διασταύρωση της γραμμής «πυρετός» με την στήλη «3η παρατήρηση» πόσο συχνά επαναλαμβάνεται η συγκεκριμένη λέξη στην 3η παρατήρηση.

Στην επεξεργασία ενός ερωτηματολογίου με κλειστές ερωτήσεις μετατρέπουμε το ερωτηματολόγιο σε ένα πίνακα που εκφράζεται με μηδέν και ένα. Αν θεωρήσουμε την ερώτηση «Αγαπάτε το θέατρο» η οποία να δέχεται τρεις απαντήσεις (ναι, όχι, δεν απαντώ) και ένα σύνολο ατόμων Ι που απαντούν, τότε δημιουργούμε ένα πίνακα $I \times J$ σε $(0,1)$ όπου το $K(i,j)=1$ αν το άτομο αποδέχτηκε την απάντηση και μηδέν εάν όχι.

Σε άλλη ανάλυση μπορεί οι γραμμές και οι στήλες να αντιπροσωπεύουν χώρες: σημειώνουμε τότε στη διασταύρωση της γραμμής «Ελλάδα» και της στήλης «Γαλλία» το σύνολο των εξαγωγών της πρώτης στη δεύτερη για ένα συγκεκριμένο έτος.

Σε μια έρευνα του Ινστιτούτου περιβάλλοντος της Γαλλίας ζητήθηκε από τα άτομα να συνδυάσουν ένα κατάλογο θεμάτων (διακοπές, εργασία, αγάπη, μίσος...) με διάφορα χρώματα. Η επεξεργασία της έρευνας έγινε σε ένα πίνακα του οποίου οι γραμμές ήταν τα θέματα και οι στήλες τα χρώματα: σημειώνουμε για παράδειγμα στη διασταύρωση της γραμμής «διακοπές» με την στήλη «πράσινο» τον αριθμό των ατόμων που έκαναν τον συνδυασμό αυτό.

1. Για μια θεωρητική εισαγωγή στην Ανάλυση Δεδομένων μπορεί κανείς να ανατρέξει στο θεμελιώδες έργο του J.P. Benzecri με τίτλο «L'Analyse des Données» εκδόσεις Dunod, Παρίσι 1973.

* Ερευνητής στο Εθνικό Κέντρο Κοινωνικών Ερευνών

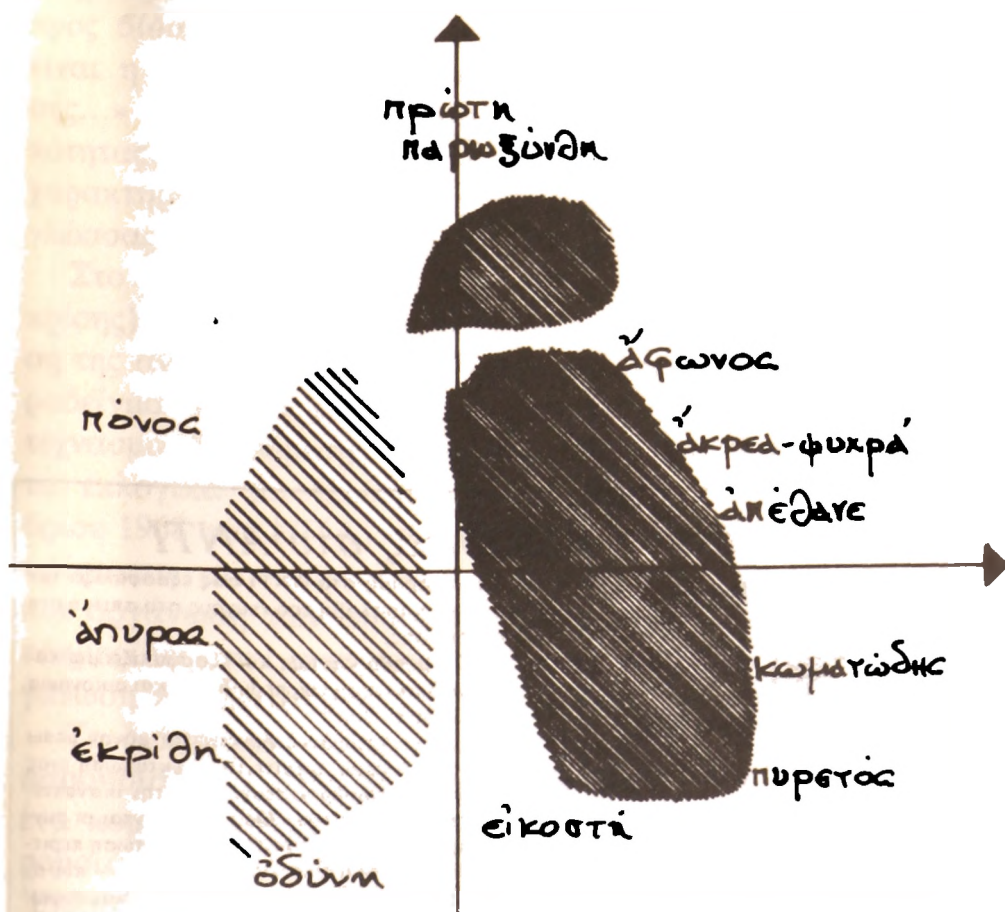


Το προγραμμα μπορεί να επεξεργαστεί πίνακες διαφόρων διαστάσεων (π.χ. 5000 γραμμές X 400 στήλες ή 100 γραμμές X 40 στήλες) ανάλογα βέβαια με τις ικανότητες (δηλαδή τη μνήμη) του υπολογιστή που χρησιμοποιούμε. Ο τύπος των αποτελεσμάτων διαφέρει ανάλογα με την μέθοδο. Θα αναφερθούμε εδώ μόνο σε δύο μεθόδους τις οποίες θεωρούμε θεμελιώδεις: την παραγοντική ανάλυση αντιστοιχιών (analyse factorielle des correspondances) και την αυτόματη ταξινόμηση (classification automatique). Με την παραγοντική ανάλυση εξάγουμε μια σειρά συναρτήσεων, τις οποίες ονομάζουμε παράγοντες. Οι παράγοντες αυτοί ορίζονται ταυτόχρονα στο σύνολο των γραμμών και στο σύνολο των στηλών.

Πολλές φορές οι δύο πρώτοι παράγοντες αρκούν να παρουσιάσουν τις ομοιότητες μεταξύ γραμμών ή μεταξύ στηλών και τις σχέσεις των μεν με τις δε. Άλλες φορές, είναι αναγκαίο να θεωρήσουμε περισσότερους παράγοντες. Σε όλες τις περιπτώσεις μπορούμε να θεωρήσουμε τα γραφήματα που ορίζουν δύο παράγοντες (p,q). Έτσι στο γράφημα (1,2) μία γραμμή i παρουσιάζεται με ένα σημείο ($F_1(i)$, $F_2(i)$) όπου $F_1(i)$ και $F_2(i)$ είναι οι τιμές του πρώτου και δεύτερου παράγοντα για το σημείο i. Παρόμοια στη στήλη C αντιστοιχεί ένα σημείο ($G_1(c)$, $G_2(c)$). Με την αυτόματη ταξινόμηση δημιουργούμε ομογενείς ομάδες στοιχείων στο εσωτερικό του ενός από τα δύο σύνολα. Την εικόνα της ταξινόμησης έχουμε και με την μορφή «δενδρογράμματος».

Οι δύο μέθοδοι συνδυάζονται αρμονικά και έτσι μπορούμε να οριοθετήσουμε στο γράφημα της παραγοντικής ανάλυσης τις ομάδες των στοιχείων που προκύπτουν απ' την ταξινόμηση.

Παράδειγμα εφαρμογής: Θα παρουσιάσουμε για παράδειγμα το γράφημα (1,2) αποτέλεσμα συνδυασμού των δύο μεθόδων, πάνω στον πίνακα αντιστοιχίας λέξεων-παρατηρήσεων που αναφέραμε ήδη.²



Ανάλυση παρατηρήσεων του Ιπποκράτη: αντιστοιχία λέξεων και ομάδων παρατηρήσεων.

Διακρίνουμε τρεις βασικές ομάδες παρατηρήσεων που αφορούν την θεραπεία, τον θάνατο και την διάρκεια της ασθένειας. Κάθε ομάδα παρατηρήσεων περιβάλλεται από τις λέξεις που την χαρακτηρίζουν.

Βλέπουμε δηλαδή ότι η ανάλυση μας έδωσε μία τυπολογία των παρατηρήσεων που αποτελεί την ανάλυση του σημασιολογικού περιεχομένου του κειμένου. Βρίσκεται δε το αποτέλεσμα αυτό σε αρμονία με την διδασκαλία του Ιπποκράτη και τις μελέτες μετέπειτα ερευνητών.

Έτσι στην συγκεκριμένη περίπτωση επιβεβαιώθηκε μια άποψη για το κείμενο αυτό, με μια μεθοδολογία άγνωστη στους φιλολογικούς ή γλωσσολογικούς κύκλους, η οποία σε καμία περίπτωση δεν βασίστηκε στην εξέταση a priori υποθέσεων.

Μεταξύ λοιπόν όλων των a priori ιδεών συχνά αντιφατικών που μας δημιουργεί ένα πρόβλημα, επιχειρείται μια «βέλτιστη» επιλογή. Θα έπρεπε να τονισθεί εδώ ότι η ιδέα που μετά την στατιστική ανάλυση των δεδομένων δηλαδή εκ των υστέρων φαίνεται ότι ήταν και εκ των προτέρων φυσική, δεν ήταν οπωσδήποτε αυτονόητη.

Λεξικολογική ανάλυση κειμένων

Δεν γνωρίζουμε κανένα μαθηματικό γλωσσολογικό ή άλλο μοντέλο που να εφαρμόζεται με επιτυχία στην ανάλυση κειμένων. Η «ανάλυση δεδομένων» με μια σειρά εφαρμογών που είχαν γίνει σε κείμενα ή που βρίσκονται σε εξέλιξη, δίνει ενθαρρυντικά αποτελέσματα για την επίλυση προβλημάτων που άπτονται της «νοημοσύνης» όπως η αναγνώριση του θέματος ενός κειμένου, ή ανάλυση του σημασιολογικού του περιεχομένου, η αναγνώριση του στίλ ή της αυθεντικότητάς του, η εύρεση λέξεων-κλειδιών κλπ.

Πρέπει να σημειώσουμε ότι ο πίνακας αντιστοιχίας λέξεων-κειμένων δημιουργείται αυτόματα με την εφαρμογή απ' ευθείας στο κείμενο –το οποίο βρίσκεται χωρίς καμία κωδικοποίηση σε ταινία ή δίσκο– ενός κατάλληλου προγράμματος. Επίσης είναι δυνατό εκτός από τις χαρακτηριστικές λέξεις ενός κειμένου να έχουμε και χαρακτηριστικές φράσεις.³

Για την συνέχεια της έρευνας στον τομέα της ανάλυσης κειμένων έχουμε να κάνουμε δύο προτάσεις για άμεσες εφαρμογές: α) Την ανάλυση των ανοιχτών ερωτήσεων σε έρευνες κοινής γνώμης και β) την ανάλυση συνεντεύξεων.

Επεξεργασία ανοιχτών ερωτήσεων

Χρησιμοποιούμε τις ανοιχτές ερωτήσεις στις κοινωνικοοικονομικές έρευνες για διάφορους λόγους: είναι κλασικό να αφήνουμε ανοιχτή μια ερώτηση, όταν το φάσμα των απαντήσεων είναι μόνο εν μέρει γνωστό. Μπορεί αντίθετα το φάσμα των απαντήσεων να είναι γνωστό αλλά αρκετά εκτεταμένο ή πολύπλοκο ή να περιλαμβάνει απαντήσεις σύνθετες ή ακόμη οι απα-

2. Για μια διεξοδική ανάπτυξη του παραδείγματος αυτού μπορεί να ανατρέξει κανείς στο τεύχος 59 της Επιθεώρησης Κοινωνικών Ερευνών, στο άρθρο του Θ. Μπεχράκη, «Η εισαγωγή του αυτοματισμού στην ανάλυση κειμένων». 1986.

3. Για μια ανάπτυξη της μεθοδολογίας επιλογής των χαρακτηριστικών φράσεων μπορεί να ανατρέξει κανείς στο τεύχος 1 του 1982 της Επιθεώρησης Consomation, στο άρθρο του L. Lebart «L' analyse statistique des réponses libres dans les enquêtes socio-économiques».

ντήσεις να μην μπορούν να προβλεφτούν τη στιγμή της δημιουργίας του ερωτηματολογίου. Σε όλες αυτές τις περιπτώσεις, οι ανοιχτές ερωτήσεις προσφέρουν πληροφορία πλούσια και εκτεταμένη, που έχει χαρακτήρα περισσότερο έκφρασης παρά αντίδρασης.

Η κλασική αντιμετώπιση των ανοιχτών ερωτήσεων είναι να δημιουργούμε για κάθε ανοιχτή ερώτηση μια ή περισσότερες κλειστές ερωτήσεις όπως κατά τη γνώμη μας προκύπτουν από ένα δείγμα των απαντήσεων του ερωτηματολογίου. Κατόπιν κωδικογράφοι απαντούν στις ερωτήσεις αυτές, κωδικοποιώντας τις ανοιχτές ερωτήσεις για το σύνολο του ερωτηματολογίου.

Η αντιμετώπιση αυτή έχει αρκετά μειονεκτήματα: Η μεσολάβηση του κωδικογράφου εισάγει αυθαιρεσία. Η πληροφορία αναγκαστικά φτωχαίνει: η ποιότητα της έκφρασης, η ιδιαιτερότητα του λεξιλογίου χάνεται όπως και οι σπάνιες και πρωτότυπες απαντήσεις.

Μπορούμε να πούμε ότι οι ανοιχτές ερωτήσεις μας δίνουν μια πληροφορία που είναι δύσκολο να χρησιμοποιηθεί με τα κλασικά σχήματα. Η επεξεργασία που μπορεί να γίνει με την Ανάλυση Δεδομένων είναι η εξής: Δημιουργούμε ένα πίνακα T, που να έχει τόσες γραμμές όσα τα άτομα που ερωτήθηκαν και τόσες κολόνες όσες λέξεις διακεκριμένες χρησιμοποιήθηκαν στην απάντηση. Κάθε απάντηση περιγράφεται με μία γραμμή από 0,1,2,... ανάλογα με το αν η αντίστοιχη λέξη είναι απύσχα, παρουσιάζεται μια, δυο φορές κλπ. στην απάντηση. Ο τρόπος αυτός παρουσίασης μας επιτρέπει να εργαστούμε σε ένα «χώρο απαντήσεων» και μας παρέχει ενδιαφέροντα εργαλεία περιγραφής. Κατόπιν μπορούμε να χωρίσουμε τα άτομα σε ομάδες, κατά το δυνατό ομογενείς, χρησιμοποιώντας την αυτόματη ταξινόμηση. Έτσι ο αρχικός πίνακας T διαστάσεων π.χ. (5000 άτομα X αριθμός λέξεων) μετατρέπεται σε ένα πίνακα C διαστάσεων (10 κλάσεις X αριθμός λέξεων), όπου είναι ο αριθμός των εμφανίσεων της λέξης στην κλάση i. Μπορούμε να έχουμε μια γραφική παρουσίαση του πίνακα αυτού με την εφαρμογή της παραγοντικής ανάλυσης. Τέλος μπορούμε να εμφανίσουμε για κάθε κλάση τις πιο χαρακτηριστικές απαντήσεις. Αυτό επιτυγχάνεται με την εισαγωγή των στοιχείων του T σαν συμπληρωματικά στην ανάλυση του πίνακα C.

Με την μεθοδολογία αυτή αντιμετώπισης των ανοιχτών ερωτήσεων πιστεύουμε να πετύχουμε τον περιορισμό της υποκειμενικότητας της επεξεργασίας και την μείωση του κόστους της έρευνας. Εφαρμόζεται με επιτυχία σε διάφορα κέντρα ερευνών κοινής γνώμης της Γαλλίας και σκοπεύουμε να την εφαρμόσουμε πειραματικά, μέσα στο 1986, σε έρευνες του ΕΚΚΕ.

Ανάλυση συνεντεύξεων

Η κλασική μεθοδολογία για την επεξεργασία συνεντεύξεων είναι η ανάλυση περιεχομένου (analyse de contenu) που συνίσταται στην εύρεση των θεμάτων που ενυπάρχουν στην συνέντευξη μετά την μελέτη που γίνεται απ' τον ερευνητή.

Αποφασίσαμε — μια μικρή διεπιστημονική ομάδα⁴ — να εφαρμόσουμε την μεθοδολογία που αναφέρθηκε στην ανάλυση συνεντεύξεων μεγάλης διάρκειας που αφορούν το ψυχολογικό βίωμα της ελληνίδας που αποφασίζει να κάνει έκτρωση⁵. Πρόκειται για συνεντεύξεις στις οποίες, αφού στην αρχή τεθεί ένα μίνιμουμ πλαίσιο συζήτησης, το άτομο ομιλεί ακολουθώντας τον ρυθμό που θέλει, προσθέτοντας ή αφαιρώντας θέματα και χω-

ρίς να υφίσταται την διαδικασία της ερώτησης-απάντησης.

Δύο είναι οι υποθέσεις οι οποίες πιστεύουμε ότι κάνουν δυνατή μια τέτοια εφαρμογή

Πρώτον ότι οι μηχανισμοί του λόγου είναι αποκαλυπτικοί των ψυχολογικών μηχανισμών άμυνας και του τρόπου διανοητικής λειτουργίας και δεύτερον ότι η επανάληψη, όχι μόνο λέξεων αλλά και ολόκληρων φράσεων καθορίζει την ομιλία κάποιου, ο οποίος μιλάει εκτός χειρόγραφου και χωρίς να έχει προετοιμαστεί.

Κάθε συνέντευξη αλλά και το σύνολο των συνεντεύξεων θα περιγραφούν από μία σειρά «κομβικών σημείων» δηλ. τόσο θεμάτων κοινών όσο και σημείων διαφοροποίησης. Έτσι η ανάλυση αυτή θα μας επιτρέψει να έχουμε μια οπτική «ενδο-ατομική» και μια οπτική «δια-ατομική». Εκτός των στοιχείων που μπορούν να μας οδηγήσουν σε μια θεματική ανάλυση, μπορούμε να έχουμε επίσης στοιχεία για μια υφολογική ανάλυση: επαναλήψεις λέξεων και φράσεων, άτυπα στοιχεία, θέση των λειτουργικών λέξεων και άλλων γραμματικών τύπων μέσα στο corpus. Εντούτοις η επιλογή της μεθόδου βρίσκεται στη φάση της έρευνας και έτσι είμαστε επιφυλακτικοί ως προς τα τελικά αποτελέσματα.

Βέβαια δεν πρόκειται παρά για ένα σχέδιο έρευνας του οποίου μόλις περιγράφουμε τις βασικές γραμμές. Αλλά το ιδιαίτερο ενδιαφέρον μας βρίσκεται στην εφαρμογή μιας νέας μεθοδολογίας ανάλυσης συνεντεύξεων πάνω σε κλινικό υλικό, καθώς επίσης στον διεπιστημονικό χαρακτήρα της προσπάθειας αυτής.

4. Η ομάδα αυτή αποτελείται από την Δ. Ναζίρη ψυχολόγο, τον Θ. Τζαβάρα ψυχίατρο, την Ο. Πανοπούλου-Μαράτου Αναπλ. καθηγήτρια ψυχολογίας στο Πανεπιστήμιο Κρήτης και τον Θ. Μπεχράκη ειδικό στην Ανάλυση Δεδομένων.
5. Για περισσότερες πληροφορίες γύρω από την έρευνα αυτή παραπέμπουμε στο άρθρο της Δ. Ναζίρη, στο ίδιο τεύχος.

Γραφτείτε συνδρομητές στο ANTI

Είναι
χρήσιμο
για σας
και
βοηθά
το
περιοδικό



Η συνδρομή στο ANTI σας εξασφαλίζει την τακτική παραλαβή του τεύχους στο σπίτι ή στη δουλειά σας.

Αλλά βοηθά και μας και εξασφαλίζει μια καλύτερη διακίνηση του περιοδικού και οικονομία στην έκδοσή του.

Όταν το περιοδικό διακινείται μονον μέσω του Πρακτορείου απαιτείται η εκτύπωση ενός μεγαλύτερου αριθμού τευχών για την ικανοποίηση της ζήτησής του. Όσο αυξάνονται οι συνδρομητές, τόσο περιορίζεται η εκτύπωση περιττών τευχών και η οικονομία στο χαρτί — που αποτελεί ένα σημαντικό στοιχείο στη διαμόρφωση του κόστους του περιοδικού.

Γραφτείτε σήμερα συνδρομητές στο ANTI.

Έχετε μια τακτική επαφή με την έκδοσή του και συγχρόνως βοηθάτε το περιοδικό.