



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

**Τμήμα Μηχανικών Παραγωγής
και Διοίκησης**

**Διπλωματική Εργασία
Χαρίλαος Α. Ανεσιάδης**

**"Μεθοδολογία Ποιοτικής Ανάλυσης της Συμπεριφοράς των
Ψηφοφόρων και Λήψη Πολιτικών Αποφάσεων"**

Εξεταστική Επιτροπή

**Ματσατσίνης Νικόλαος, Καθηγητής (επιβλέπων)
Γρηγορούδης Ευάγγελος, Επίκουρος Καθηγητής
Μαρινάκης Γιάννης, Λέκτορας**

Χανιά 2009

Ευχαριστίες

Παρουσιάζοντας το τελικό κείμενο, θα ήθελα να ευχαριστήσω όλους όσους συνέβαλαν στην πραγματοποίησή του.

Ευχαριστώ ολόψυχα τον επιβλέποντα καθηγητή κ. Ματσατσίνη Νικόλαο, χωρίς την βοήθεια του οποίου δεν θα υπήρχε η παρούσα διπλωματική. Εκφράζω την ευγνωμοσύνη μου, τόσο για την επιστημονική του καθοδήγηση όσο και για την ανθρώπινη βοήθειά του, που μου παρείχε αφειδώς σε όλη τη διαδικασία εκπόνησης της διπλωματικής μου.

Επίσης, ευχαριστώ θερμά τα άλλα δύο μέλη της τριμελούς εξεταστικής επιτροπής, τον επίκουρο καθηγητή κ. Ευάγγελο Γρηγορούδη και τον λέκτορα καθηγητή κ. Γιάννη Μαρινάκη, για τις παρατηρήσεις τους στη βελτίωση της εργασίας και για τις χρήσιμες υποδείξεις τους, κατά το τελικό στάδιο της διπλωματικής.

Οφείλω, επίσης, να εκφράσω τις ευχαριστίες μου στην κα Ευαγγελία Κρασαδάκη για την ανεκτίμητη βοήθειά της κάθε φορά που την χρειάστηκα καθώς και στον Παύλο Δελιά για τις χρήσιμες πληροφορίες και γνώσεις που αφειδώς μου προσέφερε.

Τέλος, θέλω να ευχαριστήσω τους φίλους και συμφοιτητές μου όλων των τμημάτων του Πολυτεχνείου Κρήτης και ιδιαιτέρως του τμήματος Μηχανικών Παραγωγής και Διοίκησης καθώς και τους Σακελλάρη Αλκαίο, Κούκνη Επαμεινώνδα και το προσωπικό μου περιβάλλον, για κάθε υλική και ηθική βοήθεια που μου προσέφεραν.

*Στους γονείς μου,
Ευθυμία και Αντώνη
και στην αδερφή μου,
Κασσιανή*

*Στον νουνό μου,
Βασίλη*

«είναι δύσκολο να βάλει κανείς τον εαυτό του στο μέλλον και να φανταστεί τι θα βιώσει. Είναι δύσκολο να επιτευχθεί ένα τέτοιο νοητικό ταξίδι στο χρόνο, μια τέτοια προσομοίωση που ίσως τελικά να οδηγούσε και σε καλύτερες αποφάσεις» (Wilson, T. D. & Gilbert, D. T. (2005). Affective Forecasting. Current Directions in Psychological Science, 14, 131 – 134)

ΠΕΡΙΛΗΨΗ

Στο παρόν εισαγωγικό σημείωμα θα γίνει μια συνοπτική περιγραφή της διάρθρωσης της διπλωματικής εργασίας (“Μεθοδολογία Ποιοτικής Ανάλυσης της Συμπεριφοράς των Ψηφοφόρων και Λήψη Πολιτικών Αποφάσεων”) και των στόχων που επιτυγχάνει.

Η διάρθρωση των κεφαλαίων αντικατοπτρίζει τους δύο στόχους της διπλωματικής εργασίας. Ο πρώτος στόχος είναι να γίνει μια βιβλιογραφική μελέτη της Θεωρίας των Αποφάσεων σε συνδυασμό με την Πολυκριτήρια Ανάλυση. Η μελέτη κατευθύνθηκε, έτσι ώστε να οδηγήσει στην επίτευξη του δεύτερου στόχου, που είναι η πρόταση της μεθοδολογίας της Εξόρυξης Δεδομένων (Data Mining) και η υλοποίησή της στην πρακτική εφαρμογή (case study) που έχουμε πραγματοποιήσει, προκειμένου να εξαχθούν χρήσιμα συμπεράσματα ως προς τη λήψη πολιτικών αποφάσεων.

Τα κεφάλαια της εργασίας αναπτύσσονται ως εξής:

- **Κεφάλαιο 1: Εισαγωγή**

Στο 1^ο κεφάλαιο, λοιπόν, γίνεται η εισαγωγή στο σκοπό και στην προβληματική που μας οδήγησαν στη συγγραφή αυτής της διπλωματικής εργασίας. Μελετούμε το ιστορικό και το ρόλο των δημοσκοπήσεων στην κατανόηση και την πρόβλεψη της ψήφου των πολιτών και ξεκαθαρίζουμε τις εναλλακτικές μας προτάσεις - όσον αφορά τη μελέτη και ανάλυση της ψήφου - και τους στόχους που θέλουμε να υπηρετήσουμε με την παρούσα διπλωματική εργασία.

- **Κεφάλαιο 2: Υφιστάμενη μεθοδολογία ερευνών αγοράς / κοινής γνώμης**

Στο 2^ο κεφάλαιο, πραγματοποιούμε μία πρωταρχική ανάλυση των υφιστάμενων συστημάτων-μεθόδων ερευνών αγοράς και ανάλυσης της συμπεριφοράς των πολιτών κατά την εκλογική διαδικασία. Παρουσιάζουμε τις μεταβλητές που συνήθως χρησιμοποιούνται, καθώς και τις κλίμακες μέτρησής τους. Αναφερόμαστε στα εκλογικά συστήματα και τις μεθόδους εκλογής-επιλογής καθώς και στους παράγοντες που επιδρούν, για να κατανοήσουμε τα κριτήρια επιλογής (ποιοτικά – ποσοτικά) και να αναλύσουμε τις προτιμήσεις των ψηφοφόρων.

- **Κεφάλαιο 3: Μεθοδολογικό υπόβαθρο προτεινόμενης μεθοδολογίας ποιοτικής ανάλυσης συμπεριφοράς των ψηφοφόρων**

Στο 3^ο κεφάλαιο, μελετούμε τη Θεωρία των Αποφάσεων και το ρόλο της σε θέματα διοίκησης αλλά και πολιτικής. Αρχικά μελετούμε το ορθολογικό πρότυπο απ' όπου προκύπτει ότι η απόφαση ως επιλογή μεταξύ εναλλακτικών λύσεων αποκτά νόημα – και μπορεί κατά συνέπεια να αξιολογηθεί λογικά – στο μέτρο που είναι προσανατολισμένη στην επίτευξη ορισμένων στόχων. Στη συνέχεια μελετούμε το αυξητικό πρότυπο απόφασης που προβάλλει ως μεθοδολογία το ερώτημα και την αμφισβήτηση για τη διαιρετότητα και διακρισιμότητα απ' ενός μεν στόχων και μέσων, απ' ετέρου δε αξιολογικών και πραγματικών δεδομένων, και την υπαγωγή των δεύτερων στα πρώτα. Ακολουθεί η διερεύνηση ορισμένων ενδιάμεσων θεωρήσεων της διαδικασίας των αποφάσεων και της δημόσιας πολιτικής που οριοθετούνται μεταξύ του τέλει ορθολογισμού και του καθαρού αυξητισμού. Τέλος, διατυπώνουμε τη θεωρία της Πολυκριτήριας Ανάλυσης των Αποφάσεων, τη μεθοδολογία μοντελοποίησής της, καθώς και το ρόλο της στην ανάλυση των δεδομένων. Η αποτύπωση του θεωρητικού λοιπόν υποβάθρου μας οδηγεί στην ανάλυση της προτεινόμενης μεθοδολογίας ποιοτικής ανάλυσης της συμπεριφοράς των ψηφοφόρων, δηλαδή στην Εξόρυξη Γνώσης και Δεδομένων (Data Mining). Εξηγούμε τις βασικές εργασίες της Εξόρυξης Δεδομένων καθώς και τα μοντέλα και τους αλγόριθμους που χρησιμοποιεί. Στο σημείο αυτό έχει ολοκληρωθεί η θεωρητική μελέτη που οδηγεί στην πρόταση της συγκεκριμένης μεθοδολογίας (Data Mining) για την κατανόηση και επεξήγηση της απόφασης της ψήφου. Η πρότασή μας βασίζεται σε ένα συγκερασμό του θεωρητικού υποβάθρου που ήδη έχει αναπτυχθεί και των απαιτήσεων που πρέπει να ικανοποιούνται από μια διαδικασία λήψης απόφασης και την ποιοτική ανάλυση αυτής.

- **Κεφάλαιο 4: Υλοποίηση πραγματικής εφαρμογής (case studie)**

Στο 4^ο κεφάλαιο, γίνεται η παρουσίαση της υλοποίησης της πρακτικής εφαρμογής μας (case study). Με τη χρήση ενός δομημένου ερωτηματολογίου αναζητήσαμε απαντήσεις σε δείγμα 400 πολιτών του Δήμου Σερρών. Η στατιστική ανάλυση του δείγματος, αλλά κυρίως η χρήση του υπολογιστικού συστήματος Weka για την Εξόρυξη της Γνώσης και των Δεδομένων, μας βοήθησε στο να επαληθεύσουμε, αλλά και να θέσουμε νέα ερωτήματα σε σχέση με τη μεθοδολογία που ακολουθήσαμε για την ποιοτική ανάλυση της συμπεριφοράς των πολιτών και τη λήψη πολιτικών αποφάσεων.

- **Κεφάλαιο 5: Συμπεράσματα**

Στο 5^ο κεφάλαιο, γίνεται η παρουσίαση των συμπερασμάτων της διπλωματικής μας εργασίας. Ελέγχουμε αν ανταποκριθήκαμε στους στόχους που τέθηκαν στην αρχή, αναδεικνύουμε τα κύρια σημεία της και περιγράφουμε αν η έρευνά μας κάλυψε (εξάντλησε) το θέμα καθώς και ποιες προοπτικές υπάρχουν για να αναπτυχθεί ακόμη περισσότερο (το θέμα μας), σε μελλοντικές εργασίες.

- **Βιβλιογραφία**

- **Παράρτημα**

Ερωτηματολόγιο της πρακτικής εφαρμογής μας

Τυπολόγιο δειγματοληπτικών μεθόδων

Περιεχόμενα

ΚΕΦΑΛΑΙΟ 1: Εισαγωγή.....	10
1.1 Ο ρόλος των δημοσκοπικών μεθόδων	11
1.2 Οι στόχοι της διπλωματικής εργασίας	13
ΚΕΦΑΛΑΙΟ 2:.....	14
Μεθοδολογία ερευνών αγοράς / κοινής γνώμης	14
2.1 Έρευνες αγοράς, έρευνες κοινής γνώμης: Μια μικρή αναδρομή	14
2.2 Ποσοτικές έρευνες	17
2.2.1 Σχεδιασμός της Έρευνας.....	18
2.2.2 Κατάρτιση ερωτηματολογίων και τρόποι συνέντευξης.....	18
2.2.2.1 Γενικές παρατηρήσεις για την κατάρτιση του ερωτηματολογίου ...	18
2.2.2.2 Τύποι ερώτησης	19
2.2.2.3 Προ-εξέταση του ερωτηματολογίου	19
2.2.2.4 Μέθοδοι συνέντευξης	20
2.2.3 Δειγματοληψία	22
2.2.3.1 Γενικά.....	22
2.2.3.2 Δειγματοληψία και απογραφή	22
2.2.3.3 Μέθοδοι δειγματοληψίας.....	23
2.2.4 Μεταβλητές (κριτήρια) και κλίμακες	26
2.2.4.1 Τι είναι οι μεταβλητές.....	26
2.2.4.2 Εξαρτημένες και ανεξάρτητες μεταβλητές.....	26
2.2.4.3 Μέτρηση των μεταβλητών :	27
2.2.4.4 Σχέσεις μεταξύ των μεταβλητών	27
2.2.4.5 Κλίμακες μέτρησης.....	29
2.2.5 Υλοποίηση της Έρευνας	30
2.2.5.1 Διεξαγωγή Επιτόπιας Έρευνας	30
2.2.6 Επεξεργασία Δεδομένων.....	31
2.2.7 Έλεγχος.....	31
2.2.8 Κωδικογράφηση και Εισαγωγή Στοιχείων στον Ηλεκτρονικό Υπολογιστή (H/Y)	32
2.2.9 Στατιστική Επεξεργασία των Δεδομένων	33
2.3 Ποιοτικές έρευνες	35
2.4 Το εκλογικό σύστημα	36
2.4.1 Ψηφοφορία.....	37
2.4.2 Κριτήρια επιλογής / εκλογής	39
ΚΕΦΑΛΑΙΟ 3:.....	42
Μεθοδολογικό Υπόβαθρο	42
Η Θεωρία των Αποφάσεων & η Πολυκριτήρια Ανάλυση.....	42
3.1 Μία περιγραφική προσέγγιση της θεωρίας των αποφάσεων	42
3.2 Το ορθολογικό παράδειγμα.....	44
3.3 Το αυξητικό πρότυπο απόφασης	48
3.4 Ενδεχόμενη Προσέγγιση.....	55
3.4.1 Dror	55
3.4.2 Etzioni	56
3.4.3 Lindblom.....	57
3.5 Συμπερασματική προσέγγιση της Θεωρίας των Αποφάσεων	58
3.6 Πολυκριτήρια Ανάλυση Αποφάσεων	60
3.6.1 Μεθοδολογία μοντελοποίησης στην Πολυκριτήρια Ανάλυση	62
3.6.2 Η Ανάλυση των Δεδομένων	65

3.6.3 Από την Πολυκριτήρια Ανάλυση στην Εξόρυξη Δεδομένων (Data Mining).....	68
Η Εξόρυξη Γνώσης και Δεδομένων:	69
(Προτεινόμενη μεθοδολογία ποιοτικής ανάλυσης της συμπεριφοράς των ψηφοφόρων)	69
3.7 Data Mining (Εξόρυξη Δεδομένων)	69
3.8 Εξόρυξη δεδομένων: Τι δεν μπορεί να κάνει!	71
3.9 Εξόρυξη δεδομένων και αποθήκευση δεδομένων	72
3.10 Εξόρυξη δεδομένων, Μηχανική μάθησης και Στατιστική	73
3.11 Εφαρμογές εξόρυξης δεδομένων	73
3.12 Βασικές Εργασίες του Data Mining.....	74
3.12.1 Ομαδοποίηση (Description Data Mining)	75
3.12.2 Ταξινόμηση (Predictive Data Mining).....	75
3.12.3 Παλινδρόμηση	76
3.13 Μοντέλα και Αλγόριθμοι Εξόρυξης Δεδομένων	76
3.13.1 Νευρωνικά Δίκτυα	76
3.13.2 Δέντρα Αποφάσεων	77
3.13.3 Διαφορική ανάλυση	78
3.14 Αξιολόγηση εργαλείων Data Mining	79
3.15 Το μέλλον του Data Mining.....	79
ΚΕΦΑΛΑΙΟ 4:	82
Υλοποίηση πρακτικής εφαρμογής (case study)	82
4.1 Γενικά.....	82
4.2 ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ.....	83
4.2.1 Παρουσίαση δημογραφικών στοιχείων	83
4.2.2 Παρουσίαση αποτελεσμάτων στατιστικής ανάλυσης.....	87
4.3 ΠΟΙΟΤΙΚΗ ΑΝΑΛΥΣΗ:	96
Παρουσίαση αποτελεσμάτων ποιοτικής ανάλυσης με τη χρήση του υπολογιστικού συστήματος Weka.....	96
4.3.1 Μια περιγραφική ανάλυση του υπολογιστικού συστήματος Weka: Wekato Environment for knowledge Analysis	96
4.3.2 Προ – επεξεργασία Δεδομένων	103
4.3.3 Συσταδοποίηση Δεδομένων	106
4.3.3.1 Ο αλγόριθμος K-Means για τη μη εποπτευόμενη εκμάθηση (Unsupervised Learning)	106
4.3.4 Παρουσίαση και περιγραφή αποτελεσμάτων της έρευνάς μας, για τη συσταδοποίηση των δεδομένων:.....	109
4.3.5 Κατηγοριοποίηση Δεδομένων / Ταξινομητές Απόφασης.....	118
4.3.5.1 Δέντρα αποφάσεων εποπτευόμενης μάθησης.....	119
4.3.5.2 Επεξήγηση προτύπου Δέντρου Απόφασης του Weka:	119
4.3.5.3 Μέτρηση Σφαλμάτων, παραγόμενα σφάλματα (error output):	122
4.3.6 Παρουσίαση και περιγραφή αποτελεσμάτων της έρευνάς μας, για την κατηγοριοποίηση - ταξινόμηση των δεδομένων:.....	126
4.3.7 Μία επιπλέον χρήσιμη ιδιότητα ταξινόμησης του υπολογιστικού συστήματος Weka.....	137
ΚΕΦΑΛΑΙΟ 5:	140
Συμπεράσματα	140
ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	143
ΠΑΡΑΡΤΗΜΑ	147

ΚΕΦΑΛΑΙΟ 1: Εισαγωγή

Εισαγωγή

Τα προβλήματα απόφασης είναι βασικό χαρακτηριστικό της ανθρώπινης δραστηριότητας. Καθημερινά, καλούμαστε να αποφασίσουμε για διάφορα μικρά ή μεγάλα ερωτήματα, να αποφασίσουμε για τα πιο απλά ή και πιο σύνθετα προβλήματα που προκύπτουν στη ζωή μας. Εύκολη ή δύσκολη, σωστή ή λάθος η λήψη μιας απόφασης καθορίζει τελικά τα επόμενα βήματά μας, τη σχέση μας με τους γύρω, ίσως και τη μετέπειτα πορεία μας ή και την ίδια μας τη ζωή. Ένα σημαντικό πρόβλημα απόφασης λοιπόν, που, κατά την άποψή μας, χρειάζεται περεταίρω έρευνα και μελέτη είναι και η ψήφος. Οι περισσότεροι από εμάς λαμβάνουμε υπόψη διάφορους παράγοντες και στοιχεία, πριν καταλήξουμε στο να εκφράσουμε την πολιτική μας προτίμηση, κυρίως σε μια εκλογική διαδικασία. Χρήζει, επομένως, μελέτης ο τρόπος που συμπεριφέρονται οι πολίτες, πως αντιδρούν σε εκλογικές στρατηγικές, οι προσωπικές προτιμήσεις και τελικά η επιλογή (ψήφος). Με την ανάλυση της ψήφου και των κριτηρίων επιλογής που ενδεχομένως αυτή «κρύβει», θα προέκυπταν χρήσιμα στοιχεία ή και κρίσιμες πληροφορίες που πρέπει να εξαχθούν, προκειμένου να ληφθούν και εν συνεχεία να υποστηριχθούν οι όποιες πολιτικές αποφάσεις.

Κάθε μέλος μιας κοινωνίας που έχει συμπληρώσει το 18^ο έτος της ηλικίας του είναι ένας ψηφοφόρος, όπου κάθε τέσσερα χρόνια καλείται να αποφασίσει είτε στις εθνικές - βουλευτικές εκλογές, είτε στις νομαρχιακές και δημοτικές. Ορισμένοι από εμάς είναι αγρότες, καθηγητές, ελεύθεροι επαγγελματίες, μηχανικοί, δικηγόροι, γιατροί ή βιομήχανοι, βιοτέχνες, μαγαζάτορες, άλλοι είναι μουσικοί και άλλοι τεχνίτες, όλοι όμως είναι ψηφοφόροι. Πώς γίνεται όλοι αυτοί οι άνθρωποι με διαφορετικές αφετηρίες, διαφορετικές εμπειρίες, διαφορετικούς τρόπους ζωής να καταλήγουν σε πανομοιότυπες ή εν πάση περιπτώσει «ομαδοποιημένες» επιλογές; Το συναίσθημα ή το συνειδητό εξουσιάζουν τη σκέψη κατά τη στιγμή της ψηφοφορίας; Είναι η σκέψη της στιγμής ή όταν οδηγούμαστε στην κάλπη, έχουμε ήδη αποφασίσει τι και ποιον θα ψηφίσουμε; Με ποιον τρόπο ή τι επηρεάζει τη λήψη αυτής της απόφασης; Σύμφωνα με μαρτυρίες μεγαλύτερων, παλαιότερα ένα μεγάλο ποσοστό ψήφιζε ποιον (δηλαδή βουλευτή), ενώ η επιλογή του κόμματος ερχόταν σε «δεύτερη μοίρα». Αυτό ενδεχόμενα να εξηγείται από το γεγονός ότι μετεμφυλιακά για ορισμένους ο βουλευτής ήταν και ο προστάτης τους από τις αυθαιρεσίες των οργάνων διοίκησης. Σήμερα λοιπόν τι ισχύει; Και μπορεί όλοι λίγο ή πολύ στις τοπικές μας κοινωνίες και συναναστροφές να έχουμε καταλήξει σε συμπεράσματα, άλλοτε γενικόλογα και άλλοτε πιο εμπεριστατωμένα, τι συμβαίνει, όμως, όταν πρέπει να ληφθεί η απόφαση της ψήφου και να γίνει η τελική επιλογή; Και πως αυτό αποτυπώνεται επιστημονικά, μεθοδολογικά;

1.1 Ο ρόλος των δημοσκοπικών μεθόδων

Εύλογα θα απαντήσει κάποιος «μα αυτό ήδη γίνεται». Υπάρχουν οι δημοσκοπικές μέθοδοι, οι μετρήσεις της κοινής γνώμης, τα γνωστά σε όλους γκάλοπ (Gallup), που αποτυπώνουν και καταγράφουν την άποψη του πολίτη – αποφασίζοντα. Είναι όμως έτσι; Η εκλογική συμπεριφορά έχει ταυτιστεί μονομερώς στη χώρα μας τα τελευταία χρόνια με την ανάπτυξη των πολιτικών ερευνών γνώμης και των δημοσκοπήσεων. Το γεγονός αυτό, όμως, έχει συμβάλλει σε μια σοβαρή «επιστημονική στρέβλωση», που συνίσταται στην κυριαρχία του εμπειρισμού, στην ανάδειξη «δημοσιογραφικών» κριτηρίων στην πολιτική ανάλυση, στην προτεραιότητα της «επικοινωνίας» πάνω στη θεωρία. Η «στρέβλωση» αυτή αδικεί βεβαίως και το εργαλείο των πολιτικών ερευνών γνώμης και των δημοσκοπήσεων, αφού τις αποσυνδέει από το θεωρητικό πεδίο της πολιτικής επιστήμης για να τις καταστήσει «επικοινωνιακό» προϊόν. Η Nonna Mayer (1992) σε τρία άρθρα της, που αν και έχουν γραφτεί σε διαφορετικές εποχές και συγκυρίες, συνθέτει μια συνεκτική και πλήρη παρουσίαση των θεωρητικών τάσεων και προβληματισμών που έχουν αναπτυχθεί στον κλάδο της «εκλογικής συμπεριφοράς».

Στα «Μοντέλα ανάλυσης της εκλογικής συμπεριφοράς»¹, παρουσιάζει με αναλυτικό αλλά και κριτικό τρόπο τα διαδοχικά θεωρητικά ρεύματα και τις τάσεις που αναπτύχθηκαν στο εσωτερικό της εκλογικής κοινωνιολογίας, ξεκινώντας από τα οικολογικά μοντέλα της εκλογικής γεωγραφίας, της ποσοτικής οικολογίας και της ανάλυσης συνθηκών και φτάνοντας στα κοινωνικο-ψυχολογικά μοντέλα του κοινωνικού ντετερμινισμού του Lazarsfeld, της κομματικής ταύτισης της Σχολής του Μίσιγκαν και της ορθολογικής επιλογής. Η Mayer (1992) επιμένει στη θέση ότι η μελέτη της σύγχρονης εκλογικής συμπεριφοράς είναι κατά κανόνα μια πολυσύνθετη διαδικασία, η οποία απαιτεί ουσιαστικά μια σύνθεση όλων των ιστορικών θεωρητικών τάσεων της εκλογικής κοινωνιολογίας, λαμβάνοντας, βεβαίως, υπόψη τις κοινωνικές και πολιτικές ιδιαιτερότητες και παραδόσεις, καθώς και την ιστορική συγκυρία κάθε χώρας. Η ιστορία της εκλογικής συμπεριφοράς δείχνει ότι τα διαφορετικά θεωρητικά - επιστημονικά ρεύματα εμφανίζονταν και αναπτύσσονταν κάθε φορά που οι προηγούμενες κοινωνικές, ιδεολογικές, ταξικές και πολιτισμικές συνθήκες μεταλλάσσονταν και τα «παλαιά» εργαλεία δεν ήταν ικανά να εξηγήσουν από μόνα τους τη διαδικασία της πολιτικής επιλογής. Έτσι, για παράδειγμα, η Σχολή του Μίσιγκαν «διόρθωσε» και «συμπλήρωσε» τον Lazarsfeld, ενώ με τον ίδιο τρόπο οι θεωρητικοί της ορθολογικής επιλογής αμφισβήτησαν αργότερα το δικό της υπόδειγμα της «κομματικής ταύτισης».

Στο δεύτερο άρθρο της, «Τι απομένει από τις βαριές μεταβλητές»², η ανάλυση εμβαθύνει και επικαιροποιεί τους προβληματισμούς σχετικά με το αν οι «κλασικοί» πυλώνες της σύγχρονης εκλογικής συμπεριφοράς, δηλαδή η κοινωνική τάξη και το θρήσκευμα, απότοκοι των διαιρετικών τομών από τις οποίες προέκυψαν τα σύγχρονα κομματικά συστήματα, εξακολουθούν να αποτελούν προσδιοριστικούς παράγοντες

¹ Nonna Mayer, Pascal Perrineau. (1992). *“Les comportements politiques”*, Armand Colin

² Nonna Mayer, Daniel Boy. (1997). *“L’électeur a ses raisons”*, Presses de Sciences Po

της ψήφου στις σύγχρονες δημοκρατίες. Η Mayer δείχνει ότι οι δύο αυτές παράμετροι εξακολουθούν να προσδιορίζουν την εκλογική συμπεριφορά, ωστόσο όχι με τον απόλυτο τρόπο των περασμένων δεκαετιών. Οι αλλαγές στην κοινωνική και οικονομική δομή έχουν τροποποιήσει σημαντικά τους τρόπους με τους οποίους επιδρούν στην ψήφο, ενώ παράλληλα αρχίζουν να κάνουν την εμφάνισή τους και σοβαρές εσωτερικές αντιφάσεις στους πυλώνες αυτούς.

Στο τρίτο άρθρο, τέλος, με τίτλο «Οι μεταμορφώσεις της ταξικής ψήφου»³, αναλύεται ιδιαίτερα το φαινόμενο της «ταξικής ψήφου» και αποδεικνύεται ότι αυτό που παρατηρείται στα σύγχρονα κομματικά συστήματα δεν είναι η εξάλειψή της, όπως υποστηρίζουν πολλοί ερευνητές, αλλά οι μεταμορφώσεις της κυρίως ως προς τις πολιτικές εκπροσωπήσεις. Η «ταξική ψήφος» δεν αφορά πλέον, κατά κύριο λόγο, τα κόμματα της αριστεράς, όπως ήθελε η παράδοση της εκλογικής κοινωνιολογίας, αλλά εκφράζεται πλέον είτε μέσω κομματικών μορφωμάτων, όπως είναι η νέα άκρα δεξιά στην Ευρώπη, είτε μέσω αντιπολιτικών συμπεριφορών (π.χ. αποχή από τις εκλογικές διαδικασίες). Ταυτόχρονα, διαμορφώνονται νέες ιδιόμορφες «ταξικές» σχέσεις στην πολιτική αντιπροσώπευση, όπως για παράδειγμα η εκπροσώπηση των μισθωτών στρωμάτων του δημόσιου τομέα στα κόμματα της αριστεράς, αλλά και η πολιτική δυσπιστία των μισθωτών στρωμάτων του ιδιωτικού τομέα απέναντι στα κόμματα της αριστεράς.

Τα τρία αυτά κείμενα, αν και γραμμένα σε διαφορετικές χρονικές στιγμές, παρουσιάζουν μια εντυπωσιακή μεταξύ τους συνοχή. Το πρώτο κείμενο διασχίζει οριζόντια τις θεωρητικές προσεγγίσεις και τα ρεύματα της εκλογικής ανάλυσης, το δεύτερο και το τρίτο εμβαθύνουν περισσότερο σε συγκεκριμένα «ανοικτά» ζητήματα της σύγχρονης πολιτικής συμπεριφοράς, διευρύνοντας το καθένα τους προβληματισμούς του άλλου. Η ανάλυση ανά περιόδους που επιχειρείται, αποτελεί στην ουσία ένα ταξίδι στην ιστορία της πολιτικής και εκλογικής ανάλυσης. Συνδέει τις φάσεις ανάπτυξης της τελευταίας με τον κοινωνικό, πολιτιστικό και επιστημονικό περίγυρο κάθε εποχής και κάθε χώρας, αναδεικνύει τις θεμελιακές συνεισφορές, αλλά και τα όρια κάθε ρεύματος και τάσης, επιμένει στην απαραίτητη διεπιστημονικότητα πάνω στην οποία θεμελιώνεται η επιστήμη της εκλογικής συμπεριφοράς: ιστορικό πλαίσιο και κοινωνικές αντιθέσεις, οικονομική συγκυρία, ταξική και κοινωνική διαστρωμάτωση, κομματικό σύστημα και κοινωνικές διαιρετικές τομές, εμπειρική κοινωνική έρευνα, στατιστική ανάλυση. Διασχίζοντας τα διαδοχικά ρεύματα και τάσεις στην εκλογική – πολιτική ανάλυση, η Nonna Mayer παρουσιάζει ταυτόχρονα τις αλλαγές στην ταξική διαστρωμάτωση των σύγχρονων κοινωνιών, τις ιδεολογικές επιπτώσεις που αυτή επιφέρει, τις μεταμορφώσεις και μεταλλάξεις των πολιτικών εκπροσωπήσεων. Το σημαντικότερο, ωστόσο, είναι ένα στοιχείο που υπονοείται στην ανάλυση: η συγκρότηση κάθε νέου θεωρητικού, επιστημονικού ή ερευνητικού υποδείγματος στο χώρο της εκλογικής συμπεριφοράς παραπέμπει, στην πραγματικότητα, σε μια αλλαγή της μορφής του κομματικού φαινομένου. Με άλλα λόγια, δεν υπάρχει μελέτη της εκλογικής συμπεριφοράς χωρίς θεωρία για τα πολιτικά

³ Nonna Mayer, Bruno Cautrès. (2004). *“Le nouveau désordre électoral”*, Presses de Sciences Po

κόμματα. Το «κοινωνιολογικό» υπόδειγμα του Lazarsfeld (1948) ή η θεωρία της «κομματικής ταύτισης» για παράδειγμα, παραπέμπουν σε μια μορφή κομμάτων με «καθαρές» και ανταγωνιστικές μεταξύ τους σχέσεις εκπροσώπησης, αλλά και με «καθαρά» και ανταγωνιστικά μεταξύ τους πολιτικά-προγραμματικά σχέδια. Οι πολλαπλές θεωρίες της ορθολογικής επιλογής με τη σειρά τους παραπέμπουν σε μια μορφή κομμάτων με «συγκλίνοντα» μεταξύ τους ιδεολογικά, κοινωνικά και προγραμματικά χαρακτηριστικά. Καθένας από τους δύο παραδειγματικούς αυτούς τύπους κομμάτων παραπέμπει σε διαφορετική ιστορική περίοδο, σε διαφορετική μορφή κράτους και πολιτικού συστήματος, σε διαφορετική μορφή κοινωνικών αντιθέσεων, σε διαφορετική σχέση οικονομίας / πολιτικής. Γι' αυτόν ακριβώς το λόγο ο χώρος της μελέτης της εκλογικής συμπεριφοράς δεν περιορίζεται απλώς και μόνο στις μεθόδους της εμπειρικής πολιτικής – κοινωνικής έρευνας, όπως θα ήθελε ο διαρκώς διογκούμενος εμπειρισμός των ημερών μας, αλλά αποτελεί ένα συστατικό κομμάτι της μελέτης του πολιτικού και κρατικού φαινομένου.

1.2 Οι στόχοι της διπλωματικής εργασίας

Στην παρούσα διπλωματική εργασία θα προσπαθήσουμε, λοιπόν, να κάνουμε την αρχή και να απαντήσουμε σε ορισμένα ερωτήματα και ερωτηματικά που δημιουργούνται σε σχέση με την ανάλυση της συμπεριφοράς των πολιτών – ψηφοφόρων, σήμερα. Πώς ή γιατί ο ψηφοφόρος συμπεριφέρεται όπως συμπεριφέρεται; Πώς επιλέγει να ψηφίσει το τάδε ψηφοδέλτιο και το δείνα υποψήφιο; Ποια στοιχεία οδηγούν σε ταξινομήσεις και ομαδοποιήσεις τους ψηφοφόρους – αποφασίζοντας; Για να απαντήσουμε στα ερωτήματα αυτά, αλλά πολύ περισσότερο για να καταλήξουμε σε χρήσιμα συμπεράσματα που θα βοηθήσουν στη λήψη ή και υποστήριξη πολιτικών αποφάσεων, δεν αρκεί μόνο να περιγράψουμε και να προβλέψουμε, χρειάζεται να κατανοήσουμε και να εξηγήσουμε. Θα αναζητήσουμε, λοιπόν, τα συστήματα-μεθόδους ερευνών αγοράς και ανάλυσης της συμπεριφοράς των πολιτών κατά την εκλογική διαδικασία, αναδεικνύοντας τα κριτήρια – κλίμακες και την ανάλυση προτιμήσεων – κριτήρια επιλογής (ποιοτικά – ποσοτικά) και θα καταλήξουμε σε μία προτεινόμενη μεθοδολογία ποιοτικής ανάλυσης συμπεριφοράς των ψηφοφόρων. Θεωρούμε, λοιπόν, ότι η έρευνα που επιχειρούμε και που αναλύουμε στις επόμενες σελίδες, δείχνει ότι η εφαρμογή της πολυκριτήριας ανάλυσης και η μεθοδολογία της Εξόρυξης Δεδομένων, μας επιτρέπει να κατανοήσουμε καλύτερα τη συμπεριφορά των ατόμων και συνεπώς να πάρουμε ορθολογικότερες αποφάσεις.

ΚΕΦΑΛΑΙΟ 2:

Μεθοδολογία ερευνών αγοράς / κοινής γνώμης

2.1 Έρευνες αγοράς, έρευνες κοινής γνώμης: Μια μικρή αναδρομή

Οι δημοσκοπικές έρευνες μπορούν να θεωρηθούν σαν ένα υποσύνολο των μεθόδων Έρευνας Αγοράς, οι οποίες με τη σειρά τους είναι υποσύνολο των μεθόδων Έρευνας Μάρκετινγκ. Οι μέθοδοι Έρευνας Μάρκετινγκ, πέραν των Δημοσκοπικών Ερευνών περιλαμβάνουν και Ερευνητικές – Μεθοδολογικές Προσεγγίσεις Πειραμάτων, Προσομοιώσεων, Μεθόδων Παρατήρησης και Αναλύσεων Περιεχομένου (observational and content analysis methods), Συλλογής και Επεξεργασίας Δεδομένων Σάρωσης (scanning data), Ανάλυσης Εσωτερικών Δεδομένων (internal data analysis methods), Συλλογής και Επεξεργασίας Δεδομένων Εξειδικευμένων Συσκευών, Πληροφοριακά Συστήματα Μάρκετινγκ, Γεωγραφικά Συστήματα Πληροφοριών Μάρκετινγκ, Ποιοτικές Έρευνες Συνεντεύξεων Βάθους και Ομάδων Εστίασης (depth interviews – focus groups), Εξόρυξη Δεδομένων, κ.α.

Η ανάπτυξη των πολιτικών δημοσκοπήσεων, άρρηκτα συνδεδεμένη με τη γενικότερη ανάπτυξη της εμπειρικής κοινωνικής έρευνας, διαθέτει ήδη μια μακρά ιστορία, αφετηρίες της οποίας αποτελούν αφενός το εύλογο ενδιαφέρον για τη γνώση των κοινωνικών προβλημάτων - απαραίτητο στοιχείο για την πρόοδο της κοινωνίας - κι αφετέρου η σταθερή επιθυμία των κυβερνώντων να γνωρίσουν τις απόψεις και τις τάσεις των πληθυσμών που εξουσιάζουν - προϋπόθεση για τον έλεγχο της συμπεριφοράς των «επικίνδυνων τάξεων». Η διπλή αυτή αφετηρία των πολιτικών και κοινωνικών ερευνών προσδιορίζει εξ αρχής και την καθοριστική τους αμφισημία, την οποία εικονογραφεί η ιστορία της ανάπτυξης των πολιτικών δημοσκοπήσεων στην Ελλάδα.

Όπως αναφέρει, ο κ. Νικολακόπουλος Ηλίας, αναπληρωτής καθηγητής του πανεπιστημίου Αθηνών και πολιτικός αναλυτής, στην εισήγησή του με τίτλο « Οι Πολιτικές Δημοσκοπήσεις στην Ελλάδα: Ιστορία, Προβλήματα, Προοπτικές» (1^ο συνέδριο ΣΕΔΕΑ, 2003), η πρώτη σχετική εμπειρία στον τομέα αυτό ανάγεται στην ταραγμένη περίοδο του Εμφυλίου Πολέμου και σχετίζεται με την παρουσία της «Συμμαχικής αποστολής για την παρακολούθηση των Ελληνικών εκλογών» (Allied Mission to Observe the Greek Elections – AMFOGE), η οποία ήρθε στην Ελλάδα για να επιβλέψει τη διεξαγωγή των πρώτων μεταπολεμικών βουλευτικών εκλογών της 31^{ης} Μαρτίου 1946. Η Συμμαχική Αποστολή, πραγματοποίησε, μεταξύ άλλων, δειγματοληπτική έρευνα την επομένη των εκλογών, ένα είδος, δηλαδή, πρώιμου exit poll, με προσωπικές συνεντεύξεις 1.345 ατόμων, επιλεγμένα από τους εκλογικούς καταλόγους 113 εκλογικών τμημάτων, για να ελέγξει την αξιοπιστία του εκλογικού αποτελέσματος, καθώς και το μέγεθος της «πολιτικής αποχής». Η παρουσίαση από την AMFOGE των αποτελεσμάτων της πρώτης αυτής πολιτικής δημοσκόπησης, η οποία έγινε με περιληπτική ανακοίνωση στις εφημερίδες λίγες ημέρες μετά τις

εκλογές, - το θλιβερά διάσημο 9,3% ως μέγεθος της «πολιτικής αποχής» - αποτελεί κορυφαία ένδειξη στατιστικής αλαζονείας, χαρακτηριστικό παράδειγμα για την απόλυτα χειραγωγική χρησιμοποίηση που μπορεί να γίνει στα ευρήματα μιας πολιτικής δημοσκόπησης.

Πέρασαν δέκα περίπου χρόνια έως ότου επαναληφθεί στην Ελλάδα μια εθνικού επιπέδου πολιτική δημοσκόπηση. Στις αρχές του 1957, με δεδομένο το έντονα αντιδυτικό πνεύμα που είχε διαμορφωθεί λόγω του Κυπριακού, το State Department αποφάσισε να συμπεριλάβει την Ελλάδα στα πλαίσια μιας ευρύτερης συγκριτικής ευρωπαϊκής έρευνας - τη συνολική ευθύνη της οποίας είχε αναλάβει ο George Gallup – με αντικείμενο τις στάσεις του κοινού έναντι των ΗΠΑ, της Βορειοατλαντικής Συμμαχίας, αλλά και γενικότερα ζητήματα εξωτερικής πολιτικής και διεθνών πολιτιστικών επιρροών. Τη διεξαγωγή της έρευνας, που πραγματοποιήθηκε τον Μάιο – Ιούνιο 1957 και επαναλήφθηκε τον επόμενο χρόνο, ανέλαβε το Ινστιτούτο Ερευνών Επικοινωνίας (με πρωτεργάτη τον Χρυσόστομο Παπαδόπουλο), το οποίο είχε δημιουργηθεί το 1952 και διέθετε ήδη (το 1957) μια σημαντική εμπειρία συνεργασίας με την USIS, έχοντας πραγματοποιήσει σειρά ερευνών σε ειδικά κοινά (π.χ. φοιτητές) για κοινωνικά και πολιτιστικά θέματα. Με βάση την εμπειρία αυτή το Ινστιτούτο Ερευνών Επικοινωνίας (ΙΕΕ), αμέσως μετά την επικράτηση της Ένωσης Κέντρου το 1963, ανέλαβε, με ανάθεση της Αμερικανικής Πρεσβείας, τη διεξαγωγή μηνιαίων πολιτικών ερευνών στην περιφέρεια της πρωτεύουσας, με αντικείμενο τις πολιτικές στάσεις και προτιμήσεις του κοινού. Η πρακτική αυτή συνεχίστηκε ανελλιπώς μέχρι τη Δικτατορία, ενώ, ακόμη και κατά τη διάρκειά της, επιχειρήθηκε ευκαιριακά η διεξαγωγή ερευνών, κυρίως για θέματα διεθνούς πολιτικής.

Οι εμπειρίες που αναφέρθηκαν προηγουμένως, προσφέρουν ορισμένες ενδείξεις για την προϊστορία των πολιτικών δημοσκοπήσεων στην Ελλάδα, η συστηματική μελέτη της οποίας, καθώς και η συγκρότηση ενός συνεκτικού αρχείου, παραμένει ανοικτό ερευνητικό πεδίο, αλλά και προϋπόθεση για την επιστημονική και επαγγελματική αυτογνωσία του κλάδου. Αντίστοιχα, ανοικτό ερευνητικό πεδίο αποτελεί όμως (δεν χρειάζεται) και η ουσιαστική ανάπτυξη των πολιτικών δημοσκοπήσεων, αμέσως μετά την πτώση της δικτατορίας. Ήδη στις παραμονές των εκλογών του Νοεμβρίου 1974, η ICAP πραγματοποίησε, με ανάθεση από το Υπουργείο Τύπου, μια μεγάλη (αλλά απόρρητη) έρευνα, με δείγμα 8.000 ατόμων σε τρία κύματα, η οποία και κατέγραψε με σχετική πιστότητα το προεκλογικό τοπίο. Ακολούθησε, το Μάρτιο του 1976, η πρώτη, δημόσια, εμφάνιση των πολιτικών δημοσκοπήσεων, με τη δημοσίευση από την Ελευθεροτυπία, σε πολλές συνέχειες και με ιδιαίτερη προβολή, μιας εκτεταμένης έρευνας, την οποία είχε πραγματοποιήσει το ΙΕΕ. Ο αντίκτυπος που είχε η σχετική δημοσίευση οδήγησε σε μια «δημοσκοπική έκρηξη», την περίοδο που προηγήθηκε των εκλογών της 20^{ης} Νοεμβρίου 1977, καθώς και στις πρώτες προσπάθειες για πρόβλεψη του εκλογικού αποτελέσματος. Πρέπει, όμως, να σημειωθεί ότι αυτή η σταδιακή καθιέρωση των πολιτικών δημοσκοπήσεων, με χαρακτηριστικότερο παράδειγμα τη διεξαγωγή τακτικών πανελλαδικών ερευνών για λογαριασμό του Κέντρου Πολιτικής Ερεύνης και Επιμορφώσεως (ΚΠΕΕ), δεν

συνδυάστηκε, την εποχή εκείνη, από αντίστοιχη απήχηση και προβολή από τα ΜΜΕ. Τόσο για τις ηγεσίες των πολιτικών κομμάτων, όσο και για τους εκπροσώπους των ΜΜΕ, οι επιφυλάξεις και η δυσπιστία για τις δημοσκοπήσεις εξακολουθούσαν να αποτελούν το γενικό κανόνα. Ακόμα και το 1984, στις παραμονές των Ευρωεκλογών, τα αποτελέσματα των προεκλογικών ερευνών, δημοσιευμένων και αδημοσίευστων, αγνοήθηκαν επιδεικτικά τόσο από τα δύο μεγάλα κόμματα, όσο και από την Προεδρία της Δημοκρατίας.

Το κλίμα είχε αρχίσει πάντως σταδιακά να μεταβάλλεται ήδη από τις αρχές της δεκαετίας του '80: η διεξαγωγή τακτικών ερευνών κοινής γνώμης από το Ευρωβαρόμετρο, η σχετική επιτυχία των δημοσκοπήσεων που έγιναν πριν από τις Ευρωεκλογές του 1984, η απαρχή επιστημονικών ερευνών πολιτικής κουλτούρας και συμπεριφοράς (κυρίως από το ΕΚΚΕ) καθώς και τα πρώτα βήματα απεξάρτησης των ΜΜΕ από την πλήρη συνταύτιση με πολιτικά κόμματα, αποτέλεσαν τους βασικούς παράγοντες, που συνέβαλαν στη διαμόρφωση ενός νέου ερευνητικού τοπίου, γεγονός που οδήγησε, σε πολύ σύντομο χρονικό διάστημα, σε άνθιση των πολιτικών δημοσκοπήσεων. Ιδιαίτερα από το 1989 έως το 1990 και μετά, οι πολιτικές δημοσκοπήσεις αποτελούν πλέον σταθερό σημείο αναφοράς όλων των πολιτικών συζητήσεων και η συστηματική τους χρήση γενικευμένο κανόνα για τα πολιτικά κόμματα. Ιδιαίτερη μνεία στο σημείο αυτό πρέπει να γίνει και στην τηλεοπτική προβολή των πολιτικών δημοσκοπήσεων, η οποία ακολούθησε την έκρηξη της ιδιωτικής τηλεόρασης τη δεκαετία του '90. Οι απαιτήσεις του τηλεοπτικού χρόνου συνέβαλαν καθοριστικά στη υιοθέτηση των δημοσκοπήσεων εξόδου (exit poll), οι οποίες πραγματοποιούνται την ημέρα των εκλογών, με άμεσο στόχο την έγκαιρη πρόβλεψη του εκλογικού αποτελέσματος. Η πρώτη σχετική έρευνα πραγματοποιήθηκε από το Mega Channel στις Ευρωεκλογές του Ιουνίου 1994 και στη συνέχεια επαναλήφθηκε στις Δημοτικές και Νομαρχιακές Εκλογές του Οκτωβρίου 1994 (από το Mega και τον ANT-1), για να γενικευθεί ακολούθως στις Βουλευτικές Εκλογές του Σεπτεμβρίου 1996.

Με τη σύντομη αναφορά μας στην ιστορία των πολιτικών δημοσκοπήσεων στην Ελλάδα, αναδεικνύονται ορισμένα βασικά χαρακτηριστικά, τα οποία και σήμερα ακόμη τις διακρίνουν. Το πρώτο αφορά τη διαπλοκή (ή έστω την υποψία περί διαπλοκής) των ίδιων των ερευνών με τη χρήση ή την κατάχρησή που τους γίνεται και το δεύτερο την απουσία γενικά καθιερωμένων και αποδεκτών κριτηρίων για την αξιολόγηση τόσο των ίδιων των ερευνών, αλλά κυρίως των δημοσιεύσεων που βασίζονται σε αυτές. Πρόκειται για κριτήρια τα οποία δεν μπορούν να προκύψουν παρά μόνον από μία σωρευμένη εμπειρία, η οποία να αποτελεί αντικείμενο συνεχούς και ανοιχτής επιστημονικής συζήτησης, καθώς και από την καθιέρωση ενός θεσμικού πλαισίου, το οποίο να διασφαλίζει την κοινή γνώμη από την όποια, πραγματική ή φανταστική, χειραγωγική χρήση των δημοσκοπήσεων.

Η αντιμετώπιση και η κριτική των πολιτικών δημοσκοπήσεων δεν είναι απαραίτητο να ταλαντεύεται ανάμεσα στις ακραίες εκδοχές είτε της θεοποίησής τους – ως σύγχρονη έκφραση της άμεσης δημοκρατίας – είτε της δαιμονοποίησής τους – ως χειραγωγικό μηχανισμό ουσιαστικής κατάργησης της δημοκρατίας. Όπως όλα τα

κοινωνικά φαινόμενα, έτσι και οι πολιτικές δημοσκοπήσεις έχουν αναμφίβολα μία διπλή και αντιφατική όψη: επιτρέπουν τη συστηματικότερη καταγραφή και γνώση των πολιτικών φαινομένων και ταυτόχρονα είναι δυνατόν να χρησιμοποιηθούν για τη διαμόρφωση και επιβολή ορισμένων απόψεων. Γι' αυτό και είναι απαραίτητο η υλοποίηση και η δημοσιοποίησή τους να διέπονται από ένα, κατά το δυνατόν, διαφανές θεσμικό πλαίσιο, το οποίο να διαμορφωθεί με ιδιαίτερη σύνεση και με την ευρύτερη συναίνεση ανάμεσα στους παραγωγούς και τους χρήστες των πολιτικών δημοσκοπήσεων. Και όπως είναι φυσικό, για κάθε μεθοδολογία που θέλει να ακολουθεί, έστω και εν μέρει, τους βασικούς κανόνες της επιστήμης, έτσι και οι πολιτικές δημοσκοπήσεις είναι αναπόφευκτο, σε ορισμένες περιπτώσεις, να κάνουν λάθος.

Στη μελέτη που πραγματοποιούμε, από την αναζήτησή μας για τη διαδικασία και τον τρόπο διάρθρωσης των ερευνών αγοράς και των δημοσκοπικών ερευνών, προέκυψε ότι οι περισσότερες εταιρείες του χώρου ακολουθούν τις ίδιες ή παρόμοιες μεθόδους, προκειμένου να συλλέξουν και να επεξεργαστούν τις απαντήσεις / δεδομένα που προκύπτουν από τις έρευνές τους. Έτσι, στη συνέχεια, θα παρουσιάσουμε τις μεθόδους – λύσεις που ακολουθούν και θα αναπτύξουμε ορισμένες έννοιες που παίζουν σημαντικό ρόλο στη συλλογή και ανάλυση των απαντήσεων.

2.2 Ποσοτικές έρευνες

Η διαδικασία διεξαγωγής μιας ποσοτικής κοινωνικής έρευνας ακολουθεί δύο διακριτά στάδια. Κατά το πρώτο, το στάδιο σχεδιασμού, διατυπώνονται οι σκοποί της έρευνας και προσδιορίζονται τα ζητούμενα βάσει υποθέσεων εργασίας, ακολούθως δε επιλέγεται η μέθοδος πραγματοποίησής της και σχεδιάζεται η βήμα προς βήμα υλοποίησή της. Κατά το δεύτερο, το στάδιο υλοποίησης, συλλέγονται τα απαραίτητα στοιχεία, ακολουθεί η επεξεργασία και ανάλυση των δεδομένων που προκύπτουν και γίνεται η σύνθεσή τους και διατύπωση των σχετικών συμπερασμάτων.

Τα βήματα σε μία έρευνα:

1. Καθορισμός των στόχων της έρευνας - Τι θέλουμε να μάθουμε
2. Δημιουργία του ερωτηματολογίου μας - Τι θα ρωτήσουμε
3. Επιλογή μεθοδολογίας συνέντευξης - Πώς θα πάρουμε συνέντευξη
4. Καθορισμός του δείγματός μας – Από ποιους θα πάρουμε συνέντευξη
5. Προ-εξέταση (πιλοτική εφαρμογή) του ερωτηματολογίου, εάν είναι πρακτικό κ.λπ. - Εξετάζουμε τις ερωτήσεις
6. Πραγματοποίηση συνεντεύξεων και εισαγωγή των στοιχείων - Υποβάλλουμε τις ερωτήσεις
7. Ανάλυση των στοιχείων - Συντάσσουμε τις εκθέσεις / αναφορές

2.2.1 Σχεδιασμός της Έρευνας

Η γνώση του “τι θέλει ο πελάτης” είναι ο βασικός παράγοντας επιτυχίας σε οποιοδήποτε τύπο επιχείρησης. Τα μέσα ενημέρωσης, οι κρατικές υπηρεσίες και οι υποψήφιοι πολιτικοί πρέπει να ξέρουν τι σκέφτεται το κοινό. Οι συνεταιρισμοί πρέπει να ξέρουν τι θέλουν τα μέλη τους. Οι μεγάλες επιχειρήσεις πρέπει να λαμβάνουν υπόψη τη γνώμη των υπαλλήλων τους. Ο καλύτερος τρόπος για να βρεθούν αυτές οι πληροφορίες, είναι να πραγματοποιηθεί μια έρευνα.

Ο προσδιορισμός και η διατύπωση των ζητούμενων μιας ποσοτικής κοινωνικής έρευνας στηρίζονται σε κάποιες βασικές υποθέσεις εργασίας, των οποίων η διερεύνηση είναι επιβεβλημένη, ούτως ώστε να διαπιστωθεί κατά πόσον αυτές ισχύουν και να αποτυπωθεί η υφιστάμενη κατάσταση στο συγκεκριμένο πλαίσιο αναφοράς.

2.2.2 Κατάρτιση ερωτηματολογίων και τρόποι συνέντευξης

Στις ποσοτικές έρευνες χρησιμοποιείται ευρύτατα η συμπλήρωση ερωτηματολογίων, στα οποία αποτυπώνεται το περιεχόμενο των προσωπικών συνεντεύξεων που λαμβάνονται επί τούτου. Η συνέντευξη είναι η τεχνική που έχει σκοπό να οργανώσει μία σχέση προφορικής επικοινωνίας ανάμεσα σε δύο πρόσωπα, τον συνεντευκτή και τον ερωτώμενο, έτσι ώστε να επιτρέψει στον πρώτο τη συλλογή ορισμένων πληροφοριών απ’ τον δεύτερο πάνω σ’ ένα συγκεκριμένο αντικείμενο. Η έρευνα που γίνεται με τη μέθοδο αυτή, πρέπει να έχει προετοιμαστεί με κάθε λεπτομέρεια, μια και ο συνεντευκτής οφείλει, στο ελάχιστο χρονικό διάστημα, να αποκτήσει πολυάριθμες και τις πιο σημαντικές πληροφορίες.

Αυτή η προετοιμασία οδηγεί στην κατάστρωση ερωτηματολογίου από τον ερευνητή, ο οποίος αναλαμβάνει:

- α) Να μετατρέψει τους σκοπούς που επιδιώκει η έρευνα σε επί μέρους ερωτήσεις.
- β) Να προσαρμόσει το ερωτηματολόγιο στα πρόσωπα με τα οποία θα γίνει η συνέντευξη.
- γ) Να ενημερώσει τους συνεντευκτές γι’ αυτά έτσι ώστε να μπορέσουν να εκθέσουν με σαφήνεια τις ερωτήσεις στα πρόσωπα που θα υποβληθούν στη συνέντευξη και να προδιαθέσει το ερωτώμενο πρόσωπο να μεταδώσει αυθόρμητα τις πληροφορίες που περιμένουν από αυτό.

2.2.2.1 Γενικές παρατηρήσεις για την κατάρτιση του ερωτηματολογίου

Το ερωτηματολόγιο θα πρέπει να είναι σχετικά απλό και κατανοητό. Εάν παρουσιάσουμε ένα ερωτηματολόγιο 20-σελίδων οι περισσότεροι πιθανοί

ερωτώμενοι θα σταματήσουν πριν καν αρχίσουν. Επίσης κατά τη διαδικασία σύνταξης του ερωτηματολογίου θα πρέπει να αναρωτηθούμε τι θα κάνουμε με τις πληροφορίες από κάθε ερώτηση / απάντηση και θα πρέπει να προσέξουμε η σύνταξη των ερωτήσεων να είναι “ουδέτερη”, να μην καθοδηγεί, δηλαδή, τον ερωτώμενο σε μία προφανή απάντηση.

2.2.2.2 Τύποι ερώτησης

Οι ερευνητές χρησιμοποιούν συνήθως δύο βασικούς τύπους ερωτήσεων: τις λεγόμενες ανοιχτές ή κλειστές. Γενικά οι κλειστές ερωτήσεις είναι πιο αξιόπιστες και είναι ευκολότερο να απαντηθούν, να κωδικοποιηθούν και να αναλυθούν. Από την άλλη, οι ανοιχτές ερωτήσεις είναι χρήσιμες για να καταγράφουν σκέψεις, για έρευνες που στοχεύουν σε άτομα και όχι σε ομάδες, αλλά είναι δύσκολο να συγκριθούν και να παρουσιαστούν. Κωδικοποιώντας τη χρήση των ανοιχτών και κλειστών ερωτήσεων καταλήγουμε στο ότι επιλέγουμε τις ανοιχτού τύπου ερωτήσεις όταν:

- Θέλουμε να διατυπώσουμε μόνο την ερώτηση
- Είναι σημαντικές οι διατυπώσεις των ερωτώμενων
- Οι ερωτώμενοι μπορούν και θέλουν να απαντήσουν με δικά τους λόγια
- Μπορούμε να αναλύσουμε σχόλια και σπάνιες απαντήσεις
- Θα παρουσιάσουμε ατομικές ή ομαδοποιημένες λεκτικές απαντήσεις

ενώ επιλέγουμε τις κλειστού τύπου ερωτήσεις όταν:

- Θέλουμε ταξινομημένα δεδομένα και γνωρίζουμε τη διάταξη εκ των προτέρων
- Θέλουμε απαντήσεις εκ των προτέρων τοποθετημένες σε μια κλίμακα
- Θέλουμε να επιλέξουν από προκαθορισμένες απαντήσεις
- Προτιμάμε να βασιστούμε σε εκ των προτέρων επιλογές
- Θα παρουσιάσουμε στατιστικά αποτελέσματα

2.2.2.3 Προ-εξέταση του ερωτηματολογίου

Το τελευταίο βήμα στο σχεδιασμό ερωτηματολογίων είναι να το εξετάσουμε σε έναν μικρό αριθμό συνεντευξιαζόμενων, πριν πραγματοποιήσουμε τις κύριες συνεντεύξεις μας. Το ιδανικό θα ήταν να εξετάσουμε την έρευνα στις ίδιες ομάδες ανθρώπων που θα περιλάβουμε και στην κύρια μελέτη. Εάν αυτό δεν είναι δυνατόν, θα πρέπει τουλάχιστον να δοκιμάσουμε το ερωτηματολόγιο σε μερικούς ανθρώπους. Αυτό το είδος “δοκιμαστικής λειτουργίας” μπορεί να αποκαλύψει απρόβλεπτα προβλήματα με τη διατύπωση των ερωτήσεων, την παράβλεψη και μη απάντηση ερωτήσεων κ.λπ..... Μπορεί επίσης να βοηθήσει στο να δούμε, εάν οι συνεντευξιαζόμενοι καταλαβαίνουν τις ερωτήσεις μας και αν, εν τέλει, μας δίνουν χρήσιμες απαντήσεις. Εάν αλλάξουμε οποιαδήποτε ερώτηση μετά από την προ-

εξέταση του ερωτηματολογίου, θα πρέπει να προσέξουμε ούτως ώστε να μην συνδυάσουμε τα αποτελέσματα από την προ-εξέταση με τα αποτελέσματα των κύριων συνεντεύξεων.

2.2.2.4 Μέθοδοι συνέντευξης

Η συνέντευξη στις Κοινωνικές Επιστήμες - κι αυτό την ξεχωρίζει από τη δημοσιογραφική - είναι μια «συστηματοποιημένη» συνέντευξη - συζήτηση με έναν ορισμένο αριθμό χαρακτηριστικών που καθορίζουν την τεχνική της εφαρμογής της. Οι μέθοδοι συνέντευξης, με τα αντίστοιχα πλεονεκτήματα και μειονεκτήματα που παρουσιάζει η καθεμία είναι συνήθως οι εξής:

- **Προσωπικές συνεντεύξεις:** Οι face-to-face (‘πρόσωπο - με - πρόσωπο’) έρευνες αποτελούν τρόπο τινά, την «κλασική» μέθοδο συλλογής δεδομένων στην εμπειρική κοινωνική έρευνα. Πρόκειται για έρευνες που διεξάγονται στα νοικοκυριά της χώρας, όπου αναζητείται το «κατάλληλο», ανάλογα με την έρευνα, άτομο προς διενέργεια της συνέντευξης. Η ραγδαία, τα τελευταία χρόνια, ανάπτυξη των τηλεφωνικών ερευνών μείωσε κατά πολύ τη χρήση της μεθοδολογίας αυτής και περιόρισε την κατάχρησή της, η οποία είχε αυξηθεί σε υπερβολικό βαθμό. Σήμερα οι face-to-face έρευνες χρησιμοποιούνται με περισσότερο μέτρο και όταν το αντικείμενο και ο στόχος της έρευνας το επιτρέπουν. Στην ουσία, **η μεθοδολογία αυτή ταιριάζει σε έρευνες προσδιορισμού συμπεριφοράς** και όχι απλώς σε έρευνες καταγραφής γνώμων.

- **Τηλεφωνικές έρευνες:** Η διεξαγωγή τηλεφωνικών ερευνών γνώμης έχει κερδίσει πλέον και στην Ελλάδα σημαντικό έδαφος ως μέθοδος συλλογής δεδομένων. Η τηλεφωνική μέθοδος συλλογής δεδομένων έχει τρία βασικά πλεονεκτήματα σε σχέση με την face-to-face μέθοδο. Πρώτον, απαιτεί σημαντικά μικρότερο χρόνο πραγματοποίησης, δεύτερον, δίνει καλύτερη γεωγραφική και δημογραφική κάλυψη (διασπορά του δείγματος) και διεισδύει σε νοικοκυριά ή και σε ολόκληρες περιοχές / συνοικίες που η face-to-face (‘πρόσωπο-με-πρόσωπο’) έρευνα είναι σχεδόν απαγορευτική και τρίτον, στο μέτρο που το ποσοστό των νοικοκυριών που δεν διαθέτουν σταθερό τηλέφωνο, είναι πλέον σχεδόν ανύπαρκτο, δεν δημιουργεί μεροληψίες υποαντιπροσώπησης πληθυσμών. Πολλές φορές, βεβαίως, έχει διαπιστωθεί σε τηλεφωνικές έρευνες υπεραντιπροσώπηση ατόμων ανώτερης μόρφωσης και υψηλότερων εισοδημάτων και υποαντιπροσώπηση ατόμων που ανήκουν στις χαμηλότερες κοινωνικοοικονομικές τάξεις ή και μειονότητες.

- **Έρευνες ταχυδρομείου:** Οι έρευνες ταχυδρομείου επιτρέπουν τη συμπλήρωση του ερωτηματολογίου στον ελεύθερο χρόνο του συνεντευξιζόμενου. Όμως, χρειάζονται πολύ περισσότερο χρόνο για να συμπληρωθούν και να αποσταλούν

πίσω, ώστε να ξεκινήσει η ανάλυση των απαντήσεων / δεδομένων. Αυτό βεβαίως δεν ωφελεί σε περιπτώσεις όπου χρειάζονται άμεσα αποτελέσματα, όπως για παράδειγμα στις έρευνες κοινής γνώμης (δημοσκοπικές).

- **Έρευνες ηλεκτρονικού ταχυδρομείου:** Οι έρευνες ηλεκτρονικού ταχυδρομείου είναι και πολύ οικονομικές και πολύ γρήγορες. Οι περισσότεροι άνθρωποι συνήθως χρησιμοποιούν το ηλεκτρονικό τους ταχυδρομείο περισσότερο από μια πλήρη περιήγηση στο Διαδίκτυο. Αυτό κάνει τις έρευνες ηλεκτρονικού ταχυδρομείου καλύτερη επιλογή από μια έρευνα ιστοσελίδας, για μερικές ομάδες του πληθυσμού. Αφ' ετέρου όμως, οι έρευνες ηλεκτρονικού ταχυδρομείου περιορίζονται στα απλά ερωτηματολόγια, ενώ οι έρευνες ιστοσελίδας μπορούν να περιλάβουν πιο σύνθετες λογικές.

- **Έρευνες διαδικτύου (ιστοσελίδας):** Οι έρευνες διαδικτύου κερδίζουν συνεχώς έδαφος. Έχουν σημαντική ταχύτητα, μικρό κόστος και πλεονεκτήματα ευελιξίας, αλλά έχουν και σημαντικούς περιορισμούς δειγματοληψίας (απαντούν μόνο όσοι έχουν internet, πολλές φορές οι ερωτήσεις μπορούν να απαντηθούν παραπάνω από μία φορά κ.λπ.). Αυτοί οι περιορισμοί καθιστούν την επιλογή λογισμικού ιδιαίτερα σημαντική και περιορίζουν και τις ομάδες που μπορούμε να μελετήσουμε αλλά και τη χρησιμοποίηση αυτής της τεχνικής.

Θα πρέπει να επισημανθεί ότι η γνώση των ιδιαίτερων χαρακτηριστικών όσων δεν απαντούν σε μία έρευνα αποτελεί πολύ σημαντική συνιστώσα, ώστε να προσδιοριστεί η ορθή ή όχι γενίκευση των αποτελεσμάτων της έρευνας στον υπό εξέταση πληθυσμό. Οι παραπάνω μέθοδοι συνέντευξης που χρησιμοποιούνται στις ποσοτικές έρευνες, διακρίνονται σε δύο βασικές μορφές, σε δύο βασικές κατηγορίες, που είναι:

Η δομημένη συνέντευξη:

Με τον όρο αυτό εννοούμε τη συνέντευξη εκείνη όπου ο ερωτώμενος προτρέπει στο να απαντήσει σε μια σειρά ερωτήσεων που ο αριθμός, η σειρά και το περιεχόμενο προκαθορίζεται από το έντυπο της συνέντευξης. Οι απαντήσεις καταγράφονται ή λέξη προς λέξη ή κωδικοποιημένες.

Η εντοπισμένη συνέντευξη (Focused Interview):

Ο συνεντευκτής θέτει το γενικό πλαίσιο και εντοπίζει τα σημεία ιδιαίτερου ενδιαφέροντος, ούτως ώστε εκεί να εστιασθεί η ανάπτυξη του θέματος (ημι-κατευθυνόμενη συνέντευξη). Η τεχνική αυτή χρησιμοποιείται κυρίως σε πληροφοριοδότες - κλειδιά, για τους οποίους εκ των προτέρων θεωρούμε πως έχουν ιδιαίτερη γνώση επί του διερευνώμενου θέματος. Μετά την κατάστρωση του κατάλληλου ερωτηματολογίου, γίνεται «πιλοτική έρευνα» (προ - έρευνα) για να προσδιορισθεί η λειτουργικότητα του ερωτηματολογίου και να διαμορφωθεί οριστικά η δομή του. Κατ' αυτή την διαδικασία, σ' ένα βαθμό γίνεται χρήση των τεχνικών της

ποιοτικής προσέγγισης. Εν συνεχεία ακολουθεί η επιλογή δείγματος από το σύνολο του πληθυσμού, στο οποίο θα πραγματοποιηθεί η επιτόπια έρευνα με τη διεξαγωγή συνεντεύξεων μέσω ερωτηματολογίων.

2.2.3 Δειγματοληψία

2.2.3.1 Γενικά

Η μαθηματική δειγματοληψία στηρίζεται στη θεωρία των πιθανοτήτων και τη στατιστική. Και στους δύο αυτούς κλάδους των μαθηματικών η έννοια του «τυχαίου» είναι βασική και αυστηρά ορισμένη.

Επειδή, όπως θα δούμε, το αντιπροσωπευτικό υποσύνολο, το δείγμα, δεν είναι παρά ένα «τυχαίο υποσύνολο» του πληθυσμού, πρέπει να διευκρινίσουμε πως το «τυχαίο» δε συμπίπτει με αυτό που στη καθημερινή ζωή λέμε «στην τύχη». Μόνο οι πίνακες τυχαίων αριθμών, αν χρησιμοποιηθούν σωστά, εγγυώνται την «τυχειότητα» της επιλογής, γιατί είναι φτιαγμένοι με τέτοιο τρόπο ώστε ν' ανταποκρίνονται στο μαθηματικό ορισμό του τυχαίου. Αντίθετα, δηλαδή, από αυτό που θα περίμενε κανείς, η επίτευξη του τυχαίου απαιτεί την πραγματοποίηση μιας πολύ συστηματικής διαδικασίας.

2.2.3.2 Δειγματοληψία και απογραφή

Ο βασικότερος παράγοντας που μας κάνει να επιλέγουμε τη δειγματοληψία αντί της απογραφής, είναι οπωσδήποτε η αδυναμία να πραγματοποιηθεί απογραφή είτε για λόγους αντικειμενικούς είτε για λόγους οικονομίας, μια και το κόστος μιας έρευνας είναι συνήθως ανάλογο του αριθμού των στοιχείων που εξετάζονται. Τα κύρια πλεονεκτήματα της δειγματοληψίας έναντι της απογραφής είναι:

- α) Ταχύτητα. Η απογραφή ενός πολύ μεγάλου πληθυσμού απαιτεί πολύ χρόνο. Αν κάποια στοιχεία επείγει να γίνουν γνωστά, η δειγματοληπτική μέθοδος αποτελεί τη μόνη λύση.
- β) Ευρύτερο πεδίο έρευνας: Ακριβώς επειδή απευθύνεται σε λιγότερα άτομα, υπάρχει η δυνατότητα να συγκεντρωθούν πληροφορίες για πολύ περισσότερα χαρακτηριστικά του ερωτώμενου πληθυσμού.
- γ) Ακρίβεια: Όσο κι αν φαίνεται περίεργο είναι δυνατό με μια δειγματοληπτική έρευνα να έχουμε πιο ακριβή αποτελέσματα. Πράγματι, η περιορισμένη έκταση της επιτρέπει να χρησιμοποιηθεί ειδικευμένο προσωπικό, γεγονός που θα μειώσει τα σφάλματα παρατήρησης.

Μια από τις κρισιμότερες μεθοδολογικές και επιστημονικές προϋποθέσεις για τη διενέργεια μιας ποσοτικής έρευνας είναι λοιπόν ο δειγματοληπτικός σχεδιασμός

της. Ουσιαστικά, η επιλογή και η εκπόνηση δειγματοληπτικών σχεδίων (sampling design) έχουν σημαίνοντα ρόλο στη διαδικασία της έρευνας, ώστε να ανταποκρίνεται στις ανάγκες και τις απαιτήσεις κάθε ερευνητικού προγράμματος που καλείται να υλοποιήσει. Ως δείγμα (sample) σε μια ποσοτική έρευνα, εννοούμε ένα υποσύνολο του πληθυσμού (population) που μας ενδιαφέρει να εξετάσουμε. Το μέγεθος του δείγματος που θα επιλέξουμε, παίζει σημαντικότατο ρόλο στην ακρίβεια και αξιοπιστία των αποτελεσμάτων που θα εκδώσουμε. Επίσης, για να μπορούμε να γενικεύσουμε τα αποτελέσματα που θα προκύψουν από τη μελέτη του δείγματος, θα πρέπει να έχουμε ορίσει σωστά τον υπό μελέτη πληθυσμό, να έχουμε βρει, ή κατασκευάσει ένα όσο το δυνατόν πιο πλήρες δειγματοληπτικό πλαίσιο (sampling frame) και να έχουμε επιλέξει, με την κατάλληλη τεχνική, το δείγμα, ώστε να είναι αντιπροσωπευτικό του πληθυσμού.

Ως πιθανό θεωρητικό δείγμα (probability sample) εννοούμε αυτό στο οποίο σε κάθε μονάδα που επιλέγεται στο δείγμα αντιστοιχεί σε μια γνωστή μη μηδενική - πιθανότητα, με την οποία περιέχεται στο δείγμα. Οι συνηθέστερες τεχνικές επιλογής πιθανοθεωρητικών δειγμάτων είναι η απλή τυχαία δειγματοληψία (simple random sample), η στρωματοποιημένη δειγματοληψία (stratified random sample), η δειγματοληψία κατά συστάδες (cluster sample), η συστηματική δειγματοληψία (systematic random sample), καθώς και συνδυασμοί των παραπάνω τεχνικών. Αναλόγως του συγκεκριμένου προβλήματος κάθε φορά (ειδικότερες παραμέτρους της έρευνας, την ύπαρξη ικανού ή όχι δειγματοληπτικού σχεδίου), επιλέγουμε την καταλληλότερη τεχνική.

Για να πετύχει μια δειγματοληπτική έρευνα πρέπει ο πληθυσμός όχι μόνο να είναι σαφώς ορισμένος, αλλά και αρκετά ομοιογενής. Αν η τιμή του χαρακτηριστικού που μετράμε, παίρνει πολύ μικρές και πολύ μεγάλες τιμές, θα πρέπει οι ακραίες περιπτώσεις ν' απομονωθούν και να μελετηθούν χωριστά, γιατί αλλιώς τα συμπεράσματα μας θα είναι εντελώς λανθασμένα.

Τέλος, αν δεν δοθεί προσοχή στα σφάλματα παρατήρησης, όσο σωστά κι αν έχει επιλεγεί το δείγμα, η προσπάθεια θα πάει χαμένη. Αποδεικνύεται πως τα λάθη παρατήρησης επηρεάζουν περισσότερο μια δειγματοληψία παρά μια απογραφή και μάλιστα πολύ περισσότερο όταν το δείγμα είναι μικρότερο.

2.2.3.3 Μέθοδοι δειγματοληψίας

Θα αναφέρουμε εν συντομία τις μεθόδους δειγματοληψίας που ακολουθούνται κατά τη διαδικασία των ερευνών και θα αναλύσουμε τις δύο πιο σημαντικές μεθόδους, αυτές που συνήθως επιλέγονται, προκειμένου να πραγματοποιηθεί η δειγματοληψία μιας έρευνας. Η ανάλυση των στατιστικών τύπων και των πιθανοτήτων των περισσότερων δειγματοληπτικών μεθόδων παρατίθενται στο Παράρτημα της διπλωματικής μας (δεν χρειάζεται) εργασίας (τυπολόγιο δειγματοληπτικών μεθόδων). Είναι λοιπόν:

- **Απλή τυχαία δειγματοληψία:**

Κάθε υποκείμενο στον πληθυσμό έχει την ίδια πιθανότητα να επιλεγεί στο δείγμα.

- **Τυχαία δειγματοληψία σε στοιβάδες:**

Στοιβάδα είναι μια φυσικά ορισμένη μονάδα (σχολείο, πανεπιστήμιο, νοσοκομείο, πόλη, νομός). Επιλέγουμε τυχαία τις στοιβάδες και τοποθετούμε στο δείγμα όλα τα υποκείμενα (ή ένα δείγμα από κάθε στοιβάδα) των στοιβάδων που επιλέχθηκαν.

- **Cluster and Stratified sampling:**

Cluster: Από την αλυσίδα 10 υποκαταστημάτων της Ψ εταιρείας επιλέγονται πέντε και μετέχουν στο δείγμα όλοι οι υπάλληλοι.

Stratified: Οι υπάλληλοι της X εταιρείας ομαδοποιούνται με κριτήριο το τμήμα τους και επιλέγονται τυχαία 10 από κάθε τμήμα.

- **Πολύ - σταδιακή δειγματοληψία:**

Επιλέγονται στοιβάδες και σε κάθε μια ένα τυχαίο δείγμα. Αυτό μπορεί να γίνει σε οποιοδήποτε στάδιο. Δεν μπορείς εύκολα να καταγράψεις όλους τους Έλληνες που ταξιδεύουν στην Ευρώπη αλλά μπορείς να βρεις τα ταξιδιωτικά γραφεία! Περισσότερες στοιβάδες σημαίνει μικρότερο δείγμα σε κάθε μία!

- **Δειγματοληψία ευκολίας:**

Άτομα που είναι διαθέσιμα. Όσοι επιλέγονται μόνοι τους ή συμφωνούν να συμμετέχουν μπορεί να μην είναι αντιπροσωπευτικό δείγμα του γενικού πληθυσμού. Γενικεύσιμο στον πληθυσμό με τα ίδια χαρακτηριστικά!

- **Χιονοστιβάδα:**

Άτομα που ανήκουν σε μια ομάδα συστήνουν άλλα άτομα που ανήκουν σε αυτή. Όταν δεν μπορεί να καταγραφεί ο πληθυσμός. Π.χ. στέλνουμε ένα ερωτηματολόγιο σε 50 αποφοίτους Μαθηματικών και τους ζητούμε να υποδείξει ο καθένας πέντε άλλους που θα ήταν διατεθειμένοι να συμμετέχουν.

- **Αναλογική:**

Χωρίζουμε τον πληθυσμό σε υποσύνολα, υπολογίζουμε τις αναλογίες στον πληθυσμό και διατηρούμε τις αναλογίες στο δείγμα (διαφέρει από τα στρώματα γιατί (στα στρώματα) σε αυτά/εκεί δεν διατηρούμε τις αναλογίες).

- **Focus groups:**

6 με 10 άτομα με ειδικά χαρακτηριστικά που αντιπροσωπεύουν μια ομάδα του πληθυσμού.

- **Συστηματική Δειγματοληψία:**

Στη συστηματική δειγματοληψία προκειμένου να σχηματίσουμε ένα δείγμα n μονάδων από πληθυσμό μεγέθους N , εργαζόμαστε ως εξής:

α) Επιλέγουμε έναν τυχαίο αριθμό λ μεταξύ 1 και K , όπου $K = N/n$, ο αντίστροφος του κλάσματος δειγματοληψίας ονομάζεται διάστημα δειγματοληψίας

β) Το δείγμα μας θα αποτελείται από τα στοιχεία του πληθυσμού με αριθμούς $\lambda, \lambda+K, \lambda+2K, \dots, \lambda+(-1)K$

Πρακτική εφαρμογή των ανωτέρω έχουμε στο ακόλουθο παράδειγμα:

Έστω μια Πανεπιστημιακή Σχολή, όπου φοιτούν 9.200 φοιτητές και θέλουμε δείγμα από 400. Τότε το διάστημα δειγματοληψίας θα είναι: $9200/400$ δηλαδή το $K=23$.

Μπορούμε τώρα να ενοποιήσουμε τα μητρώα των ετών της σχολής και αφού βρούμε ένα αριθμό μεταξύ 1 και 23 π.χ το 12 να πάρουμε τους φοιτητές με τη σειρά αρχίζοντας από τον 12

12, 35, 58, 81,9.189

Αυτοί οι αριθμοί ανταποκρίνονται στη συνεχή αρίθμηση των φοιτητών που κάναμε μετά την ενοποίηση των καταλόγων.

• Δειγματοληψία κατά στρώματα:

Συνηθισμένη μέθοδος είναι και η δειγματοληψία κατά στρώματα, της οποίας αν και η αναλυτική παρουσίαση, όπως και των περισσότερων παραπάνω μεθόδων δειγματοληψίας, γίνεται στο Παράρτημα (τυπολόγιο δειγματοληπτικών μεθόδων), θα αναφερθούμε στα βασικά πλεονεκτήματά της.

Η αρχή της κατά στρώματα δειγματοληψίας στηρίζεται στην απλή ιδέα πως αν καταφέρουμε να διαιρέσουμε τον πληθυσμό σε ομάδες (στρώματα) σχετικά ομοιογενείς ως προς το υπό μελέτη χαρακτηριστικό, θα αρκεί ένα μικρό δείγμα από κάθε στρώμα για να εκτιμήσουμε την τιμή του χαρακτηριστικού. Στην οριακή περίπτωση, που η ομοιογένεια κάθε στρώματος ήταν απόλυτη, θα αρκούσε δείγμα μιας μονάδας από κάθε στρώμα. Στη συνέχεια, οι εκτιμήσεις αυτές συνδυάζονται για να μας δώσουν την εκτίμηση για το σύνολο του πληθυσμού.

Στις περιπτώσεις της κατά στρώματα δειγματοληψίας τα χαρακτηριστικά βάσει των οποίων εξετάζεται η ομοιογένεια των στρωμάτων, μπορεί να είναι:

- η γεωγραφική θέση (περιφέρεια)
- το είδος του οικισμού (αστικός, ημιαστικός, αγροτικός)
- το φύλο
- οι ηλικιακές ομάδες
- η κατάσταση απασχόλησης (εργαζόμενοι ή άνεργοι) κ.λπ., γενικότερα, δηλαδή, στοιχεία ταυτότητας από τα οποία αναμένεται να προκύπτουν ομοιογενή ή διαφορετικά γνωρίσματα για κάθε στρώμα.

Από τα παραπάνω, βγαίνει εύκολα το συμπέρασμα πως δεν υπάρχει περίπτωση μία κατά στρώματα δειγματοληψία να δώσει λιγότερο ακριβή αποτελέσματα από την απλή τυχαία δειγματοληψία. Στην χειρότερη περίπτωση, αν όλα τα στρώματα είναι εξ' ίσου ανομοιογενή με τον πληθυσμό ως προς το χαρακτηριστικό που μας ενδιαφέρει, οι εκτιμήσεις θα έχουν την ίδια ακρίβεια με την απλή τυχαία δειγματοληψία, για το ίδιο μέγεθος δείγματος.

2.2.4 Μεταβλητές (κριτήρια) και κλίμακες

2.2.4.1 Τι είναι οι μεταβλητές

Οι μεταβλητές είναι οντότητες που μετράμε, ελέγχουμε, ή χειριζόμαστε στην έρευνα. Διαφέρουν από πολλές απόψεις και ειδικότερα στο ρόλο που δίνουν στην έρευνά μας και στον τύπο μέτρων που μπορεί να εφαρμοστεί σε αυτές.

Συσχετιστική και πειραματική έρευνα

Η εμπειρική έρευνα ανήκει σαφώς σε μια από αυτές τις δύο γενικές κατηγορίες. Στη συσχετιστική έρευνα προσπαθούμε να μην επηρεάζουμε τις μεταβλητές, αλλά μόνο τις μετράμε και ψάχνουμε τις σχέσεις (συσχετισμούς) μεταξύ κάποιου συνόλου μεταβλητών, όπως η πίεση αίματος και το επίπεδο χοληστερόλης.

Στην πειραματική έρευνα, χειριζόμαστε μερικές μεταβλητές και μετράμε έπειτα τα αποτελέσματα αυτού του χειρισμού σε άλλες μεταβλητές, παραδείγματος χάριν, ένας ερευνητής τεχνητά μπορεί να αυξήσει την πίεση αίματος και να καταγράψει έπειτα το επίπεδο χοληστερόλης.

Η ανάλυση στοιχείων στην πειραματική έρευνα οδηγεί στον υπολογισμό των συσχετισμών μεταξύ των μεταβλητών, συγκεκριμένα, εκείνων που χειρίζονται και εκείνων που επηρεάζονται από το χειρισμό. Εντούτοις, τα πειραματικά στοιχεία μπορούν ενδεχομένως να παρέχουν τις ποιοτικά καλύτερες πληροφορίες: Μόνο τα πειραματικά στοιχεία μπορούν αποφασιστικά να καταδείξουν τις αιτιώδεις σχέσεις μεταξύ των μεταβλητών. Παραδείγματος χάριν, εάν διαπιστώσαμε ότι όποτε αλλάζουμε τις μεταβλητές A, οι μεταβλητές B αλλάζουν, κατόπιν μπορούμε να καταλήξουμε στο συμπέρασμα ότι το A επηρεάζει το B. Τα στοιχεία από τη συσχετιστική έρευνα μπορούν μόνο να ερμηνεύσουν τους αιτιώδεις όρους, βασισμένους σε μερικές θεωρίες που έχουμε, αλλά τα συσχετιστικά στοιχεία δεν μπορούν αποφασιστικά να αποδείξουν την αιτιότητα.

2.2.4.2 Εξαρτημένες και ανεξάρτητες μεταβλητές

Οι ανεξάρτητες μεταβλητές είναι εκείνες που επεξεργαζόμαστε, ενώ οι εξαρτημένες μεταβλητές μετρούνται μόνο ή καταχωρούνται. Αυτή η διάκριση εμφανίζεται ως ιδιάζουσα, αφού δημιουργεί σύγχυση σε πολλούς, επειδή, όπως μερικοί “ερευνητές” λένε, όλες οι μεταβλητές εξαρτώνται από κάτι. Η ορολογία «εξαρτημένη και ανεξάρτητη μεταβλητή» ισχύει συνήθως για την πειραματική έρευνα, όπου μερικές μεταβλητές επιδέχονται επεξεργασία, και από αυτή την άποψη είναι ανεξάρτητες κατά τα αρχικά σχέδια αντίδρασης, τα χαρακτηριστικά γνωρίσματα, τις προθέσεις, κ.λπ. των θεμάτων. Μερικές άλλες μεταβλητές αναμένονται να θεωρηθούν εξαρτημένες από την επεξεργασία τους, το χειρισμό ή τις

πειραματικές συνθήκες. Δηλαδή, εξαρτώνται από την απάντηση στην ερώτηση "τι θα κάνει το υποκείμενο".

Κάπως αντίθετα προς τη φύση αυτής της διάκρισης, αυτοί οι όροι χρησιμοποιούνται επίσης στις μελέτες όπου δεν χειριζόμαστε κυριολεκτικά τις ανεξάρτητες μεταβλητές, αλλά μόνο ορίζουν ιδιότητες στις πειραματικές ομάδες που ορίζονται από μερικές προϋπάρχουσες ιδιότητες των θεμάτων. Παραδείγματος χάριν, εάν σε ένα πείραμα, τα αρσενικά συγκρίνονται με τα θηλυκά σχετικά με τα χρωμοσώματα τους το φύλο θα μπορούσε να κληθεί ως ανεξάρτητη μεταβλητή και το χρωμόσωμα ως εξαρτημένη μεταβλητή.

2.2.4.3 Μέτρηση των μεταβλητών :

- Κλίμακα λόγου (ratio scale): Περιλαμβάνει όλα τα ποσοτικά χαρακτηριστικά, αλλά χρησιμοποιείται κυρίως σε δεδομένα όπου υπάρχει νόημα για αναφορά στη μηδενική τιμή.
- Κλίμακα διαστήματος (interval scale): Αναφέρεται σε δεδομένα που οι τιμές τους μπορούν να συγκριθούν ως προς τις μεταβολές τους, αλλά όχι μεταξύ τους. Χαρακτηριστικό παράδειγμα είναι η θερμοκρασία. Μπορούμε να συζητήσουμε για αύξηση της θερμοκρασίας, αλλά ποτέ δε θα πούμε για ποσοστιαία τιμή.
- Κλίμακα ιεράρχησης (ordinal scale): Αφορά τις τιμές των χαρακτηριστικών που αντικείμενό τους είναι να ιεραρχήσουν τις παρατηρήσεις με αύξουσα ή φθίνουσα σειρά. Η βαθμολογία αποτελεί χαρακτηριστικό παράδειγμα.
- Ονομαστική κλίμακα (nominal scale): Όταν χρειαζόμαστε απλή ταξινόμηση των παρατηρήσεων που κατηγοριοποιούν τα δεδομένα (π.χ. Φύλο, οικογενειακή κατάσταση κ.λπ.).

2.2.4.4 Σχέσεις μεταξύ των μεταβλητών

Ανεξάρτητα από τον τύπο τους, δύο ή περισσότερες μεταβλητές συσχετίζονται εάν σε ένα δείγμα των παρατηρήσεων, οι τιμές εκείνων των μεταβλητών κατανέμονται κατά τρόπο συνεπή. Με άλλα λόγια, οι μεταβλητές συσχετίζονται, εάν οι τιμές τους αντιστοιχούν συστηματικά η μια στην άλλη για αυτές τις παρατηρήσεις.

Παραδείγματος χάριν, το ύψος συσχετίζεται με το βάρος, επειδή χαρακτηριστικά τα ψηλά άτομα είναι βαρύτερα από τα κοντά, ή ο δείκτης νοημοσύνης συσχετίζεται με τον αριθμό λαθών σε μια δοκιμή, αφού οι άνθρωποι με τον υψηλότερο δείκτη νοημοσύνης κάνουν λιγότερα λάθη.

Γιατί οι σχέσεις μεταξύ των μεταβλητών είναι σημαντικές;

Γενικά, ο απώτερος στόχος κάθε έρευνας ή της επιστημονικής ανάλυσης είναι να βρει τις σχέσεις μεταξύ των μεταβλητών. Η φιλοσοφία της επιστήμης μας διδάσκει ότι δεν υπάρχει κανένας άλλος τρόπος για να αναζητήσουμε τη σημασία ενός φαινομένου εκτός από τη συσχέτιση ποσοτικών και ποιοτικών χαρακτηριστικών, επομένως την αναζήτηση σχέσεων μεταξύ των μεταβλητών. Κατά συνέπεια, η πρόοδος της επιστήμης πρέπει πάντα να περιλαμβάνει την εύρεση των νέων σχέσεων μεταξύ των μεταβλητών. Η συσχετιστική έρευνα περιλαμβάνει τη μέτρηση τέτοιων σχέσεων με τον απλούστερο τρόπο. Εντούτοις, η πειραματική έρευνα δεν είναι καθόλου διαφορετική από αυτή την άποψη. Παραδείγματος χάριν, το προαναφερθέν πείραμα που συγκρίνει χρωμοσώματα στα αρσενικά και τα θηλυκά, μπορεί να περιγραφεί, ψάχνοντας έναν συσχετισμό μεταξύ δύο μεταβλητών: φύλο και χρωμοσώματα. Οι στατιστικές μας βοηθούν να αξιολογήσουμε τις σχέσεις μεταξύ των μεταβλητών.

Δύο βασικά χαρακτηριστικά γνωρίσματα κάθε σχέσης μεταξύ των μεταβλητών:

Οι δύο πιο στοιχειώδεις ιδιότητες κάθε σχέσης μεταξύ των μεταβλητών είναι το μέγεθος της σχέσης και η αξιοπιστία της.

1. **Μέγεθος.** Το μέγεθος είναι πολύ ευκολότερο να το καταλάβουμε και να το μετρήσουμε σε σχέση με την αξιοπιστία. Παραδείγματος χάριν, εάν κάθε αρσενικό στο δείγμα μας βρέθηκε να είναι ψηλότερο από οποιοδήποτε θηλυκό στο δείγμα, θα μπορούσαμε να πούμε ότι το μέγεθος της σχέσης μεταξύ των δύο μεταβλητών (φύλο και ύψος) είναι πολύ υψηλό στο δείγμα μας. Με άλλα λόγια, θα μπορούσαμε να προβλέψουμε το ένα βασιζόμενοι στο άλλο (τουλάχιστον μεταξύ των μελών του δείγματός μας).
2. **Αξιοπιστία.** Η αξιοπιστία μιας σχέσης είναι μια πολύ λιγότερο διαισθητική έννοια, αλλά εξαιρετικά σημαντική. Αναφέρεται στη "αντιπροσωπευτικότητα" του αποτελέσματος που βρίσκεται στο συγκεκριμένο δείγμα μας για το σύνολο του πληθυσμού. Με άλλα λόγια, λέει πόσο πιθανό είναι ότι θα εντοπιζόταν μια παρόμοια σχέση, εάν το πείραμα ξαναγινόταν με άλλα δείγματα που προήλθαν από τον ίδιο πληθυσμό. Θυμηθείτε ότι δεν ενδιαφερόμαστε σχεδόν ποτέ "τελικά" μόνο για αυτό που γίνεται στο δείγμα μας, αλλά ενδιαφερόμαστε για το δείγμα μόνο στην έκταση που μπορεί να παρέχει τις πληροφορίες για τον πληθυσμό. Εάν η μελέτη μας ικανοποιεί μερικά ειδικά κριτήρια, κατόπιν η αξιοπιστία μιας σχέσης μεταξύ των μεταβλητών που παρατηρούνται στο δείγμα μας μπορεί να υπολογιστεί ποσοτικά και να αντιπροσωπευθεί χρησιμοποιώντας ένα τυποποιημένο μέτρο (τεχνικά αποκαλούμενο ως p-value ή στατιστικό επίπεδο σημαντικότητας).

Τι είναι η p-value:

Η στατιστική σημαντικότητα ενός αποτελέσματος είναι η πιθανότητα ότι η παρατηρηθείσα σχέση (π.χ., μεταξύ των μεταβλητών) ή της διαφοράς (π.χ., μεταξύ

των μέσων) σε ένα δείγμα εμφανίστηκε κατά καθαρή τύχη και ότι στον πληθυσμό από τον οποίο το δείγμα προήλθε, καμία τέτοια σχέση ή διαφορά δεν υπάρχει.

Χρησιμοποιώντας λιγότερο τεχνικούς όρους, κάποιος θα μπορούσε να πει ότι η στατιστική σημασία ενός αποτελέσματος μας λέει κάτι για το βαθμό στον οποίο το αποτέλεσμα είναι "αληθινό" (από την άποψη της ύπαρξης "αντιπροσώπευσης του πληθυσμού"). Πιο τεχνικά, η τιμή της p-value αντιπροσωπεύει έναν δείκτη της αξιοπιστίας ενός αποτελέσματος.

Συγκεκριμένα, η p-value αντιπροσωπεύει την πιθανότητα του λάθους που περιλαμβάνεται στην αποδοχή του παρατηρηθέντος αποτελέσματος μας ως έγκυρου, δηλαδή ως "η αντιπροσώπευση του πληθυσμού". Παραδείγματος χάριν, μια p-value του 0.05 (δηλ., 1/20) δείχνει ότι υπάρχει μια πιθανότητα 5% ότι η σχέση μεταξύ των μεταβλητών που βρίσκονται στο δείγμα μας μπορεί να είναι "ψευδής". Με άλλα λόγια, υποθέτοντας ότι στον πληθυσμό δεν υπήρξε καμία σχέση μεταξύ εκείνων των μεταβλητών, και επαναλαμβάνοντας τα πειράματά μας, θα μπορούσαμε να αναμείνουμε ότι περίπου σε κάθε 20 επαναλήψεις του πειράματος θα υπήρχε ένα στο οποίο η σχέση μεταξύ των εν λόγω μεταβλητών θα ήταν ίση ή ισχυρότερη απ' ότι στους δικούς μας υπολογισμούς. Σε πολλούς τομείς της έρευνας, η p-value του 0.05 είναι συνήθως η διαχωριστική γραμμή ως αποδεκτό "επίπεδο λάθους".

2.2.4.5 Κλίμακες μέτρησης

Μία κλίμακα μέτρησης είναι μία διαδικασία συνδυασμού δεικτών σε μια μετρίσιμη διάσταση η οποία αντιπροσωπεύει την ευρύτερη έννοια. Οι κλίμακες δίνουν την δυνατότητα ανάπτυξης της έννοιας (βασίζονται στην συσσώρευση των δεικτών μιας μεταβλητής). Η γενική κλίμακα αποτελείται από μια σειρά ερωτήσεων οι οποίες ενοποιούνται και μετρώνται ως σύνολο. Ο συνολικός βαθμός που προκύπτει από τον ερωτώμενο αντιπροσωπεύει την γενική του στάση / τακτική κλίμακα. Οι σπουδαιότερες κλίμακες είναι : Likert, Guttman, Thurstone, Semantic Differential, Borgadus, τις οποίες και θα παραθέσουμε παρακάτω:

♦ Κλίμακα Likert

1. Βαθμός συμφωνίας ή διαφωνίας σε κάποιο θέμα που διατυπώνεται από το ερωτηματολόγιο. Η κλίμακα παρουσιάζεται ευθεία γραμμή με σταθερές διαιρέσεις (συνήθως 5)/ ευρεία χρήση.
2. Μία πρόταση όπου ο ερωτώμενος καλείται να απαντήσει εάν συμφωνεί ή διαφωνεί, ως σύνολο, όχι μία- μία.

♦ Κλίμακα Guttman

1. Μετράται η ένταση και η βαρύτητα της έννοιας.
2. Εξετάζει το διαφορετικό βαθμό, όπου ένα άτομο συνδέεται με μια άλλη ομάδα.

3. Οι ερωτήσεις τίθενται σε ιεραρχική διάταξη και εκφράζουν μειωμένη ή αυξημένη στήριξη.
4. Κάθε ερώτηση συνοδεύεται από δύο απαντήσεις/ οι απαντήσεις μπορούν να συνοψιστούν σε ένα κοινό αριθμό/ αντιπροσωπεύει το πιο κοντινό αποδεκτό αποτέλεσμα.

♦ **Κλίμακα Thurstone**

1. Κατασκευάζει την κλίμακα με την συμμετοχή των ίδιων των ερωτώμενων.
2. Εκφράζει την αντίληψη των ερωτώμενων περισσότερο από τις άλλες.
3. Ερευνητής διατυπώνει 50-100 ερωτήσεις/ τις διανέμει/ ποιες έχουν την μεγαλύτερη συχνότητα;

2.2.5 Υλοποίηση της Έρευνας

Μετά την επιλογή του δείγματος έχουμε εντοπίσει τα ερωτώμενα άτομα από τα οποία θα ληφθούν οι προβλεπόμενες συνεντεύξεις υπό μορφή δομημένων ερωτηματολογίων.

2.2.5.1 Διεξαγωγή Επιτόπιας Έρευνας

Υπό την καθοδήγηση της ερευνητικής ομάδας και τον έλεγχο των εποπτών διεξάγεται τότε (δεν χρειάζεται) η επιτόπια έρευνα, κατά την οποία εκπαιδευμένοι στοιχειολήπτες - συνεντευκτές συμπληρώνουν τα ερωτηματολόγια.

Συχνά (κυρίως στις μεγάλης κλίμακας ποσοτικές έρευνες) μεγάλο βάρος δίδεται στα δημογραφικά στοιχεία, ούτως ώστε να ομαδοποιούνται μετά τα ερωτηματολόγια για την περαιτέρω επεξεργασία τους.

Στην εμπειρική έρευνα η παρουσίαση της δημογραφικής δομής του ερευνώμενου πληθυσμού αποτελεί βασική αρχή, διότι τα δημογραφικά χαρακτηριστικά είναι οι βασικές ανεξάρτητες μεταβλητές με τις οποίες, συσχετιζόμενες οι στάσεις και αντιδράσεις των ερωτώμενων ως προς το ερευνώμενο αντικείμενο, καταδεικνύουν την ύπαρξη ή μη κάποιας εξάρτησης από αυτές.

Βασικά δημογραφικά χαρακτηριστικά είναι:

- Φύλο
- Ηλικία
- Επίπεδο εκπαίδευσης
- Απασχόληση (επάγγελμα, κλάδος οικιακής δραστηριότητας, θέση στο επάγγελμα)
- Οικογενειακή κατάσταση
- Αριθμός παιδιών
- Αριθμός μελών νοικοκυριού και σχέση με τον ερωτώμενο.

Για τη τήρηση των κανόνων της δειγματοληψίας και την αποφυγή μεροληπτικών σφαλμάτων ιδιαίτερη προσοχή πρέπει να δίδεται στις πιθανές απουσίες ερωτώμενων, στις αρνήσεις συμμετοχής στην έρευνα και στις αντικαταστάσεις ερωτώμενων.

Όταν το επιλεγόμενο άτομο ενός νοικοκυριού απουσιάζει (σε τρεις διαδοχικές και σε διαφορετική χρονική στιγμή επισκέψεις) γίνεται αντικατάσταση κατά τον εξής τρόπο: Στο αμέσως προηγούμενο νοικοκυριό αναζητούμε άτομο με παρόμοια δημογραφικά χαρακτηριστικά, δηλαδή φύλο, ηλικία και σύνθεση νοικοκυριού. Εάν υπάρχει άτομο που ν' ανταποκρίνεται στα ζητούμενα χαρακτηριστικά, πηγαίνουμε στο αμέσως προηγούμενο κ.ο.κ., έως ότου βρεθεί το κατάλληλο άτομο προκειμένου να αποφευχθεί παραβίαση της τυχαιότητας του δείγματος. Στις περιπτώσεις κατηγορηματικής άρνησης του ατόμου που έχει επιλεγεί να συμμετάσχει στην έρευνα, θεωρούμε το νοικοκυριό τυπικά εκτός δείγματος και δεν προχωρούμε σε αντικατάσταση του συγκεκριμένου ατόμου. Εν συνεχεία, πραγματοποιείται ο προβλεπόμενος αριθμός συνεντεύξεων και ολοκληρώνεται έτσι η διαδικασία της επιτόπιας έρευνας.

2.2.6 Επεξεργασία Δεδομένων

Μετά τη συλλογή των δεδομένων αρχίζει το στάδιο της επεξεργασίας τους. Μπορεί μάλιστα κανείς να πει ότι η φάση της επεξεργασίας αρχίζει πολύ πριν από τη συλλογή των δεδομένων και συνεχίζεται κατά τη διάρκειά της. Με αυτό εννοούμε ότι κατά τη φάση του σχεδιασμού, ήδη έχει αποφασισθεί ένα «πλάνο» επεξεργασίας περισσότερο ή λιγότερο λεπτομερές, ανάλογα με τη γνώση του αντικειμένου της έρευνας, τη φύση της έρευνας (διερευνητική, περιγραφική, πειραματική) και την οργάνωσή της και ότι κατά τη διάρκεια της συλλογής των στοιχείων η πρώτη πράξη της επεξεργασίας, δηλαδή ο έλεγχος των ερωτηματολογίων, ήδη συντελείται. Επίσης, η κωδικογράφηση γίνεται κατά μεγαλύτερο ή μικρότερο ποσοστό παράλληλα με τη συλλογή των στοιχείων για οικονομία χρόνου. Αν μάλιστα οι απαντήσεις είναι προκωδικογραφημένες 100%, η κωδικογράφηση τελειώνει σχεδόν ταυτόχρονα με τη συλλογή των στοιχείων.

Η φάση της επεξεργασίας αποτελείται από τα εξής διακεκριμένα στάδια:

- έλεγχος
- κωδικογράφηση
- μηχανογραφική επεξεργασία

2.2.7 Έλεγχος

Ο έλεγχος πρέπει φυσικά να ασκείται σε όλη τη διάρκεια της έρευνας, για κάθε δραστηριότητα: έλεγχος της σωστής διατύπωσης των στόχων της έρευνας στο

ερωτηματολόγιο, έλεγχος στη σωστή εκτύπωση των εντύπων, έλεγχος στην επιλογή των συνεντευκτών. Ο κυρίως, όμως, έλεγχος αφορά την ορθή συμπλήρωση και την τήρηση των κανόνων δειγματοληψίας.

Η στοιχειοληψία επιτελείται από ομάδες συνεντευκτών (μικρότερη ή μεγαλύτερη ανάλογα με το μέγεθος του ερωτηματολογίου και τη γεωγραφική διασπορά της έρευνας), υπό την εποπτεία υπευθύνων ατόμων που πραγματοποιούν τους προαναφερόμενους ελέγχους.

2.2.8 Κωδικογράφηση και Εισαγωγή Στοιχείων στον Ηλεκτρονικό Υπολογιστή (H/Y)

Με τον όρο κωδικογράφηση εννοούμε την μετατροπή των απαντήσεων σε αριθμούς ή σύμβολα, δηλαδή το ποιοτικό στοιχείο (ολόκληρες φράσεις, ένα όνομα, μια κατάφαση ή άρνηση κ.λπ.) σε ποσοτικό ή ποιοτικό-συμβολικό.

Φυσικά η απάντηση μπορεί να έχει ένα ήδη αριθμό, οπότε δεν χρειάζεται μετατροπή. Αυτό όμως στις περισσότερες κοινωνικές έρευνες (κυρίως διερευνητικές) αφορά ένα μικρό ποσοστό των απαντήσεων.

Η κωδικογράφηση, λοιπόν, μετατρέπει τις απαντήσεις σε μορφή κατάλληλη για μηχανογραφική επεξεργασία. Η κωδικογράφηση μπορεί να προετοιμασθεί (κατά μεγαλύτερο ή μικρότερο βαθμό) στη φάση του σχεδιασμού, με την κωδικοποίηση των προβλεπόμενων απαντήσεων. Με κωδικοποίηση εννοούμε την πρόβλεψη των δυνατών κατηγοριών απάντησης σε κάθε ερώτηση. Σε αυτήν την περίπτωση οι απαντήσεις είναι προκωδικογραφημένες και οι ερωτήσεις χαρακτηρίζονται «κλειστές», ενώ αντίθετα στις «ανοιχτές» ερωτήσεις δεν υπάρχει πρόβλεψη απάντησης (δεν είναι προκωδικογραφημένες). Στην περίπτωση αυτή, οι απαντήσεις ομαδοποιούνται, εκ των υστέρων, κατά κατηγορίες, χαρακτηρίζονται με κωδικούς οι ομάδες απαντήσεων και κωδικογραφούνται.

Με την ολοκλήρωση των προβλεπόμενων ελέγχων και της κωδικογράφησης, τα στοιχεία των ερωτηματολογίων είναι έτοιμα για την εισαγωγή τους στον Ηλεκτρονικό Υπολογιστή (H/Y). Παλαιότερα, η εισαγωγή στον H/Y γινόταν με αποκλειστικό σκοπό τη στατιστική επεξεργασία των στοιχείων, έτσι ώστε να αποτελέσουν το υλικό για την ανάλυση των αποτελεσμάτων και τη διατύπωση των πορισμάτων της έρευνας. Σήμερα, για την εισαγωγή των στοιχείων στον H/Y δημιουργείται μια Βάση Δεδομένων, η οποία συμπληρώνεται με τα στοιχεία των ερωτηματολογίων και αποτελεί αυτοτελές υλικό. Για την περαιτέρω στατιστική επεξεργασία των δεδομένων χρησιμοποιούνται εξειδικευμένα Στατιστικά Προγράμματα, στα οποία αναφερόμαστε παρακάτω.

2.2.9 Στατιστική Επεξεργασία των Δεδομένων

Τα προγράμματα διαχείρισης βάσεων δεδομένων (ΒΔ) δεν παρέχουν επαρκείς λειτουργίες για τη στατιστική επεξεργασία των δεδομένων. Για το σκοπό αυτό, υπάρχουν ειδικά προγράμματα, που προσφέρονται για απλές και σύνθετες στατιστικές επεξεργασίες.

Στα πλαίσια της ανάλυσης των δεδομένων διερευνώνται οι αλληλεξαρτήσεις των διαθέσιμων δημογραφικών και μη χαρακτηριστικών, καθώς και των κύριων στάσεων και απόψεων σχετικά με τα υπό εξέταση κάθε φορά θέματα, με τις κατάλληλες τεχνικές (X²-test, Fisher's exact tests, Spearman correlation coefficients, κλπ). Σε κάθε περίπτωση, οι τεχνικές στατιστικής ανάλυσης προσαρμόζονται πλήρως στις υποθέσεις εργασίας κάθε συγκεκριμένης έρευνας. Επίσης, ένα από τα βασικά προβλήματα που διερευνάται κάθε φορά είναι αυτό της 'μη απόκρισης' και των αντικαταστάσεων του δείγματος. Ειδικότερα, ως προς τη 'μη-απόκριση' (non-response) κάποιας μερίδας του δείγματος στο σύνολο του ερωτηματολογίου ή σε ορισμένες ερωτήσεις, επιχειρείται ειδική ανάλυση, προκειμένου να διαπιστωθεί αν αυτή είναι μεροληπτική (αν δηλαδή συνιστά συστηματική απόκρυψη απόψεων) και συνεπώς η προσέγγιση των πορισμάτων της έρευνας χρειάζεται συγκριτική και ποιοτική επανατοποθέτηση.

Χρησιμοποιούνται, εκτενώς, οι στατιστικές μέθοδοι και τεχνικές της απλής και πολυδιάστατης ανάλυσης δεδομένων, όπως η Παραγοντική ανάλυση δεδομένων (Factor analysis), η Διακρίνουσα ανάλυση (Discriminant analysis), οι Μέθοδοι αυτόματης ταξινόμησης (Cluster analysis), η Πολυδιάστατη Ταξινόμηση (Multidimensional scaling), η Ανάλυση Πολλαπλών Αντιστοιχιών (Correspondence analysis), η Πολυκατηγορική Λογιστική Παλινδρόμηση (Multinomial Logistic Regression).

Παρουσιάζοντας εν συντομία μερικές από τις παραπάνω χρήσιμες και συνήθεις εργασίες που εκτελούν αυτά τα Στατιστικά Προγράμματα (ΣΠ), έχουμε:

Μετατροπή (Recode)

Ορισμένες μεταβλητές για να τύχουν στατιστικής επεξεργασίας χρειάζονται μετατροπές. Τα Σ.Π. μας επιτρέπουν να μετατρέψουμε τα στοιχεία μας, συνδυάζοντας διάφορες κατηγορίες σε μία ή ομαδοποιώντας σε λιγότερες κατηγορίες τις τιμές μιας μεταβλητής.

Παράδειγμα:

Η χρήση της μεταβλητής «έτος γέννησης», που ίσως έχουμε, και που περιέχει το έτος γέννησης των καθηγητών της υποθετικής μας ΒΔ (:), χρειάζεται τις εξής μετατροπές:

α) Να αφαιρεθεί το έτος γέννησης από το τρέχον έτος, για να βρεθεί η ηλικία των καθηγητών σε έτη και

β) Να ομαδοποιηθεί σε ορισμένες κατηγορίες π.χ. 21-25, 26-30, 31-35 κλπ.

Με τον τρόπο αυτό, δημιουργείται μια νέα μεταβλητή : «Ομάδες Ηλικιών»,

την οποία μπορούμε να χειριστούμε πιο εύκολα για σύνθετες στατιστικές επεξεργασίες σε συνδυασμό με άλλες μεταβλητές.

Συχνότητες

Στις συχνότητες βλέπουμε με μια πρώτη ματιά το πώς κατανέμονται τα δεδομένα μας. Είναι κατάλληλες για ορισμένο τύπο στοιχείων. Κυρίως όταν τα στοιχεία είναι ομαδοποιημένα σε λίγες μεγάλες κατηγορίες, για παράδειγμα το φύλο των καθηγητών (τόσοι άνδρες, τόσες γυναίκες), για την Ομάδα Ηλικιών. Δεν είναι κατάλληλες για στοιχεία που έχουν μια συνεχή σειρά, για παράδειγμα το έτος γέννησης κάθε καθηγητή.

Πινακοποίηση

Οι πιο απλές διασταυρώσεις μεταξύ δύο ή περισσότερων μεταβλητών παρουσιάζονται συνήθως με τη μορφή πινάκων, όπου στον οριζόντιο άξονα είναι η μια μεταβλητή και στον κάθετο η άλλη. Και σε αυτή την περίπτωση η διασταύρωση έχει νόημα, όταν οι μεταβλητές είναι ομαδοποιημένες σε λίγες κατηγορίες.

Παλινδρόμηση

Η διαδικασία αυτή εκτιμά τη σχέση μεταξύ δύο μεταβλητών. Αν συσχετίσουμε για παράδειγμα τη μεταβλητή «μισθός» με τη μεταβλητή «χρόνια προϋπηρεσίας» ενός καθηγητή είναι πολύ πιθανόν να βρούμε ότι υπάρχει μια πολύ ισχυρή θετική σχέση μεταξύ των δύο μεταβλητών (δηλαδή, ο μισθός να αυξάνει με τα χρόνια προϋπηρεσίας). Τη διαδικασία αυτή πιθανόν να θέλουμε να την εφαρμόσουμε για να ελέγξουμε το κατά πόσο σχετίζεται η απόδοση ενός μαθητή με το μορφωτικό επίπεδο των γονιών του κ.λπ.

Αφήσαμε για το τέλος, όσο κι αν φαντάζει αυτονόητο, τη γραφική απεικόνιση όλων αυτών των στοιχείων που έχουμε συλλέξει. Μια εικόνα μιλά όσο χίλιες λέξεις και μια από τις απαραίτητες πλέον λειτουργίες των Στατιστικών Πακέτων είναι η δημιουργία γραφικών. Η γραφική απεικόνιση των στοιχείων είναι ένα βοηθητικό μέσο για τη συναγωγή γενικών συμπερασμάτων από πίνακες που μπορεί να είναι αρκετά πολύπλοκοι. Τα συνηθέστερα γραφικά που χρησιμοποιούμε είναι το διάγραμμα, οι στήλες και οι πίτες. Το γραμμικό διάγραμμα χρησιμοποιείται συνήθως για μη ομαδοποιημένα στοιχεία, οι στήλες για δεδομένα που είναι ομαδοποιημένα σε αρκετές κατηγορίες, ενώ οι πίτες για δεδομένα ομαδοποιημένα σε πολύ λίγες κατηγορίες.

2.3 Ποιοτικές έρευνες

Αυτοτελώς ή συμπληρωματικά προς τις ποσοτικές τεχνικές η ποιοτική προσέγγιση στην έρευνα στοχεύει στη διερεύνηση και κατανόηση σε βάθος των κοινωνικών φαινομένων. Παρέχοντας τη δυνατότητα στον ερευνητή να αντλήσει πλούσιες πληροφορίες για το υπό εξέταση θέμα, η ποιοτική έρευνα αποτελεί την ενδεδειγμένη μεθοδολογία για να απαντηθούν τα ερωτήματα που σχετίζονται με το “Γιατί;” και το “Πώς;” των φαινομένων. Η ποιοτική προσέγγιση αποτελεί μια κατά βάση διερευνητική (exploratory) μέθοδο. Στοχεύει περισσότερο στην ανάδυση νέων τυποποιήσεων και θεωρητικών μοντέλων παρά στην επαλήθευση υποθέσεων. Το βασικό πλεονέκτημα των ποιοτικών μεθόδων που εξυπηρετεί αυτή τη στόχευση είναι η ευελιξία που χαρακτηρίζει την ερευνητική διαδικασία. Πιο συγκεκριμένα:

- Ο ερευνητικός σχεδιασμός βασίζεται σε ευέλικτες στρατηγικές που επιτρέπουν την ανατροφοδότηση των σταδίων της έρευνας, όπου αυτό κρίνεται αναγκαίο.
- Τα ερωτήματα που απευθύνονται στους συμμετέχοντες είναι ανοιχτού τύπου (μη-δομημένο ερωτηματολόγιο), γεγονός που αποτρέπει την επιβολή προκατασκευασμένων νοηματικών πλαισίων στα υποκείμενα της έρευνας, αλλά και αφήνει τη δυνατότητα να ανακύψουν θέματα που πιθανώς δεν είχαν προβλεφθεί κατά το σχεδιασμό.
- Η ανάλυση των δεδομένων βασίζεται σε δυναμικά εργαλεία, όπως είναι η ανοιχτή κωδικοποίηση των δεδομένων και η συνεχής σύγκριση, που επιτρέπουν στον ερευνητή να «επιστρέφει» στα δεδομένα σε όλες τις φάσεις της αναλυτικής διαδικασίας.

Η ποιοτική έρευνα αποτελεί την κατάλληλη μεθοδολογική επιλογή για να διερευνηθούν σε βάθος οι αναπαραστάσεις, οι στάσεις, οι αντιλήψεις, τα κίνητρα, καθώς και τα συναισθηματικά και τα συμβολικά/φαντασιακά δεδομένα της συμπεριφοράς των ατόμων. Στόχος της ποιοτικής διερεύνησης δεν αποτελεί απλά η περιγραφή μιας στάσης ή μιας συμπεριφοράς, αλλά η ολιστική κατανόησή τους. Η ποιοτική έρευνα διερευνά την εμπειρία των ατόμων και τα υποκειμενικά νοήματα που τη συγκροτούν, εστιάζοντας πάντα στο ευρύτερο κοινωνικό και πολιτισμικό (αξιακό και ιδεολογικό) πλαίσιο (context) στο οποίο εγγράφεται. Ταυτόχρονα, η ποιοτική έρευνα δίνει ιδιαίτερη έμφαση στη δυναμική διάσταση των φαινομένων. Παρέχει τη δυνατότητα να περιγραφούν λεπτομερώς οι διαδικασίες που συγκροτούν μια κοινωνική κατάσταση, επιτρέποντας να διατυπωθούν υποθέσεις ως προς την ερμηνεία της. Η έμφαση στη δυναμική διάσταση των φαινομένων καθιστά την ποιοτική έρευνα κατάλληλη για να διερευνηθούν μεταβατικά φαινόμενα καθώς και οι πιθανές εκβάσεις τους (περίοδοι κοινωνικής και πολιτικής κρίσης). Τέλος, η χρήση των κατάλληλων ποιοτικών τεχνικών έμμεσης προσέγγισης (Προβολικές τεχνικές, έμμεσες ερωτήσεις) των ερευνητικών υποκειμένων, δίνει τη δυνατότητα στον ερευνητή να υπεισέλθει στα βαθύτερα δεδομένα της συμπεριφοράς,

προσπελάζοντας τις ορθολογικές ερμηνείες και τις στερεοτυπικές απαντήσεις που προβάλλουν συνήθως τα άτομα.

Περιγράψαμε παραπάνω τη μεθοδολογία πραγματοποίησης των ερευνών αγοράς και ανάλυσης της συμπεριφοράς των πολιτών κατά την εκλογική διαδικασία, τον τρόπο δειγματοληψίας, τις μεταβλητές απόφασης και τα κριτήρια - κλίμακες που χρησιμοποιούνται καθώς και τον τρόπο ανάλυσης των δεδομένων που προκύπτουν. Τι συμβαίνει όμως στο πραγματικό πεδίο της εκλογικής διαδικασίας; Ποια είναι τα εκλογικά συστήματα που αναδεικνύουν το δικαίωμα των πολιτών / ψηφοφόρων να επιλέγουν το ποιοι θα τον εκπροσωπήσουν και θα τον κυβερνήσουν; Ποιοι είναι οι παράγοντες που επιδρούν στο θέμα της επιλογής της ψήφου και ποια τα κριτήρια επιλογής; Θα προσπαθήσουμε, εν συντομία, να αναλύσουμε τις παραπάνω έννοιες, προκειμένου να ολοκληρώσουμε το συγκεκριμένο κεφάλαιο ανάλυσης της υφιστάμενης κατάστασης στο χώρο των ερευνών.

2.4 Το εκλογικό σύστημα

Εκλογικό σύστημα είναι η μέθοδος μετατροπής των ψήφων σε έδρες. Δηλαδή, με ποιόν τρόπο η βούληση του εκλογικού σώματος, που εκφράστηκε στις εκλογές με την ψήφο των εκλογέων θα μεταφραστεί σε κατανομή των βουλευτικών εδρών. Υπάρχουν πολλά εκλογικά συστήματα και πολύ περισσότερες παραλλαγές τους. Τα δύο βασικά, πάντως, συστήματα είναι το πλειοψηφικό (κατ' αυτό, σε κάθε εκλογική περιφέρεια εκλέγεται ο συνδυασμός ή ο υποψήφιος που λαμβάνει την πλειοψηφία των ψήφων) και το αναλογικό (κατά το οποίο οι έδρες κάθε εκλογικής περιφέρειας κατανέμονται στους συνδυασμούς και στους μεμονωμένους υποψηφίους ανάλογα με τις ψήφους που έλαβαν).

Στην Ελλάδα, έχουν εφαρμοστεί και τα δύο βασικά εκλογικά συστήματα. Έτσι, από το 1844 μέχρι το 1923, οι βουλευτικές εκλογές γίνονταν με το πλειοψηφικό σύστημα (και μάλιστα με "σφαιρίδια" και όχι με ψηφοδέλτια). Από το 1926 έως το 1956 εναλλάσσονται τα δύο συστήματα, πλειοψηφικό και αναλογικό. Μετά το 1956, βασικό εκλογικό σύστημα είναι το αναλογικό, σε πολλές παραλλαγές του. Αξίζει να σημειωθεί ότι οι τρεις τελευταίες βουλευτικές εκλογές έχουν διεξαχθεί με την ίδια, κατά βάση, παραλλαγή του αναλογικού συστήματος. Επίσης, ότι σύμφωνα με το αναθεωρημένο άρθρο 54 παρ. 1 του Συντάγματος, το εκλογικό σύστημα ορίζεται με νόμο που ισχύει από τις μεθεπόμενες εκλογές, εκτός εάν προβλέπεται η ισχύς του άμεσα από τις επόμενες εκλογές με ρητή διάταξη που ψηφίζεται με (από) την πλειοψηφία των δύο τρίτων του όλου αριθμού των βουλευτών.

2.4.1 Ψηφοφορία

A) Η αρχή της καθολικής ψηφοφορίας. Σύμφωνα με την αρχή αυτή, από το εκλογικό σώμα αποκλείονται μόνον οι πολίτες εκείνοι που δεν συγκεντρώνουν τις ελάχιστες προϋποθέσεις που θέτει, με αποκλειστικό τρόπο, το Σύνταγμα. Ο κοινός νομοθέτης δεν μπορεί να προβλέψει επί πλέον λόγους στέρησης του δικαιώματος της ψήφου.

B) Η αρχή της ισότητας της ψήφου. Η αρχή αυτή εξειδικεύεται σε δύο επί μέρους αρχές: ότι κάθε πολίτης διαθέτει μόνον μία ψήφο και ότι όλες οι ψήφοι είναι νομικά ισοδύναμες.

Γ) Η αρχή της άμεσης ψηφοφορίας. Κατά την αρχή αυτή, δεν μεσολαβεί κάποια άλλη βούληση μεταξύ του εκλογέα και του αποτελέσματος της εκλογής. Με άλλα λόγια, δεν είναι δυνατόν οι εκλογείς να εκλέξουν κάποιους "εκλέκτορες", οι οποίοι με τη σειρά τους θα εκλέξουν τους βουλευτές.

Δ) Η αρχή της μυστικότητας της ψήφου. Με την αρχή αυτή εξασφαλίζεται το ότι η εκλογική βούληση του εκλογέα δεν θα γίνει γνωστή σε τρίτους.

E) Η αρχή της υποχρεωτικής ψηφοφορίας. Σύμφωνα με την αρχή αυτή, η άσκηση του εκλογικού δικαιώματος είναι υποχρεωτική. Πρέπει πάντως να σημειωθεί ότι με τη συνταγματική αναθεώρηση του 2001 καταργήθηκε η πρόβλεψη για νόμο που θα επιβάλει ποινικές κυρώσεις στον εκλογέα που δεν θα λάβει μέρος στις εκλογές.

ΣΤ) Η αρχή της ταυτόχρονης διενέργειας των εκλογών σε ολόκληρη την Επικράτεια. Με το αναθεωρημένο άρθρο 51 παρ. 4 του Συντάγματος, η αρχή αυτή μπορεί να καμφθεί για εκλογείς που βρίσκονται στο εξωτερικό, αρκεί να τηρείται η αρχή της ταυτόχρονης καταμέτρησης και ανακοίνωσης των αποτελεσμάτων.

Z) Η αρχή της αυτοπρόσωπης άσκησης του εκλογικού δικαιώματος ισχύει, πλέον, μόνον για τους εκλογείς που βρίσκονται στην Επικράτεια. Το αναθεωρημένο άρθρο 51 παρ. 4 του Συντάγματος προβλέπει τη δυνατότητα στους εκλογείς που βρίσκονται στο εξωτερικό να ασκούν το εκλογικό τους δικαίωμα "με επιστολική ψήφο ή άλλο πρόσφορο μέσο".

Σε ένα πολιτικό σύστημα ανοικτής εξουσίας, οι κανόνες του εκλογικού ανταγωνισμού είναι ζωτικής σημασίας για την ομαλή λειτουργία του. Όταν λέμε κανόνες του εκλογικού ανταγωνισμού, εννοούμε μια δέσμη συνταγματικών, νομοθετικών και κανονιστικών διατάξεων που ρυθμίζουν, κατά κύριο λόγο, την εκλογή των κυβερνώντων, το μέγεθος των εκλογικών περιφερειών, της βουλευτικής εκλογιμότητας, την κατανομή των βουλευτικών εδρών μεταξύ των κομμάτων και τη διανομή των εδρών αυτών μεταξύ των κομμάτων και των υποψηφίων του ιδίου συνδυασμού, το δικαστικό έλεγχο του κύρους των εκλογών κ.λπ. Οι ρυθμίσεις που περιέχουν οι κανόνες αυτοί, εκφράζουν τόσο τη φιλοσοφία του πολιτικού

συστήματος, όσο και το πολιτιστικό επίπεδο της κοινωνίας από την οποία προέρχονται και στην οποία απευθύνονται.

Το ελληνικό Σύνταγμα του 1975, όπως άλλωστε και τα προηγούμενα - με μοναδική εξαίρεση το Σύνταγμα του 1925, που κατοχύρωνε το αναλογικό εκλογικό σύστημα - δεν περιέχει διάταξη που να αναφέρεται γενικά στο εκλογικό σύστημα, αλλά περιορίζεται στο να προβλέψει ότι το εκλογικό σύστημα και οι εκλογικές περιφέρειες ορίζονται με νόμο (άρθρο 54 παρ. 1 Σ.). Κατ' εξαίρεση, όμως, καθιερώνει το αναλογικό σύστημα για τους Βουλευτές Επικρατείας, ορίζοντας στο άρθρο 54 παρ. 3 ότι αυτοί εκλέγονται σε συνάρτηση με τη συνολική εκλογική δύναμη κάθε κόμματος στην Επικράτεια, όπως ο νόμος ορίζει.

Σχετικά με το νόμο, γενικά, με τον οποίο θα καθορισθεί το εφαρμοστέο εκλογικό σύστημα, το Σύνταγμα περιορίζεται στο να προβλέπει μόνο (άρθρο 72 παρ. 1) ότι ψηφίζεται από την Ολομέλεια της Βουλής. Και παρόλο που όλες οι εκλογές, από το 1958 και μετά, έγιναν με αναλογικό σύστημα, αναμφισβήτητο είναι ότι πάντοτε το σύστημα, όσο και αν εξακολουθούσε να είναι βασικά αναλογικό, μεταρρυθμιζόταν νομοθετικά, κατά κανόνα, τις παραμονές των εκλογών σε ορισμένα σημεία, ανάλογα με τις επιθυμίες της προεκλογικής κυβέρνησης. Στο σημείο αυτό, δεν πρέπει να διαφεύγει της προσοχής μας το γεγονός ότι η εκάστοτε αντιπολίτευση κατηγορούσε τις προεκλογικές κυβερνήσεις, που προέβαιναν σε τέτοιες ενέργειες, ότι δηλαδή μέσα στο πεδίο της διακριτικής τους ευχέρειας προσπαθούσαν να βρουν σκόπιμες ή πολιτικά επιθυμητές λύσεις, για να διατηρηθούν στην εξουσία. Ακόμα και αν οι θέσεις αυτές της αντιπολίτευσης θεωρηθούν ότι είναι σε ένα σημείο υπερβολικές, από μια επιστημονική άποψη, μπορεί να πει κανείς ότι ο εκλογικός νόμος αποτελεί μια από τις πιο εμφανείς πρακτικές που διαθέτει το πολιτικό σύστημα για να αφομοιώσει, ενσωματώσει ή και εξουδετερώσει εκείνες τις επιλογές που θεωρούνται αντίθετες προς τις επιλογές των κυρίαρχων κοινωνικών συμφερόντων.

Πραγματικά, ο ρόλος του εκλογικού συστήματος στη λειτουργία του πολιτεύματος είναι πρωταρχικός. Η κοινοβουλευτική μας δημοκρατία έχει ανάγκη από θεσμικές ισορροπίες για τη διατήρηση της κοινωνικής ειρήνης, στηριζόμενη στη συνοχή συγκεκριμένων διαύλων έκφρασης και πρακτικής ως εγγυητικών οργανικών όρων πλαισίωσης, σταθεροποίησης και αναπαραγωγής της κοινωνικής και οικονομικής δομής. Οι θεσμικές αυτές ισορροπίες αναφέρονται πρωτίστως ανάμεσα στους πραγματικούς πόλους του πολιτικού γίνεσθαι, δηλαδή στα πολιτικά κόμματα και δευτερευόντως ανάμεσα στον Αρχηγό του Κράτους και την Κυβέρνηση. Πρωτίστως στα πολιτικά κόμματα, γιατί αυτά αναλαμβάνουν την ευθύνη της λειτουργίας του πολιτεύματος, ανεξάρτητα από την ανάδειξη ή όχι απόλυτης κοινοβουλευτικής πλειοψηφίας. Η ρήση του μεγάλου νεοκλασικού της συνταγματικής θεωρίας G. Leibholz, αναφορικά με τη σημερινή μορφή του κοινοβουλευτισμού, για το «κράτος των πολιτικών κομμάτων», ισχύει αναμφισβήτητα και για τα δικά μας δεδομένα. Δευτερευόντως στον Αρχηγό του Κράτους και τη Κυβέρνηση, γιατί ο Ανώτατος Άρχοντας έχει συμβολικό ρόλο που αποτυπώνει την εθνική ενότητα και οργανώνει με διακριτικό και ήπιο τρόπο τον κομματικό διάλογο, που στοχεύει στη διαμόρφωση της αναγκαίας κάθε φορά

ρυθμιστικής συναίνεσης. Εξάλλου, το Σύνταγμα παρέχει το πλαίσιο μιας σχετικά ασφαλούς πολιτικής πρόγνωσης, γιατί οριοθετεί τις πιθανές λύσεις, τυποποιεί τους χειρισμούς και προβλέπει ένα συγκεκριμένο χρονοδιάγραμμα.

Ο εκλογικός νόμος αναφέρεται άμεσα στις κομματικές ισορροπίες, στην κυβερνητική εξουσία και στο προφίλ της αντιπολίτευσης. Επομένως, λειτουργεί ως καθοριστική παράμετρος κατά την ερμηνεία και την εφαρμογή κρίσιμων συνταγματικών διατάξεων στις οποίες στηρίζεται η λειτουργία του πολιτεύματος. Πρόκειται για μια δέσμη διατάξεων, που ρυθμίζουν τους βασικούς μηχανισμούς ενός κοινοβουλευτικού συστήματος : το διορισμό και την παύση της κυβέρνησης (άρθρα 37 και 38), τη διάλυση της Βουλής (άρθρο 41) και τη θέση της αντιπολίτευσης και το ρόλο του Κοινοβουλίου σε σχέση προς την Κυβέρνηση (άρθρο 84).

Το Σύνταγμα, με μια σειρά ρυθμίσεων, επιδιώκει ευθέως την κοινοβουλευτική σταθερότητα (δυσχέρειες στη διάλυση της Βουλής) και την κυβερνητική σταθερότητα (διαδικασίες επίλυσης των κυβερνητικών κρίσεων), κάτι που δεν ισχύει για με τη διασφάλιση της σχετικής, έστω, ισοδυναμίας της ψήφου. Η σχετική ισοδυναμία της ψήφου συνάγεται συμπερασματικά, από την αρχή της λαϊκής κυριαρχίας, η οποία όμως οριοθετείται από το ίδιο το Σύνταγμα και η οποία περιορίζεται από την κοινοβουλευτική μορφή του πολιτεύματος (άρθρο 1). Άλλωστε μη ξεχνάμε τη ρήση του M. Miaille ότι: «Στην πραγματικότητα τίποτα δεν είναι πιο πολιτικό από την επιλογή του εκλογικού συστήματος».

2.4.2 Κριτήρια επιλογής / εκλογής

Η δυνατότητα πρόβλεψης της έκβασης μιας εκλογικής μάχης, εξαρτάται από το πόσο ακριβές είναι το μοντέλο εκλογικής συμπεριφοράς που χρησιμοποιούν οι εκλογολόγοι στις αναλύσεις τους. Σκοπός της συγκεκριμένης ενότητας είναι να συγκρίνει το σημερινό εκλογικό τοπίο με αυτά του παρελθόντος, να εντοπίσει τους άγνωστους συντελεστές διαμόρφωσης του αποτελέσματος και να κωδικοποιήσει τους άξονες του σημερινού μοντέλου εκλογικής συμπεριφοράς. Όσο μεγαλύτερη είναι η προσέγγιση της πραγματικότητας, τόσο πιο επιτυχείς είναι και οι εκλογικές προβλέψεις.

Στις σύγχρονες εκλογικές αναμετρήσεις συγκρούονται με τα κόμματα, και δύο διαφορετικές σχολές πολιτικής σκέψης. Η «παραδοσιακή σχολή» υποστηρίζει ότι τα εκλογικά αποτελέσματα έχουν διαμορφωθεί πολύ πριν από την έναρξη της προεκλογικής καμπάνιας από μακρο - πολιτικούς παράγοντες, ενώ οι οπαδοί της «μοντέρνας σχολής» ισχυρίζονται ότι είναι οι προεκλογικές εκστρατείες αυτές που βγάζουν τις κυβερνήσεις.

Πιο σωστή θέση είναι η σύνθεση των δύο προαναφερθέντων απόψεων. Η πολιτική αντιπαράθεση στις «μεσο-εκλογικές περιόδους» παίζει όντως σημαντικό ρόλο, αλλά η σπουδαιότητά της φθίνει συνεχώς και αν το προβάδισμα του ενός κόμματος δεν είναι πολύ μεγάλο, οι σύγχρονες εκστρατείες είναι ικανές να το ανατρέψουν εις βάρος του άλλου.

Τα τέσσερα (4) κριτήρια, τα οποία παραδοσιακά καθόριζαν την εκλογική συμπεριφορά των πολιτών:

- i. η κομματική ταύτιση,
- ii. η ιδεολογική τοποθέτηση,
- iii. η κοινωνική θέση και
- iv. η οικονομική κατάσταση

αποδυναμώθηκαν σημαντικά τα τελευταία δέκα χρόνια. Σε αυτό συνετέλεσαν τόσο οι κοσμογονικές εξελίξεις σε διεθνές επίπεδο, όπως η κατάρρευση του Ανατολικού Μπλοκ και η παγκοσμιοποίηση των αγορών, όσο και οι επακόλουθες εξελίξεις στο εσωτερικό της χώρας. Επίσης, σήμερα ανεξάρτητα από το γεγονός της εκ νέου ανάπτυξης έντονων φαινομένων κοινωνικής ανισότητας, ένα μεγάλο κομμάτι του ελληνικού πληθυσμού αισθάνεται ότι υπάρχουν ευκαιρίες ατομικής καταξίωσης ανεξάρτητα από το ποιο από τα δύο κόμματα εξουσίας θα κυβερνήσει.

Για την πληρότητα της ανάλυσής μας, οφείλουμε να σταθούμε σε μερικά ενδιαφέροντα στοιχεία που αφορούν τους αναποφάσιστους ψηφοφόρους μια και αυτοί κρίνουν πλέον τα αποτελέσματα των εκλογών. Ομάδες, όπως οι νέοι, οι πτυχιούχοι ανωτέρων και ανωτάτων σχολών και οι κάτοικοι των αστικών κέντρων, παρουσιάζουν μειούμενη κομματική ταύτιση, γεγονός που εντείνει τη διακομματική κινητικότητα. Το ίδιο συμβαίνει επίσης και με την ιδεολογική τους τοποθέτηση, η οποία με την άμβλυνση των διαφορών που υπήρξε μεταξύ των δύο (2) κομμάτων εξουσίας, έχει γίνει ποιο συγκεχυμένη και παίζει ολοένα και μικρότερο ρόλο στην τελική τους επιλογή.

Όλοι οι παραπάνω παράγοντες συντείνουν στην ισχυροποίηση νέων παραμέτρων εκλογικής συμπεριφοράς:

i. Η ψήφος παρόρμησης (impulsive voting), η οποία βασίζεται στις εντυπώσεις που σχηματίζουν οι ψηφοφόροι από τις εικόνες (κυρίως τηλεοπτικές) της προεκλογικής καμπάνιας.

ii. Η επιλογή της τελευταίας στιγμής (last minutes voting). Με βάση τα exit polls των δύο τελευταίων εκλογικών αναμετρήσεων, υπολογίζεται ότι 7-8% του εκλογικού σώματος, περίπου 500.000 ψηφοφόροι, κάνουν την τελική τους επιλογή κυριολεκτικά μέσα στο παραβάν του εκλογικού κέντρου. Ο χρόνος τελικής επιλογής είναι αντιστρόφως ανάλογος τόσο της ηλικίας του ψηφοφόρου (όσο πιο νέος, τόσο πιο κοντά στην κάλπη αποφασίζει), όσο και της κομματικής τοποθέτησής του (όσο πιο μικρό το κόμμα που στηρίζει, τόσο πιο αργά κάνει την τελική επιλογή).

iii. Η αποϊδεολογικοποιημένη επιλογή, η οποία βασίζεται σε κριτήρια που δεν έχουν πλέον σχέση με την παραδοσιακή ιδεολογική ταύτιση και κομματική πειθαρχία προηγούμενων γενεών.

iv. Η θεματική επιλογή (issue voting), η οποία διαμορφώνεται από κριτήρια που κυριαρχούν στη θεματολογία των ΜΜΕ και όχι απαραίτητα από τη θεματολογία που προσπαθούν να επιβάλλουν τα κόμματα την περίοδο των εκλογών (agenda setting). Για παράδειγμα, η ανεργία, η παιδεία, η υγεία και η εγκληματικότητα, ή αλλιώς τα «προβλήματα της καθημερινότητας», κυριάρχησαν στις εκλογές του 2000,

εξοστρακίζοντας από την εκλογική ατζέντα θέματα, όπως η ONE (ολογράφως), γιατί συνέπιπταν με τα κοινωνικά θέματα που αποτελούσαν το «κύριο μενού» των τηλεοπτικών δελτίων ειδήσεων, τα οποία αποτελούν πλέον τον ισχυρότερο διαμορφωτή της κοινής γνώμης.

Είναι χαρακτηριστικό και δεν πρέπει να υποτιμάται το γεγονός, ότι παρά την κρίση που προκάλεσε το τρομοκρατικό χτύπημα της 11ης Σεπτεμβρίου στις ΗΠΑ και τον επακόλουθο πόλεμο στο Αφγανιστάν, η προσοχή της κοινής γνώμης μετά από μια σύντομη περίοδο μερικών ημερών, οπότε και εγκλωβίστηκε από τον καταγιγισμό των διεθνών ειδήσεων, στράφηκε στις παραδοσιακές της επιλογές, τα reality shows και το ποδόσφαιρο!

Τέλος, ένα ακόμη σημαντικό στοιχείο είναι ότι το ποσοστό των πολιτών που «δε νοιώθει απολύτως βέβαιο τι θα ψηφίσει», είναι υπερδιπλάσιο του ποσοστού των πολιτών που δηλώνουν «αναποφάσιστοι». Αυτό τι σημαίνει; Ότι εκτός από τους παραδοσιακούς «αμφιταλαντευόμενους», υπάρχουν άλλοι τόσοι «ευάλωτοι», δηλαδή ψηφοφόροι, που κατά τη διάρκεια της καμπάνιας υπάρχει σοβαρό ενδεχόμενο να αλλάξουν κομματική προτίμηση. Το ποσοστό αυτό είναι ακόμα ένα αποδεικτικό μιας αυξημένης ρευστότητας στη δυναμική της εκλογικής συμπεριφοράς.

Κριτήρια ψήφου, λοιπόν, τόσο διαφορετικά από αυτά των αρχών της προηγούμενης δεκαετίας, κάνουν επιτακτική τη διερεύνηση του δεύτερου παρανομαστή εκλογικής συμπεριφοράς, την «ποιοτική μετάλλαξη» των ψηφοφόρων. Μια μετάλλαξη, η οποία σε συνδυασμό με τη φθίνουσα πορεία των παραδοσιακών κριτηρίων ψήφου, οδηγεί τα παραδοσιακά μοντέλα ανάλυσης σε απορρύθμιση. Υπάρχει κάποια μεθοδολογία ποιοτικής ανάλυσης της συμπεριφοράς των ψηφοφόρων που μπορεί να έρθει και να καλύψει αυτό το διαφαινόμενο κενό που παρατηρείται; Αυτό θα προσπαθήσουμε να καλύψουμε στο επόμενο κεφάλαιο, περιγράφοντας και αναλύοντας τη Θεωρία των Αποφάσεων και την Πολυκριτήρια Ανάλυση ως το μεθοδολογικό εκείνο υπόβαθρο που θα μας οδηγήσει στην “Εξόρυξη Δεδομένων” ως την προτεινόμενη μεθοδολογία ποιοτικής ανάλυσης της συμπεριφοράς των ψηφοφόρων.

ΚΕΦΑΛΑΙΟ 3:

Μεθοδολογικό Υπόβαθρο

Η Θεωρία των Αποφάσεων & η Πολυκριτήρια Ανάλυση

3.1 Μία περιγραφική προσέγγιση της θεωρίας των αποφάσεων

Ενώ ο Weber (1954) και άλλοι εκπρόσωποι της κλασικής ή παραδοσιακής προσέγγισης στο πλαίσιο της θεωρίας των οργανώσεων (Fayo, Taylor, Gulick, Urwick, M.P.Follett) ενασχολήθηκαν κυρίως με τον εξορθολογισμό της οργανωτικής δομής, πιο πρόσφατες κατευθύνσεις ή προσεγγίσεις ερευνούν άλλες όψεις της οργανωτικής συμπεριφοράς, όπως τη διαδικασία λήψης των αποφάσεων και επιλογής αξιών. Κύριος εκπρόσωπος της προσέγγισης αυτής υπήρξε ο Hebert Simon (1960) που θεώρησε ότι το «αποφασίζει» διατρέχει το σύνολο της οργανωτικής συμπεριφοράς και οφείλει, κατά συνέπεια, να αποτελέσει κύριο αντικείμενο μελέτης της διοικητικής επιστήμης.

Η διαδικασία λήψης αποφάσεων μπορεί να έχει, και έχει, μελετηθεί στο πλαίσιο της διοικητικής επιστήμης, είτε από την άποψη της δύναμης και της εξουσίας (ποιος και τι αποφασίζει) είτε από την άποψη της εκλογίκευσης και του ορθολογισμού των αποφάσεων (πώς αποφασίζει κανείς, βελτιστοποίηση των αποφάσεων) – χωρίς αυτές να είναι οι μόνες δυνατές οπτικές ή προσεγγίσεις ούτε αναγκαστικά αμοιβαία αποκλειόμενες, όπως θα φανεί στη συνέχεια.

Ορισμένες θεωρήσεις επεξεργάζονται το πρότυπο ή ιδεότυπο του πλήρως ορθολογικά δρώντος (Homo Economicus), που περιλαμβάνει τη δυνατότητα εξειδίκευσης και συγκεκριμενοποίησης των επιδιωκόμενων στόχων και αξιών, την κατάταξή τους σε σαφή κλίμακα προτιμήσεων καθώς και την ικανότητα της επακριβούς εκτίμησης όλων των επιπτώσεων και συνεπειών καθεμιάς από τις εναλλακτικές επιλογές που γίνονται αντιληπτές για την επίτευξη των στόχων και την προώθηση των αξιών. Σε αυτή την περίπτωση, ο δρων λειτουργεί σύμφωνα με τη μεγιστοποιητική αρχή (minimax principle), επιδιώκει, δηλαδή, την ανεύρεση της άριστης επιλογής (optimal), εκείνης που οδηγεί ή επιφέρει τη μεγιστοποίηση των στόχων ή του οφέλους, ελαχιστοποιώντας ταυτόχρονα το σχετικό κόστος ή θυσία. Το πρότυπο του ορθολογικά δρώντος, σύμφωνα μ' έναν σχολιαστή, καθορίζει σε σημαντικό βαθμό τις αποφάσεις (διαγραμματικών πλεγμάτων) του οικονομολόγου, και μεγάλο μέρος της οικονομικής θεωρίας και πρακτικής και είναι θεμελιωμένο στην ορθολογική παραδοχή της συμπεριφοράς, δηλαδή ο άνθρωπος συνήθως επιδιώκει τη μεγιστοποίηση της ικανοποίησής του και πράττει ανάλογα σε δεδομένες οικονομικές περιστάσεις.

Αν το πρότυπο ή το παράδειγμα του εξορθολογισμού διέπει τουλάχιστον κατά το ήμισυ την επικράτεια των επιστημών της συμπεριφοράς, υπόκειται όμως σε

σθεναρή αντίσταση από μέρους όλων εκείνων των θεωρήσεων ή αντιλήψεων στην κοινωνική ψυχολογία που έλκουν τη προέλευσή τους από το φροϋδικό έργο και υπάγουν το γνωστικό στο υποσυνείδητο ή θυμικό επίπεδο της ανθρώπινης συμπεριφοράς (Homo Psychologicus ή Affectivus).

Στο ερμηνευτικό αυτό δίπολο, τα ακραία σημεία του οποίου ορίζονται από το ορθολογικό και το ανορθολογικό παράδειγμα αντίστοιχα, ο Simon (1960) φιλοδόξησε να παρεμβάλει τη δική του συνθετική άποψη για το Homo Administrativus που επιδιώκει να αποφασίζει και ενεργεί ορθολογικά, μέσα όμως στα όρια του εφικτού και του δυνατού. Ο Homo Administrativus συμπεριφέρεται κατά το δυνατόν ορθολογικά, και έτσι διαφοροποιείται τόσο από τον εξ ορισμού απόλυτα ορθολογικό homo Economicus όσο και προς τον «αρνητικό» του τελευταίου Homo Psychologicus ή Affectivus.

Το κύριο, κατά συνέπεια, αντικείμενο της μελέτης της διοικητικής επιστήμης οφείλει να είναι, κατά τον Simon (1960), η διερεύνηση των ορίων μεταξύ των ορθολογικών και ανορθολογικών στοιχείων της ανθρώπινης συμπεριφοράς. Η επιδιωκόμενη, αλλά περιορισμένη ορθολογικότητα του Homo Administrativus έχει ως συνέπεια, μεταξύ άλλων, την από μέρους του – και συνεκδοχικά και από τη διοικητική πρακτική – υιοθέτηση κριτηρίων που αν δεν μεγιστοποιούν το όφελος (όπως θα ήταν ίσως δυνατόν, αν δεν υπήρχαν όρια στην ανθρώπινη λογική), τουλάχιστον επιδρούν βελτιωτικά στην υφιστάμενη κατάσταση.

Ο Homo Administrativus επιλέγει εκείνες τις εναλλακτικές εκδοχές που φαίνονται ικανοποιητικές ή αρκετά καλές (satisfactory or good enough), αδυνατώντας να εντοπίσει την άριστη επιλογή. Αυτό που ο John Rawls (1971) έχει αποκαλέσει στοχαστική ή επιδιωκόμενη λογικότητα (deliberative rationality) – την αναζήτηση, δηλαδή, ικανοποιητικών ή βελτιωτικών επιλογών – δεν περιλαμβάνει εξαντλητική θεώρηση της πραγματικότητας σε όλη της την πολυπλοκότητα ούτε μελέτη και αξιολόγηση όλων των πιθανών εναλλακτικών επιλογών. Ο Rawls (1971) συμφωνεί κατ' ουσίαν με τον Simon (1960) στο ότι είναι ασυνήθης στην κοινή ανθρώπινη εμπειρία η αναζήτηση τέλει λύσεων ή επιλογών που μεγιστοποιούν το όφελος ή το επιδιωκόμενο αποτέλεσμα (optimizing).

Είναι αντίθετα συνεπέστερη προς την πραγματικότητα η υπόθεση ότι συνήθως οι άνθρωποι, όταν πρόκειται να λάβουν (δύσκολες) αποφάσεις είναι ευτυχείς, αν εντοπίσουν μια ικανοποιητική λύση ή επιλογή που να ανταποκρίνεται σε ορισμένα ελάχιστα μέτρα ή κριτήρια (βελτιοδοξία). Η υιοθέτηση, αντίθετα, κριτηρίων που μεγιστοποιούν το όφελος και μόνο φαίνεται μάλλον εξωπραγματική, δεδομένου ότι η απεριόριστη αναζήτηση της τέλει επιλογής συνεπάγεται τεράστιο κόστος. Είναι γεγονός ότι συνήθως οι άνθρωποι εγκαταλείπουν την παραπέρα έρευνα και αναζήτηση, εφόσον τα αναμενόμενα αποτελέσματα ακυρώνονται από το κόστος επίτευξής τους. Στον άνθρωπο ως έλλογο ον δεν απουσιάζει η αίσθηση και επίγνωση των ορίων της λογικότητάς του και του πεπερασμένου ενεργειακού του δυναμικού. Αποφεύγει συνεπώς να κατασπαταληθεί στην προσπάθεια επίτευξης του αδύνατου.

Ο Simon (1960) μάλιστα ισχυρίζεται ότι η επιδιωκόμενη, αλλά πεπερασμένη λογικότητα (bounded rationality) του Homo Administrativus μπορεί να έχει συνέπεια

την από μέρους του υιοθέτηση μιας δραστικά ίσως απλουστευμένης εικόνας της πολυπλοκότητας του πραγματικού κόσμου, ώστε έτσι να καθίσταται δυνατή η αποτελεσματική παρέμβασή του και ο έλεγχος αυτής της πραγματικότητας.

3.2 Το ορθολογικό παράδειγμα

Ως απόφαση ο Simon (1960) θεωρεί τη (πολύπλοκη και σύνθετη) διαδικασία της επιλογής μεταξύ εναλλακτικών που παρουσιάζονται ως περισσότερο ή λιγότερο πρόσφορες για την επίτευξη ορισμένων στόχων. Ειδικότερα, ο Simon (1960) χρησιμοποιεί αδιακρίτως τους όρους «απόφαση» ή «επιλογή» για να περιγράψει τη διαδικασία εκείνη σύμφωνα με την οποία από ένα σύνολο πιθανών εναλλακτικών λύσεων για την επίτευξη ενός στόχου επιλέγεται εκείνη που, τελικά, υλοποιείται.

Προκύπτει από τα παραπάνω ότι η απόφαση ως επιλογή μεταξύ εναλλακτικών λύσεων αποκτά νόημα – και μπορεί κατά συνέπεια να αξιολογηθεί λογικά – στο μέτρο που είναι προσανατολισμένη στην επίτευξη ορισμένων στόχων. Είναι, επομένως, σαφές ότι μόνο το πρόβλημα της ορθολογικής απόφασης και επιλογής ανακύπτει, προκειμένου περί σκόπιμης (δηλαδή προσανατολισμένης στην επίτευξη σκοπών) δράσης. Αν, αντίθετα, απουσιάζει το κριτήριο της σκοπιμότητας, η δράση απογυμνώνεται από το «λογικό» περιεχόμενό της και γίνεται νοηματικά απροσπέλαστη από διοικητικής σκοπιάς. Η σκοπιμότητα, επομένως, όχι μόνο «προσανατολίζει» τη διοικητική δράση, αλλά αποτελεί επίσης προϋπόθεση της λογικότητάς της και κριτήριο της αξιολόγησής της (π.χ. βαθμός επίτευξης στόχων – τελεολογική συνέπεια).

Στη διαδικασία της απόφασης η επιλογή αναφέρεται σε εναλλακτικές μεθόδους για την επίτευξη ορισμένων στόχων. Δυνατότητα επιλογής υφίσταται, επομένως, μόνο εάν διαφαίνονται περισσότερες από μία εναλλακτικές λύσεις, προτάσεις, μέθοδοι προσέγγισης των στόχων που επιδιώκονται. Ορθολογική είναι η απόφαση που οδηγεί, καταλήγει στην επιλογή της εναλλακτικής πρότασης που συνεπάγεται την αποτελεσματικότερη επίτευξη των στόχων συγκριτικά με τις λοιπές εναλλακτικές επιλογές στις δεδομένες συνθήκες. Αν, λόγου χάρη, προκειμένου να μεταβεί κανείς από την Αθήνα στη Θεσσαλονίκη διαφαίνονται τρεις (3) ή τέσσερις (4) εναλλακτικές λύσεις (πεζή, οδικώς, αεροπορικώς, ακτοπλοϊκώς), ορθολογική θα είναι η επιλογή εκείνης της εναλλακτικής μεθόδου που θα επιτυγχάνει το στόχο αποτελεσματικότερα από τις άλλες.

Το κριτήριο, βέβαια, της αποτελεσματικότητας συναρτάται με τη διερεύνηση των συνεπειών κάθε εναλλακτικής επιλογής, καθώς και με τους διαθέσιμους πόρους από μέρους του αποφασίζοντος, γιατί είναι φανερό ότι κάθε εναλλακτική επιλογή έχει διαφορετικές συνέπειες, κόστος και προϋποθέσεις. Ανάλογα με την πολυπλοκότητα του προβλήματος και της απόφασης είναι φυσικό να πολλαπλασιάζονται οι εναλλακτικές και να δυσχεραίνεται η εκτίμηση των συνεπειών και προϋποθέσεών τους.

Σε τέτοιες ή ανάλογες ακριβώς περιστάσεις ανακύπτει η ανάγκη και η χρησιμότητα ενός προτύπου (μοντέλου) ή θεωρίας απόφασης που να διευκολύνει τον

δρώντα στην απλούστευση, την ταξινόμηση και τον έλεγχο της πολυπλοκότητας που αντιμετωπίζει. Ένα πρότυπο ή θεωρία απόφασης που να είναι πρακτικά χρήσιμο, με την έννοια ότι η εφαρμογή του εξοπλίζει τον δρώντα με τη δυνατότητα να βελτιώσει την ποιότητα των αποφάσεων και των ενεργειών του.

Το ορθολογικό πρότυπο απόφασης ή το παράδειγμα του ορθολογικά δρώντος προβλέπει μια ακολουθία ή σειρά σταδίων ή φάσεων στη διαδικασία της (ορθολογικής) απόφασης, που στη διαδοχή και διαπλοκή τους συμβάλουν στη βελτιστοποίηση της απόφασης και τη μεγιστοποίηση της επίτευξης των επιδιωκόμενων σκοπών ή στόχων.

Από μία πληθώρα εκδοχών αναλυτικών παραδειγμάτων ορθολογικής απόφασης διαλέγουμε να περιγράψουμε σύντομα ένα συνθετικό πρότυπο που προϋποθέτει τη συστημική οπτική και μεθοδολογία και έχει εφαρμογή στο χώρο της δημόσιας πολιτικής. Σύμφωνα με το πρότυπο αυτό στη διαδικασία της ορθολογικής απόφασης περιλαμβάνονται:

- 1) Δόμηση ή αποτύπωση του προβλήματος, με την έννοια της καταγραφής και εννοιολόγησης της προβληματικής κατάστασης και των κυριότερων συνιστωσών της. Η ορθή δόμηση του προβλήματος κρίνεται ως ιδιαίτερης σημασίας, δεδομένου ότι συχνά η αναποτελεσματική ή άστοχη ενέργεια οφείλεται στη λανθασμένη τοποθέτηση του προβλήματος. Μπορεί εξάλλου να υποστηριχθεί ότι είναι προτιμότερο να αντιμετωπίζεται έστω και ατελώς το ορθό (κρίσιμο) πρόβλημα παρά να «επιλύεται ορθά» το λάθος πρόβλημα. Ο Wildavsky (1962) έχει παρατηρήσει σχετικά ότι είναι προτιμότερο να πάρει κανείς το αργό τρένο προς τη σωστή κατεύθυνση παρά το γρήγορο τρένο προς τη λάθος κατεύθυνση. Η αποτύπωση του προβλήματος περιλαμβάνει την αναγνώριση και καταγραφή των κυριότερων από τους παράγοντες ή συντελεστές που μετέχουν, δηλαδή υφίσταται άμεσα τις συνέπειες, στην προβληματική κατάσταση και των αξιολογικών επιδιώξεών τους, των μεταξύ τους αλληλεξαρτήσεων και των συνεπειών και επιπτώσεων από τις ενέργειες ή τις επιδιώξεις τους.
- 2) Διερεύνηση και αναζήτηση των εναλλακτικών λύσεων του προβλήματος. Στο στάδιο αυτό επιχειρείται η εκτενέστερη δυνατή χαρτογράφηση των πιθανών αντιμετώπισεων του προβλήματος, με τη χρησιμοποίηση όλων των διαθέσιμων μεθόδων και τεχνικών που μπορεί να περιλαμβάνουν ακόμα και την ενόραση ή τη διαίσθηση.
- 3) Επιλογή σκοπού, αξίας και εναλλακτικής λύσης από το σύνολο των προοπτικών που έχουν ήδη καταγραφεί με τη χρησιμοποίηση διαφόρων κριτηρίων (αποτελεσματικότητα, αποδοτικότητα, δικαιοσύνη) και μεθόδων ή τεχνικών (όπως ανάλυση κόστους – ωφέλειας, ταξινόμηση στόχων και σύνθεση αξιών).
- 4) Εκτέλεση και υλοποίηση της εναλλακτικής πρότασης που επιλέχθηκε με τη λειτουργία μηχανισμών παρακολούθησης (monitoring) του τρόπου και της διαδικασίας εφαρμογής.

- 5) Αξιολόγηση, ανατροφοδότηση πληροφοριών, έλεγχος, απολογισμός και διορθώσεις.

Στο καταληκτικό αυτό στάδιο της ορθολογικής απόφασης, σύμφωνα με το συστημικό πρότυπο, διερευνάται και εντοπίζεται ο βαθμός ή το μέτρο απόκλισης μεταξύ επιδιωκόμενων στόχων και συγκεκριμένων αποτελεσμάτων, επισημαίνονται τα αίτια των αποκλίσεων, αξιολογείται αν πράγματι επιλύθηκε ή μάλλον κατά πόσο αντιμετωπίστηκε το πρόβλημα, εξετάζονται οι συνέπειες και οι επιπτώσεις τόσο βραχυπρόθεσμα όσο και μακροπρόθεσμα, στο άμεσο και στο ευρύτερο περιβάλλον, των επιλογών που υιοθετήθηκαν.

Το συστημικό – ορθολογικό πρότυπο της απόφασης λειτουργεί κυκλικά, με την έννοια ότι η διάκριση μεταξύ των φάσεων ή σταδίων της διαδικασίας είναι σχετική, αναλυτική, δεδομένων των αλληλεπιδράσεων και εξαρτήσεων που υφίσταται μεταξύ τους.

Αποτελεί κοινό τόπο όλων των παραπάνω εκδοχών του ορθολογικού παραδείγματος ότι με τον τρόπο της διαδοχικής μετάβασης από το ένα στάδιο στο άλλο καθίσταται δυνατή η βελτιστοποίηση της ποιότητας των αποφάσεων και η μεγιστοποίηση της επίτευξης των επιδιωκόμενων στόχων. Η ορθολογική εξάλλου μεθοδολογία της απόφασης προσιδιάζει στην αντιμετώπιση κρίσιμων ή ουσιωδών προβλημάτων στο χώρο της οργανωτικής – διοικητικής και κοινωνικής πραγματικότητας, όταν, δηλαδή, εμπλέκονται πολλοί παράγοντες ή οπτικές στο πρόβλημα και παρατηρείται αντίθεση ή (σχετική) ασυμβατότητα μεταξύ των διάφορων αξιών, αναγκών και συνεπειών.

Η αριστοποιητική (optimizing) οπτική ή διάθεση που φαίνεται να διαπερνά και να χαρακτηρίζει το ορθολογικό παράδειγμα της απόφασης, είναι φυσικό να υπενθυμίζει ορισμένα γνωρίσματα αυτού που περιγράφηκε παραπάνω ως Homo Economicus. Ο τελευταίος επιδιώκει τη μεγιστοποίηση του οφέλους από τις ενέργειές του, ελαχιστοποιώντας ταυτόχρονα το κόστος που αυτές συνεπάγονται. Υιοθετεί, δηλαδή, κριτήρια αριστοποιητικά, εκζητεί τη βέλτιστη (optimal) επιλογή μεταξύ όλων των πιθανών λύσεων ή προσεγγίσεων στο πρόβλημά του. Η διασύνδεση του ορθολογικά δρώντος και αποφασίζοντος με τον ιδεότυπο του Homo Economicus δεν αποτελεί απλώς μια θεωρητική ιδέα.

Το κριτήριο της διοικητικής αποδοτικότητας (efficiency), εξάλλου, προϋποθέτει μια ανάλογα ορθολογική (βελτιστοποιητική) προσέγγιση από μέρους του δρώντος, δεδομένου ότι υπαγορεύει την επιλογή εκείνης της εναλλακτικής λύσης που συνεπάγεται τα καλύτερα δυνατά αποτελέσματα με σταθερό κόστος ή, εναλλακτικά, διατηρώντας σταθερά αποτελέσματα, τη μεγαλύτερη δυνατή μείωση του απαιτούμενου κόστους για την επίτευξή τους. Είναι, ίσως, προφανές ότι το κριτήριο της αποδοτικότητας είναι πληρέστερα (ή αποκλειστικά) εφαρμόσιμο σε οργανώσεις που όχι μόνο διαθέτουν ποσοτικά μετρήσιμες εισροές και εκροές, αλλά και χρηματικά μεταλλάξιμες.

Από τα παραπάνω προκύπτει συνεπώς ότι η διασύνδεση του ορθολογικού παραδείγματος με την οικονομική δραστηριότητα δεν είναι αβάσιμη ή αδικαιολόγητη. Ορισμένοι μάλιστα σχολιαστές μιλούν για την «οικονομική

προσέγγιση» ή ακόμα και για «εισβολή» της οικονομικής θεωρίας και των ποσοτικών τεχνικών στην ανάλυση της δημόσιας πολιτικής και των αποφάσεων.

Στον προσδιορισμό της έννοιας της ορθολογικής συμπεριφοράς περιλαμβάνεται, όπως έχει ήδη αναφερθεί, η επιλογή της άριστης (optimal) εναλλακτικής προοπτικής δράσης, εκείνης, δηλαδή, της οποίας οι συνέπειες μεγιστοποιούν το επιδιωκόμενο όφελος, ελαχιστοποιώντας παράλληλα το σχετικό κόστος. Η μεγιστοποιητική αυτή προσέγγιση προϋποθέτει, όμως, ορισμένες συνθήκες ή όρους που οφείλουν να διευκρινιστούν και να διαπιστωθεί στη συνέχεια κατά πόσον πρόκειται για πραγματοποιήσιμες συνιστώσες της ορθολογικής συμπεριφοράς.

Συγκεκριμένα, η ορθότητα του ορθολογικού παραδείγματος εξαρτάται από την ικανότητα, από μέρους του δρώντος, για μια συνοπτική ή πανοραμική θεώρηση ενός μεγάλου αριθμού εναλλακτικών τρόπων ή μεθόδων δράσης και αντιμετώπισης του συγκεκριμένου προβλήματος. Υποτίθεται ότι η ποιότητα της απόφασης συνδέεται άμεσα με τον αριθμό των εναλλακτικών που εξετάζονται πριν από την οριστικοποίηση των επιλογών και την ποσότητα των πληροφοριών που αφομοιώνονται σχετικά με τις συνέπειες και τις επιπτώσεις τους.

Σχετική είναι και η υπόθεση ότι ο δρών είναι σε θέση να μελετήσει και να εκτιμήσει τις συνέπειες και τις επιπτώσεις καθεμιάς από τις εναλλακτικές επιλογές, από τη σκοπιά του κόστους και οφέλους σε σχέση με τα επιδιωκόμενα αποτελέσματα. Αυτό εξάλλου προϋποθέτει την από μέρους του δρώντος κατάρτιση σαφούς κλίμακας αξιών και προτιμήσεων αναφορικά με την οποία αξιολογούνται οι εναλλακτικές προτάσεις, ανάλογα με το βαθμό χρησιμότητάς τους. Μεταξύ των εφαρμογών του ορθολογικού παραδείγματος στη δημόσια διοίκηση ενδεικτικά αναφέρονται οι (οικονομικής προέλευσης) μέθοδοι της ανάλυσης κόστους και οφέλους ή αποτελεσματικότητας (cost benefit και cost effectiveness analysis), η επιχειρησιακή έρευνα (operational research) καθώς και τα σχεδιαστικά – προϋπολογιστικά συστήματα που χρησιμοποιήθηκαν στις ΗΠΑ, κυρίως κατά τη δεκαετία του 1960.

Τα ερωτήματα που έχουν διατυπωθεί σχετικά με τη χρησιμότητα των διαφόρων ορθολογικών ή ορθολογιστικών μεθόδων και τεχνικών στο πλαίσιο του αναλυτικού παραδείγματος, αναφέρονται κυρίως στον ποιοτικό χαρακτήρα των δημόσιων αγαθών και υπηρεσιών (όπως λ.χ υγεία, παιδεία, ασφάλεια, ανάπτυξη), η οικονομική μέτρηση, η ποσοτικοποίηση και η πλήρης υπαγωγή των οποίων σε απλουστευτικά κριτήρια εκτίμησης και αξιολόγησης δεν είναι ούτε εύκολη ούτε επιθυμητή. Δευτερευόντως, αναφέρονται και στο πολυδιάστατο, την πολυκεντρικότητα και την πολυπλοκότητα της δημόσιας πολιτικής και αποφάσεων που συχνά αντικατοπτρίζουν συνθέσεις σχετικά αντιτιθέμενων αξιών, αγαθών και συμφερόντων (όπως λ.χ η προστασία του περιβάλλοντος και η οικονομική ανάπτυξη).

Επισημαίνεται σχετικά η σαφής διαφοροποίηση του ατομικού από το συλλογικό πεδίο απόφασης και συμπεριφοράς και υπογραμμίζεται η πολιτική διάσταση, οι πολιτικές συνιστώσες των προβλημάτων της δημόσιας πολιτικής. Ενώ στο προσωπικό επίπεδο είναι ίσως νοητή και θεμιτή η υπόθεση μιας σαφούς κλίμακας αξιών και προτιμήσεων, στο συλλογικό πεδίο, κάτι αντίστοιχο είναι μάλλον εξωπραγματικό, δεδομένης της έκτασης και ποικιλίας των μορφών κοινωνικής

διαστρωμάτωσης και διαφοροποίησης. Από αυτή την πραγματικότητα εκπηγάξει σε μεγάλο βαθμό το πολυδιάστατο, η σχετική απροσδιοριστία και ο ανοιχτός ή και αντιφατικός χαρακτήρας των δημοσίων σκοπών και αποφάσεων.

Ακόμη όμως και στο προσωπικό, ατομικό επίπεδο μπορεί να παρατηρηθεί ότι είναι ασυνήθης η εξαντλητική παράθεση και διερεύνηση εναλλακτικών τρόπων ή μεθόδων δράσης, οι οποίες αξιολογούνται στη βάση συγκεκριμένης κλίμακας αξιών και προτεραιοτήτων. Είναι μάλλον συνεπέστερη προς την ανθρώπινη εμπειρία η παρατήρηση ότι σπάνια οι άνθρωποι επιδιώκουν σαφείς αξίες και μάλιστα με αριστοποιητικά κριτήρια. Αντίθετα, αποφεύγουν να αναλώνονται στην προσπάθεια της ανεύρεσης της τέλει επιλογής και υιοθετούν κριτήρια που απέχουν πολύ από τον μεγιστοποιητικό χαρακτήρα που υπαγορεύεται από το ορθολογικό παράδειγμα.

Η ουσιώδης κριτική που ενασκείται στην αναλυτική μεθοδολογία αφορά τον ουτοπικό ή ανεδαφικό της χαρακτήρα, με την έννοια ότι δεν αντιστοιχεί στη συνήθη πρακτική και εμπειρία είτε στο ατομικό είτε στο συλλογικό πεδίο δράσης και συμπεριφοράς. Το ορθολογικό παράδειγμα συνιστά προφανώς κανονιστική (δεοντολογική) πρόταση και είναι ελάχιστα περιγραφικό της πραγματικής κατάστασης του αποφασίζοντος, οι βασικές δυσκολίες του οποίου εκπηγάζουν από την αντιφατικότητα των αξιολογικών επιδιώξεων, την πρακτική σύγχυση ή σύμφυση σκοπών και μέσων, την αβεβαιότητα των συνεπειών και επιπτώσεων και τη γενική ρευστότητα του περιβάλλοντος.

3.3 Το αυξητικό πρότυπο απόφασης

Η ριζοσπαστικότερη κριτική στο αναλυτικό – ορθολογικό παράδειγμα (το οποίο μάλιστα ο Lindblom (1979) αποκαλεί «συννοπτικό ιδανικό») έχει ενασκηθεί από την αυξητική (incremental) θεωρία της απόφασης, οι γνωσιολογικές θεμελιώσεις της οποίας ανευρίσκονται στην «αντισυννοπτική» διανοητική παράδοση όσο και σε αντίστοιχα, πλουραλιστικά, πολιτικά πλαίσια κοινωνικής – συλλογικής συμπεριφοράς.

Υπενθυμίζεται ότι το αναλυτικό πρότυπο προβλέπει ότι προκειμένου να καταστεί ορθολογική η διαδικασία του αποφασίζειν οφείλει να διέρχεται ορισμένα στάδια ή φάσεις που περιλαμβάνουν τον καθορισμό των στόχων ή αξιολογικών επιδιώξεων, την εξαντλητική θεώρηση εναλλακτικών λύσεων ή επιλογών και την υιοθέτηση εκείνης της εναλλακτικής που μεγιστοποιεί τη πιθανότητα επίτευξης των στόχων.

Στη σειριακή αυτή θεώρηση της διαδικασίας του αποφασίζειν, η αυξητική μεθοδολογία προβάλλει καταρχήν το ερώτημα και την αμφισβήτηση για τη διαιρετότητα και διακρισιμότητα αφ' ενός μεν στόχων και μέσων, αφ' ετέρου δε αξιολογικών και πραγματικών δεδομένων, και την υπαγωγή των δευτέρων στα πρώτα.

Από τη σκοπιά των συνεπειών, εξάλλου, η εφαρμογή του ορθολογικού παραδείγματος μπορεί να επιφέρει ριζικές μετατοπίσεις ή αναθεωρήσεις της υφιστάμενης κατάστασης (Root Method), ενώ αντίθετα στο πλαίσιο της αυξητικής

μεθοδολογίας μόνο επιμέρους ή περιορισμένης έκτασης και χαρακτήρα μεταβολές του status quo επιτυγχάνονται (Branch Method).

Σύμφωνα με τον κορυφαίο εκπρόσωπο της αντισυνοπτικής θεώρησης στο χώρο της διοικητικής επιστήμης Charles Lindblom (1979), το ορθολογικό – συνοπτικό πρότυπο χαρακτηρίζεται από ορισμένες βασικές δυσχέρειες ή δυσκαμψίες, όπως:

- 1) έλλειψη προσαρμογής στις περιορισμένες γνωστικές ικανότητες και δυνατότητες του ανθρώπου στην ανεπάρκεια των πληροφοριών και το κόστος των αναλύσεων,
- 2) αδυναμία σύνθεσης ικανοποιητικής αξιολογικής μεθόδου από άτομα και ομάδες,
- 3) πραγματική διασύνδεση και αλληλεπίδραση μεταξύ των αξιολογικών και πραγματικών δεδομένων
- 4) πολλαπλότητα και πολυπλοκότητα των συνεπειών και των επιπτώσεων κάθε εναλλακτικής επιλογής που είναι σχεδόν αδύνατο να προβλεφθούν με ακρίβεια εκ των προτέρων, και
- 5) πολλαπλότητα των μορφών και των τρόπων που εμφανίζονται τα προβλήματα στο χώρο της δημόσιας πολιτικής.

Η αντισυνοπτική θεώρηση προβάλλει καταρχήν τον ισχυρισμό, σε αντίθεση προς την ορθολογική μεθοδολογία, ότι ο προσδιορισμός των επιδιωκόμενων στόχων και αξιών δεν μπορεί να διακριθεί απόλυτα και να συνιστά προϋπόθεση της εμπειρικής ανάλυσης και εκτίμησης των συνεπειών καθεμιάς από τις ενδεχόμενες εναλλακτικές προτάσεις. Αντίθετα, αξιολογικά και πραγματικά δεδομένα, μέσα και στόχοι συνεξετάζονται και αλληλοκαθορίζονται κατά τη διαδικασία της απόφασης. Η υπόθεση της πρωτοκαθεδρίας των στόχων και της υπαγωγής των μέσων σε αυτούς κρίνεται από τους αντισυνοπτιστές ως ουτοπική και εξωπραγματική. Στην πράξη, δεδομένης της στενότητας των πόρων και του απεριόριστου των αναγκών, μέσα και στόχοι διαπλέκονται με εξαιρετικά περίπλοκους και πολλαπλούς τρόπους. Οι στόχοι της συμπεριφοράς, ισχυρίζονται οι αντισυνοπτιστές, δεν διαμορφώνονται ανεξάρτητα από τα μέσα που διατίθενται για την υλοποίησή τους, αλλά προσαρμόζονται σε αυτά, στο πλαίσιο μιας αμφίδρομης διαπλοκής κατά την οποία οι στόχοι συνεχώς αναπροσδιορίζονται.

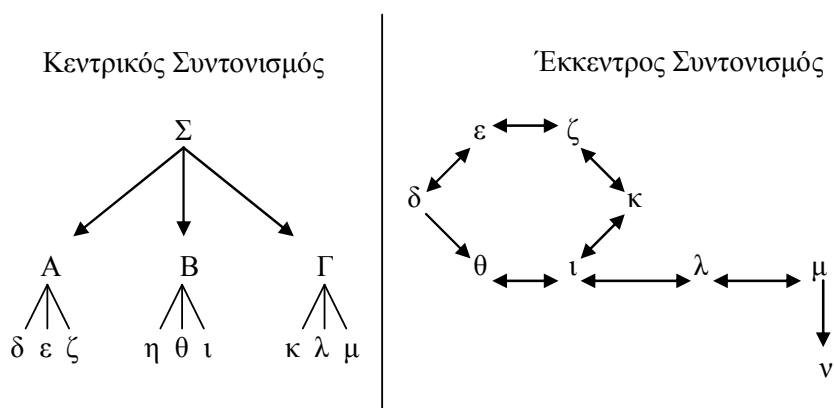
Ένας θεμελιώδης άξονας της αντισυνοπτικής θεώρησης αφορά την πλειονότητα των δρώντων, που μετέχουν στη διαδικασία της απόφασης, κανείς από τους οποίους δεν είναι σε θέση να αναλάβει συνοπτική (ορθολογική), εξαντλητική θεώρηση όλων των ενδεχόμενων εναλλακτικών λύσεων και των συνεπειών τους, ούτε να επιβάλει μονομερώς τη δική του προοπτική στο πρόβλημα. Η υπόθεση της πλειονότητας των δρώντων ή μετόχων του προβλήματος συνεπάγεται εξάλλου τη σχετική διαφοροποίηση των επιμέρους απόψεων ή συμφερόντων του καθενός από αυτούς. Δεν μπορεί συνεπώς να μιλά κανείς για συγκεκριμένο κέντρο αποφάσεων, αλλά μάλλον για έναν αριθμό φαινομενικά ασύνδετων ή ασυντόνιστων «αποφασιστικών» σημείων (fragmented – disjointed decisional points) που διαπλέκονται και αλληλεπιδρούν στη διαδικασία της συνολικής απόφασης. Κανείς

από τους συμμετέχοντες σε αυτή τη διαδικασία δεν είναι σε θέση (από πλευράς τόσο διανοητικών όσο και υλικών δυνατοτήτων) να συλλάβει και να εκτιμήσει το σύνολο των εναλλακτικών λύσεων και των συνεπειών τους.

Στο πιθανό ερώτημα πώς εξασφαλίζεται ο γενικός συντονισμός και η ισορροπία της συμπεριφοράς των δρώντων παρά την απουσία γενικού συντονιστή, ο Lindblom (1979) απαντά ότι ο συντονισμός δεν αποτελεί απόλυτο μέγεθος ούτε αναγκαστικά αντικείμενο κεντρικού ρόλου, αλλά πρακτική συνέπεια (και αποτέλεσμα της εσωτερικής λογικής) της τακτικής των ασύνδετων προσανυξήσεων και των συμπληρωματικών μεταξύ τους προσαρμογών. Ασυντόνιστη είναι η συμπεριφορά εκείνη που τα μέρη της δεν χαρακτηρίζονται από συμπληρωματικότητα.

Ο μηχανισμός των επιμέρους αμοιβαίων προσαρμογών (partisan mutual adjustments) αναφέρεται ακριβώς στη διαδικασία του πρακτικού συντονισμού της συμπεριφοράς όσων συμμετέχουν στη διαδικασία της απόφασης είτε με τη μορφή της απλής προσαρμογής (adaptive adjustment) είτε με τη μορφή της πρόκλησης της επιθυμητής συμπεριφοράς (manipulated adjustments). Είναι προφανές ότι κατά τη διαδικασία των αμοιβαίων προσαρμογών της συμπεριφοράς των επιμέρους δρώντων, οι τελευταίοι (partisans) επιδιώκοντας ο καθένας τους δικούς του σκοπούς εμπλέκονται σε συναλλαγές, διαπραγματεύσεις, αμοιβαίες διεκδικήσεις και υποχωρήσεις. Οι δρώντες τηρούν στάση προσαρμοστική (αποδέχονται και συμμορφώνονται με το υφιστάμενο πλαίσιο – parametric), υποχωρητική (αποφεύγουν ό,τι οι άλλοι δεν ανέχονται – deferential) ή υπολογιστική (ελίσσονται στο υφιστάμενο πλαίσιο – calculated).

Τα είδη ή οι μορφές του κεντρικού και του έκκεντρου (μέσω αμοιβαίων προσαρμογών) συντονισμού μπορεί να παρασταθούν διαγραμματικά ως ακολούθως:



Σχήμα 3.3.1

Όπως διαφαίνεται και στο Σχήμα 3.3.1, ο έκκεντρος συντονισμός χαρακτηρίζεται: α) από μεγαλύτερη οικονομία επιπέδων, εφόσον οι μέτοχοι συντονίζονται μεταξύ τους και έτσι αποφεύγεται η παρεμβολή ενδιάμεσων συντονιστικών βαθμίδων, και β) από μεγαλύτερη ποικιλία σχέσεων διαπλοκής (σύμμετρες και ασύμμετρες). Σύμμετρη είναι η σχέση, όταν υφίσταται

αλληλεπηρεασμός και αλληλεξάρτηση ($\delta \leftrightarrow \varepsilon$). Ασύμμετρη είναι η σχέση, όταν υφίσταται μονομερής επηρεασμός και εξάρτηση ($\mu \rightarrow \nu$).

Στο κυβερνητικό πεδίο μπορεί κανείς να παρατηρήσει ότι συμβαίνει κάτι αντίστοιχο αναφορικά με την πλειονότητα των δρώντων και την έκταση της μεταξύ τους διαπλοκής. Ως δρώντες μπορεί να νοηθούν εδώ μεγάλοι υπηρεσιακοί και οργανωτικοί φορείς, όπως είναι, λόγου χάρη, τα υπουργεία. Πολλά κυβερνητικά προγράμματα ή πολιτικές διαπερνούν το οργανωτικό πλαίσιο ενός αριθμού υπουργείων (π.χ. οικονομική πολιτική) και, αντίστροφα, πολλά από αυτά (υπουργεία) είναι επιφορτισμένα με την προώθηση και υλοποίηση περισσότερων από μίας ή και μερικά αντιτιθέμενων στρατηγικών επιλογών και επιδιώξεων, όπως λ.χ. η οικονομική ανάπτυξη και η προστασία του περιβάλλοντος, η ασφάλεια και η προσωπική ελευθερία, η εκπαίδευση και η απασχόληση.

Η επικέντρωση της προσοχής των δρώντων σε εναλλακτικές που λίγο διαφέρουν από τις ήδη υφιστάμενες επιλογές, αποτελεί εξάλλου συνέπεια του πεπερασμένου της αντιληπτικής και αφομοιωτικής ικανότητας του ανθρώπου, της ανεπάρκειας των απαιτούμενων πληροφοριών και του σχετικού κόστους ή των θυσιών που θα έπρεπε κανείς να υποβληθεί, αν επρόκειτο να καταβάλει την υπεράνθρωπη προσπάθεια για συνοπτική θεώρηση και πλήρη ορθολογισμό. Ο Lindblom (1979) μάλιστα ισχυρίζεται ότι η σκόπιμη ή ενσυνείδητη παραμέληση (deliberate disregard) ενός αριθμού συνεπειών και επιπτώσεων των επιμέρους εναλλακτικών επιλογών συνιστά «αξιοσημείωτη μεθοδολογία απόφασης» (noteworthy problem solving tactic), στο μέτρο που έτσι ελέγχεται η δαπάνη του ήδη περιορισμένου διανοητικού και πληροφοριακού αποθεματικού του αποφασίζοντος.

Ένα άλλο αξιοπρόσεκτο χαρακτηριστικό του αυξητικού προτύπου αφορά τα κριτήρια της απόφασης που, κατ' αντίθεση προς το ορθολογικό παράδειγμα, δεν μπορούν να είναι «αριστοποιητικά», αλλά απλώς «βελτιωτικά – ικανοποιητικά» (satisfying). Δεδομένης της συνθετότητας και πολυπλοκότητας των προβλημάτων, ο αποφασίζων δεν επιδιώκει την εφάπαξ οριστική και «άριστη» επίλυσή τους, αλλά μάλλον την «ικανοποιητική» αντιμετώπιση και διαχείρισή τους με τη διαδικασία της συνεχούς και σταδιακής ανάλυσης (serial analysis) και αξιολόγησης από πολλαπλά κέντρα ή σημεία απόφασης. Στην πρακτική ενάσκηση της δημόσιας πολιτικής, σημασία έχει ο περιορισμός και ο έλεγχος άμεσων και προφανών παθολογιών και όχι τόσο η διατύπωση ιδανικής κατάστασης (remedial orientation). Κατά τους αντισυνοπτιστές, προκειμένου για πολύπλοκα προβλήματα δεν είναι ορθό να μιλά κανείς για την επίλυσή τους, αλλά μάλλον για έλεγχο και αντιμετώπισή τους (problem management). Ο Lindblom (1979) προβάλλει τον ισχυρισμό ότι το αυξητικό πρότυπο συνιστά πρακτική μεθοδολογία έλλογης συμπεριφοράς που, παρά τις πιθανές ατέλειές της, υπερέχει της μάταιης απόπειρας για πλήρη (υπεράνθρωπο) ορθολογισμό που χαρακτηρίζει το ορθολογικό παράδειγμα.

Συμπερασματικά μπορεί να γίνει αποδεκτό ότι το αυξητικό ή προσθετικό πρότυπο είναι περισσότερο πραγματιστικό ή περιγραφικά ακριβέστερο (πώς πράγματι λαμβάνονται οι αποφάσεις) από το συνοπτικό παράδειγμα. Η παραδοχή αυτή στηρίζεται στο γεγονός ότι στο πλαίσιο της αντισυνοπτικής θεώρησης καταδεικνύεται

η πολυπλοκότητα και συνθετότητα των προβλημάτων της δημόσιας πολιτικής και ταυτόχρονα το πεπερασμένο της ανθρώπινης λογικότητας, η πολλαπλότητα των παραγόντων (και αντίστοιχων σημείων επιρροής) που μετέχουν στη διαδικασία της απόφασης, ο ικανοποιητικός ή απλώς βελτιωτικός και όχι αριστοποιητικός χαρακτήρας των κριτηρίων που υιοθετούνται, η απουσία βεβαιότητας σχετικά με τις συνέπειες και επιπτώσεις των εναλλακτικών που γίνονται ορατές από τον αποφασίζοντα, η στενότητα των διατιθέμενων πόρων και βέβαια η πολλαπλότητα, πολυπλοκότητα και η σχετική αντιφατικότητα των επιδιωκόμενων σκοπών και αξιών, που μάλιστα δεν διαχωρίζονται από τα μέσα επίτευξης και ικανοποίησής τους. Στο σημείο αυτό πιθανόν θα πρέπει να διερευνηθεί αν υφίσταται «πολύπλοκο πρόβλημα απόφασης» (complex decision problem), εφόσον ισχύουν οι ακόλουθες συνθήκες:

- i. πλειονότητα των δρώντων ή συμμετεχόντων στη διαδικασία της απόφασης με τις σχετικές μεταξύ τους διαφορές επιμέρους συμφερόντων, προοπτικών και αξιολογικών προσλαμβανουσών· με άλλα λόγια παρατηρούνται πολλαπλές διασυνδέσεις και αλληλεξαρτήσεις τόσο μεταξύ των μερών του ίδιου προβλήματος όσο και με άλλα σχετιζόμενα προβλήματα.
- ii. Διατυπώνονται δύο τουλάχιστον (σχετικά) αντίθετες αξιολογικές διεκδικήσεις (αντίθεση αξιών και υποκειμενικότητα προοπτικών) με συνέπεια η (μερική έστω) ικανοποίηση της μιας να επιφέρει αντίστοιχη μείωση της άλλης.
- iii. Υπάρχει αβεβαιότητα σχετικά με τον χαρακτήρα και τις συνέπειες των εναλλακτικών τρόπων δράσης που γίνονται αντιληπτές από τον αποφασίζοντα· υφίσταται, δηλαδή, μικρός βαθμός προβλεψιμότητας για ενδεχόμενες συνέπειες και επιπτώσεις, ο διαχρονικός, όμως, και πιθανώς μη αντιστρέψιμος χαρακτήρας των οποίων δύσκολα μπορεί να αμφισβητηθεί (διαχρονία και όχι παροδικότητα).

Η αμφισβήτηση του ορθολογικού – αναλυτικού προτύπου και η αντίστοιχη ορθότητα της αντισυννοπτικής θεώρησης διεκδικείται τόσο στο ατομικό όσο και στο συλλογικό πεδίο απόφασης και συμπεριφοράς. Στο ατομικό επίπεδο, το ορθολογικό παράδειγμα έχει υποστεί σοβαρή κριτική και αμφισβήτηση από τη γνωστική ψυχολογία (cognitive psychology) και την κυβερνητική θεωρία της απόφασης, στη βάση κυρίως των εσφαλμένων υποθέσεων των συνοπτιστών για την ικανότητα επεξεργασίας πληροφοριών που διαθέτει ο ανθρώπινος εγκέφαλος. Υπενθυμίζεται ότι αποτελεί θεμελιώδες αξίωμα του ορθολογικού παραδείγματος πως η επίλυση σύνθετων προβλημάτων απαιτεί αντίστοιχης – ανάλογης πολυπλοκότητας μηχανισμούς και ικανότητες από μέρους του αποφασίζοντος. Παρατηρείται, όμως, ότι στη πράξη πολλές φορές σύνθετα προβλήματα και πολύπλοκες λειτουργίες αντιμετωπίζονται και επιτελούνται χωρίς παράλληλα να διαθέτει ο δρων απεριόριστες υπολογιστικές και λογικές δυνατότητες ή να προβαίνει σε εξαντλητική εξέταση όλων των πιθανών εναλλακτικών τρόπων δράσης των συνεπειών και επιπτώσεών τους. Και αυτό ισχύει και για διαφορετικής πολυπλοκότητας οργανισμούς όσο και συμπεριφορές. Οι μέλισσες, λόγου χάρη, επιτελούν την εξαιρετικά πολύπλοκη

λειτουργία της άγρευσης γύρης χωρίς να προβαίνουν σε διαδικασίες αναλυτικού λογισμού. Ο παίχτης της αντισφαίρισης, εξάλλου (;), επιτυγχάνει υψηλά επίπεδα συντονισμού των μελών του χωρίς σχεδόν να σκέπτεται.

Εκείνο που συμβαίνει σε αυτές τις περιπτώσεις, παρατηρεί ο Steinbruner (1974), είναι η χρησιμοποίηση κλειστών κυκλωμάτων ανατροφοδότησης πληροφοριών μέσω των οποίων περιορίζεται η αβεβαιότητα και αποφεύγεται η εξαντλητική θεώρηση εναλλακτικών. Το χαρακτηριστικό παράδειγμα είναι η γάτα που κοιμάται κοντά στην εστία και την πλησιάζει, όταν η θέρμανση χαμηλώνει, ενώ απομακρύνεται, όταν αυτή ανεβαίνει. Σε αυτή την περίπτωση δεν μπορεί να υποθεθεί ότι η γάτα προβαίνει σε διαδικασία απόφασης που προσομοιάζει στο αναλυτικό παράδειγμα. Εκείνο που, κατά πάσα πιθανότητα, συμβαίνει είναι ότι διαθέτει μηχανισμούς που επιτρέπουν τη διατήρηση ορισμένων κρίσιμων μεταβλητών (critical variables) σε ορισμένα επίπεδα ή συσχετισμούς, και αλλάζει θέση μόνο αν οι μεταβλητές αυτές (π.χ. η θερμότητα) υπερβούν τα καθορισμένα ανεκτά όρια – όπως ακριβώς συμβαίνει με το θερμοστάτη. Πρόκειται για απλό μηχανισμό, ικανό όμως να επιτύχει το δύσκολο στόχο της διατήρησης της θερμοκρασίας μέσω της ανατροφοδοτικής λειτουργίας (feedback) που επιτρέπει την επισήμανση μεταβολών στη θερμοκρασία.

Ενδεικτικά παραδείγματα σχετικής αποτυχίας και προσαρμογής στην πραγματικότητα προσπαθειών για συνοπτική (πλήρως ορθολογική) μεθοδολογία στο χώρο της δημόσιας πολιτικής μπορεί να αναφερθούν πρώτον, η συστηματική ιεράρχηση στρατηγικών σκοπών, στόχων, των απαιτούμενων πόρων και των διαφαινόμενων εναλλακτικών επιλογών (Program Analysis and Preview – PAP), που «έζησε» περίπου μια δεκαετία (1970 – 79). Δεύτερον, το γνωστό Think Tank, Central Policy Review Staff, που λειτούργησε ως μηχανισμός στρατηγικής ανάλυσης και σχεδιασμού στη βρετανική δημόσια διοίκηση και καταργήθηκε σχετικά σύντομα. Και τρίτον, οι σχεδιαστικές απόπειρες στο βρετανικό Εθνικό Σύστημα Υγείας που δεν κατόρθωσαν να εκλογικεύσουν την πρακτική των συμβιβασμών, διαπραγματεύσεων και προσαρμογών μεταξύ της σχετικής ασυμβατότητας συμφερόντων και διεκδικήσεων.

Το γενικό συμπέρασμα από την εμπειρία των παραπάνω εφαρμογών είναι ότι απόπειρες ορθολογικής ανάλυσης προσκρούουν σε μεθοδολογικά, οργανωτικά ή πολιτικά εμπόδια που μοιραία οδηγούν στη σταδιακή αδρανοποίηση και αφομοίωσή τους από σταθερές συνήθειες και πρακτικές. Η εμπειρία από την αποτυχία ορισμένων μεθόδων του ορθολογικού τύπου θέτει το ερώτημα κατά πόσο ο αυξητισμός συνιστά τη μόνη πρακτικά δυνατή μεθοδολογία λήψης αποφάσεων.

Από τα δύο αλληλένδετα, ως ένα σημείο τουλάχιστον, ερωτήματα, ανακύπτει ο ειδικότερος προβληματισμός περί της ισχύος του αυξητισμού ανεξάρτητα:

- από το είδος του προβλήματος που αντιμετωπίζεται, και
- από τις συγκεκριμένες περιβαλλοντικές συνθήκες, καταστάσεις και αξιολογικές διεκδικήσεις στο πλαίσιο των οποίων λαμβάνεται η απόφαση.

Όσον αφορά το πρώτο σκέλος του ερωτήματος (είδος του προβλήματος), το ζήτημα είναι αν η αυξητική μεθοδολογία προσιδιάζει για την αντιμετώπιση κάθε μορφής προβλημάτων τόσο πολύπλοκων όσο και απλούστερων στο ατομικό και συλλογικό πεδίο. Υπενθυμίζεται ότι η αντισυνοπτική θεώρηση υιοθετεί τη μεθοδολογία της διάσπασης του προβλήματος στα κύρια μέρη του, καθένα από τα οποία αντιμετωπίζεται με κριτήρια σχετικής ορθότητας (ικανοποιητικά – suboptimal), από μια πληθώρα μετόχων ή δρώντων. Με τον τρόπο, όμως, αυτό αγνοείται ή υποβαθμίζεται η εσωτερική πολυπλοκότητα και ο υπερκείμενος ή περιεκτικός χαρακτήρας ορισμένων προβλημάτων που δεν συνιστούν το τυπικό άθροισμα των μερών τους και συνεπώς δύσκολα αντιμετωπίζονται με τη μεθοδολογία του ασυντόνιστου αυξητισμού. Στην ιατρική παθολογία, λόγου χάρη, η ανάλυση ενός συμπτώματος ανεξάρτητα από άλλα και η αποφυγή διασταυρώσεως των ενδείξεων μπορεί να εμποδίσει την ορθή διάγνωση και τελικά τη θεραπεία. Στο χώρο της δημόσιας πολιτικής, εξάλλου, η διαχείριση προβλημάτων μεγάλης κλίμακας και συνθετότητας, όπως η εθνική άμυνα, η παιδεία, η κοινωνική πρόνοια και υγεία, η οικονομική ανάπτυξη, που χαρακτηρίζονται από εξαιρετική πολυπλοκότητα και πολυκεντρικότητα, απαιτεί συντονισμένη προσπάθεια και ένα μέτρο συνοπτισμού που δεν παρέχεται από την αυξητική μεθοδολογία.

Όσον αφορά το δεύτερο σκέλος του προβληματισμού (περιβαλλοντικές συνθήκες ή καταστάσεις), μπορεί να παρατηρηθεί ότι η υιοθέτηση βελτιωτικών ή ικανοποιητικών κριτηρίων απόφασης, στο πλαίσιο της αντισυνοπτικής θεώρησης, προϋποθέτει μια κατάσταση ή ένα περιβάλλον γενικής σταθερότητας ή και στατικότητας, κοινά αποδεκτών αξιών και επιδιώξεων με αναλογικά ποσοστά εξουσίας και επιρροής, διανεμημένα ισόρροπα μεταξύ των μελών της κοινωνικής ομάδας. Σε τέτοιες συνθήκες, μεγάλα ή κρίσιμα θέματα και διαχωριστικές ιδεολογίες δύσκολα ανακύπτουν. Επιπλέον, η υιοθέτηση κριτηρίων σχετικής ικανοποίησης, που δεν συνεπάγονται ριζική διαφοροποίηση από την υφιστάμενη κατάσταση, κατά λογική συνέπεια σημαίνει ότι είναι κανείς γενικά ικανοποιημένος με την υφιστάμενη κατάσταση και απλώς επιδιώκει να επιφέρει ορισμένες βελτιωτικές τροποποιήσεις σε αυτή. Φαίνεται, συνεπώς, ότι η αυξητική στρατηγική προσιδιάζει στην αντιμετώπιση καταστάσεων που χαρακτηρίζονται από χαμηλά ποσοστά κρίσης και αμφισβήτησης. Αντίθετα, η αντιμετώπιση κρίσιμων προβλημάτων (πόλεμος, σεισμός, πυρκαγιά) με την υιοθέτηση της αυξητικής μεθοδολογίας και ικανοποιητικών κριτηρίων μπορεί όχι μόνο να προκαλεί οργανωτική αδράνεια και απάθεια αλλά και να αποβεί μοιραία.

Η αντιμετώπιση κρίσεων με την αυξητική μεθοδολογία ελέγχεται, εξάλλου, και από πλευράς επιστημολογικής στρατηγικής, δεδομένου ότι είναι δύσκολο ή λογικά αντιφατικό να αποπειράται κανείς την «επίλυση» ή υπέρβαση μιας κρίσης, ενώ παραμένει εντός του πλαισίου των συνθηκών που δημιουργήσαν αυτή την κατάσταση. Ανάλογα βέβαια με την ένταση και το βάθος της κρίσης, εκείνο που απαιτείται είναι μια συνολική στρατηγική υπέρβασης που ίσως συνεπάγεται ρήξη με ορισμένες από τις συνθήκες ή τους παράγοντες που συνέβαλαν στη διαμόρφωσή της.

Συμπερασματικά, μπορεί συνεπώς (όχι) να παρατηρηθεί ότι η αντισυνοπτική θεωρία των αποφάσεων εξαντλεί την όποια κανονιστική της εγκυρότητα προκειμένου

για θέματα που δεν χαρακτηρίζονται ως μεγάλα ή κρίσιμα. Στο μέτρο, όμως, που τέτοια θέματα ή προβλήματα πράγματι ανακύπτουν, ακόμα και στο πλαίσιο κοινωνιών τα γενικά χαρακτηριστικά των οποίων συνιστούν τις εννοιολογικές προϋποθέσεις ισχύος της αυξητικής θεωρίας, και αντιμετωπίζονται με συνοπτική μεθοδολογία, κλονίζεται αναλογικά και η περιγραφική εγκυρότητα της αντισυνοπτικής προσέγγισης.

3.4 Ενδεχόμενη Προσέγγιση

Στο πλαίσιο αυτής της ενότητας θα διερευνηθούν ορισμένες ενδιάμεσες θεωρήσεις της διαδικασίας των αποφάσεων και της δημόσιας πολιτικής που οριοθετούνται μεταξύ του τέλει ορθολογισμού και του καθαρού αυξητισμού. Συγκεκριμένα, θα εξεταστούν το κανονιστικό αριστοποιητικό πρότυπο (normative optimum model) του Dror, η μικτή ανίχνευση (mixed scanning) του Etzioni, και η στρατηγική ανάλυση (strategic analysis) του Lindblom. Το κεφάλαιο θα ολοκληρωθεί με ορισμένες συμπερασματικές κρίσεις και παρατηρήσεις σχετικά με τη σημασία και εγκυρότητα μιας ενδεχομενικής προσέγγισης στη θεωρία των αποφάσεων και την ανάλυση της δημόσιας πολιτικής.

3.4.1 Dror

Το κανονιστικό βελτιστοποιητικό πρότυπο του Dror (1964) επιχειρεί τη διεύρυνση τόσο των ορθολογικών όσο και των εξω-ορθολογικών ή μετα-ορθολογικών συνιστωσών ή συντελεστών στη διαδικασία της απόφασης. Η ουσία του επιχειρήματος του Dror (1964), σε αυτό το σημείο, έγκειται, πρώτον, στην αναγνώριση του γεγονότος ότι τέτοια στοιχεία πράγματι παρεισφρύουν στη διαδικασία της απόφασης και, δεύτερον, στην εκτίμηση της λειτουργικής χρησιμότητάς τους στο μέτρο που όντως συμβάλλουν στη μείωση της αβεβαιότητας του περιβάλλοντος και των απρόβλεπτων συνθηκών. Σε ένα άλλο σχετικό κείμενό του, ο Dror (1964) κάνει επίκληση για την αναγκαιότητα υπέρβασης της «οικονομίστικης» λογικής και την ένταξη στην ανάλυση της δημόσιας πολιτικής περισσότερο δημιουργικών δυνατοτήτων, όπως η φανταστική σκέψη (imaginative, futuristic, speculative thinking) και η έμπειρη ενόραση (trained intuition) μέσα από κατάλληλες μεθόδους.

Ως λογικά στοιχεία της απόφασης, ο Dror (1964) θεωρεί την επιλεκτική (και όχι εξαντλητική) μελέτη και εκτίμηση των εναλλακτικών επιλογών και των συνεπειών τους που γίνονται αντιληπτές από τον αποφασίζοντα, και τη σχετική (και όχι απόλυτη) διευκρίνιση και συγκεκριμενοποίηση των αξιολογικών δεδομένων, των σκοπών και στόχων της συμπεριφοράς.

Το κανονιστικό βελτιστοποιητικό πρότυπο, όπως διαφαίνεται από τα παραπάνω, χαράζει μια ενδιάμεση πορεία μεταξύ της συνοπτικής και της αντισυνοπτικής προσέγγισης. Συγκεκριμένα, ενώ ο Dror (1964) αναγνωρίζει την

περιγραφική εγκυρότητα της αυξητικής μεθοδολογίας, επιδιώκει ταυτόχρονα την ενίσχυση και βελτίωση της ποιότητας των αποφάσεων – ένα προβληματισμό που συμμερίζεται με (όχι) την ορθολογική προσέγγιση.

3.4.2 Etzioni

Την αναζήτηση μιας ενδιάμεσης προσέγγισης επιδίωξε και ο Etzioni (1968), η μικτή ή συνθετική ανίχνευση του οποίου φιλοδοξεί επίσης να γεφυρώσει τα πλεονεκτήματα του ορθολογισμού και του αυξητισμού, αποφεύγοντας ταυτόχρονα τα μειονεκτήματα αμφοτέρων. Κατά την άποψη του Etzioni (1968), η αυξητική μεθοδολογία κρίνεται ανεπαρκής, κυρίως στα ακόλουθα σημεία. Πρώτον, αγνοείται ή υποβαθμίζεται το γεγονός της ανισότητας των σχέσεων δύναμης και επιρροής (power differentials) μεταξύ των μετεχόντων, δεύτερον, η περιγραφική εγκυρότητα του αυξητισμού συνδέεται συνήθως (ή εξαρτάται από) με την ομαλότητα των συνθηκών και κυρίως την απουσία κρίσης ή σύγκρουσης αξιών, και τρίτον, η αντισυνοπτική μεθοδολογία δεν παρέχει ικανοποιητικό μηχανισμό αντιμετώπισης θεμάτων ή προβλημάτων μείζονος, θεμελιώδους σημασίας και εμβέλειας, με συνέπεια να μη διακρίνονται όσον αφορά τον τρόπο αντιμετώπισής τους από θέματα ή αποφάσεις ελάσσονος σημασίας και σπουδαιότητας.

Η μικτή θεώρηση συνίσταται στη διάκριση δύο τύπων ή μεθοδολογιών αποφάσεων, ανάλογα με το είδος του προβλήματος και την έκταση και σημασία των συνεπειών των επιλογών. Θεμελιώδεις ή μεγάλης κλίμακας αποφάσεις και διαδικασίες είναι αυτές που θέτουν τις βασικές κατευθύνσεις πολιτικής και συμπεριφοράς και διαμορφώνουν το πλαίσιο εντός του οποίου λαμβάνονται οι μικρότερης κλίμακας και σημασίας αποφάσεις. Οι τελευταίες είτε προετοιμάζουν τη λήψη θεμελιωδών αποφάσεων είτε τις υλοποιούν, και αυτό μπορεί να συμβαίνει σύμφωνα με την αυξητική, αντισυνοπτική μεθοδολογία. Όσον αφορά τις υψηλού επιπέδου θεμελιώδεις αποφάσεις, αυτές διαμορφώνονται, κατά τη μικτή θεώρηση, αφού ληφθούν υπόψη και αναλυθεί ένας αριθμός κρίσιμων εναλλακτικών επιλογών στρατηγικής σημασίας, δίχως εξαντλητική μελέτη όλων των πιθανών εναλλακτικών τρόπων δράσης και των συνεπειών τους (και σε αυτό έγκειται η διαφορά της μικτής θεώρησης από το αμιγώς ορθολογικό πρότυπο).

Παρά τα αναμφισβήτητα πλεονεκτήματά της, η μικτή θεώρηση δεν παρέχει ασφαλή κριτήρια διάκρισης και ταξινόμησης των αποφάσεων σε μείζονος και ήσσονος σημασίας. Δεν είναι δύσκολο ασφαλώς να φανταστεί κανείς περιστάσεις που οι άνθρωποι διαφωνούν σχετικά με τη σημασία ή τη σπουδαιότητα συγκεκριμένου θέματος ή προβλήματος δημόσιας πολιτικής. Σε αυτές τις περιπτώσεις, η μικτή θεώρηση δεν παρέχει κανενός είδους κριτήρια διάκρισης, πέρα από την παρατήρηση ότι τα μείζονα θέματα αντιμετωπίζονται και οι θεμελιώδεις αποφάσεις λαμβάνονται με διαδικασίες που προσομοιάζουν, ως ένα σημείο, χωρίς να ταυτίζονται με τη συνοπτική μεθοδολογία. Για το λόγο αυτό η μικτή «ανίχνευση» έχει χαρακτηριστεί από ορισμένους επικριτές της το ίδιο ουτοπική, όπως και το συνοπτικό, και το ίδιο ληθαργική και συντηρητική, όπως το αυξητικό πρότυπο.

3.4.3 Lindblom

Ως ενδιάμεση προσέγγιση μπορεί να θεωρηθεί και η μεταγενέστερη ανάλυση του Lindblom (1979), ειδικότερα η μεθοδολογία της «στρατηγικής ανάλυσης». Ο Lindblom (1979) σε μεταγενέστερες αναλύσεις του φαίνεται να τροποποιεί τον αρχικό ενθουσιασμό του για την αυξητική μεθοδολογία, να υιοθετεί μια ενδιάμεση θέση και να εκφράζεται θετικά για δημιουργικές αναλύσεις που υπερβαίνουν τόσο τις συνοπτικές όσο και τις αντισυνοπτικές απόπειρες. Συγκεκριμένα, η στρατηγική ανάλυση προτείνεται ως μια εξελιγμένη μορφή αυξητικής μεθοδολογίας (και όχι πολιτικής) που προσεγγίζει αρκετά την έννοια της πεπερασμένης λογικότητας του Simon (1960). Η «στρατηγική ανάλυση» επιδιώκει την απλούστευση, και συνεπώς τον αποτελεσματικότερο έλεγχο, πολύπλοκων προβλημάτων δημόσιας πολιτικής, με την επικέντρωση της προσοχής σε μια δέσμη κρίσιμων χαρακτηριστικών και τη χρησιμοποίηση κατάλληλα επιλεγμένων «στρατηγημάτων» (stratagems). Με τον τρόπο αυτό «υπονομεύεται» η πολυπλοκότητα και συνθετότητα του προβλήματος, αποσοβείται η εξαντλητική και λεπτομερής εξέταση όλων των εναλλακτικών επιλογών και των συνεπειών τους – όπως προβλέπεται από τη συνοπτική θεώρηση. Δεδομένου ότι τα πολύπλοκα προβλήματα σπάνια αναλύονται και ελέγχονται πλήρως, η επινόηση κατάλληλων μεθόδων ανάλυσης και αντιμετώπισης που χαρακτηρίζονται από «έντεχνη ελλειπτικότητα» (skillful incompleteness) παρουσιάζεται ως προφανής εναλλακτική. Αν απεικονιστεί η πληρότητα της ανάλυσης ως ένα συνεχές, στο ένα άκρο του οποίου οριοθετείται η απόλυτα πλήρης ή συνοπτική ανάλυση και στο άλλο η απόλυτα ατελής ή ασυστηματοποίητη (αντισυνοπτικός πόλος), η στρατηγική ανάλυση καταλαμβάνει ενδιάμεσο χώρο μεταξύ των δύο πολικών σημείων και πάντως σε αρκετή απόσταση από τον πόλο της πλήρους ατελούς ανάλυσης.



Σχήμα 3.4.3.1

Η στρατηγική ανάλυση διαφοροποιείται από την πολιτική κατάσταση του πλήρους συνοπτισμού στο μέτρο που συνιστά όχι ένα λιγότερο ουτοπικό ιδανικό, αλλά μια πιο χρήσιμη κατευθυντήρια επιδίωξη που μπορεί να προσεγγίσει και να υλοποιηθεί με επιτυχία. Η στρατηγική ανάλυση δεν είναι τόσο απομακρυσμένη από την εμπειρία και δυνατότητα βελτίωσης του αποφασίζοντος, ώστε να ακυρώνει την όποια προσπάθειά του, ενώ ταυτόχρονα του παρέχει τις κατευθυντήριες αρχές ή

καλύτερα τη στρατηγική υπέρβασης του ελέγχου της πολυπλοκότητας και δυσκολίας αντιμετώπισης του προβλήματος. Η στρατηγική ανάλυση συνεπώς χρησιμεύει ως πρακτική μεθοδολογία χειρισμού και αντιμετώπισης πολύπλοκων προβλημάτων δημόσιας πολιτικής με τη σχετική απλούστευση, τυποποίηση και περιχαράκωσή τους, ώστε να καθίσταται δυνατή η ακριβέστερη δυνατή καταγραφή των δεδομένων τους και τελικά η διανοητική άλωσή τους. Κατά τον Lindblom (1979), ο ασύνδετος αυξητισμός δεν είναι παρά μια επιμέρους τακτική στο πλαίσιο της στρατηγικής ανάλυσης. Ως παρόμοιες επιμέρους τακτικές θεωρεί ο Lindblom (1979) την υιοθέτηση ικανοποιητικών και όχι αριστοποιητικών κριτηρίων από τον αποφασίζοντα (Simon), το κανονιστικό βελτιστοποιητικό πρότυπο (Dror) και τη μικτή – ανίχνευση (Etzioni). Οι ενδιάμεσες αυτές αναλυτικές στρατηγικές αντιπροσωπεύουν, κατά τον Lindblom, υψηλής ποιότητας επιστημονικά προγράμματα παρέμβασης (και βελτίωσης) στην πρακτική της δημόσιας πολιτικής.

3.5 Συμπερασματική προσέγγιση της Θεωρίας των Αποφάσεων

Συμπερασματικά, μπορεί να παρατηρηθεί ότι η σύντομη παρουσίαση των διαφόρων αναλυτικών προτύπων, που απαντούν στο χώρο της δημόσιας πολιτικής, αρκεί για να προσδώσει την εντύπωση της δυσχέρειας της επιλογής του ενός ή του άλλου. Η τοποθέτηση του θέματος διαζευκτικά ή πολωτικά (πλήρης συνοπτισμός ή αμιγής αυξητισμός) δεν φαίνεται να είναι η εγκυρότερη ούτε η πλέον χρήσιμη και αποτελεσματική πρακτική. Αντίθετα, φαίνεται ορθότερη η άποψη ότι, προκειμένου για διαφορετικής υφής, χαρακτήρα και σημασίας θέματα και αποφάσεις, ίσως προσιδιάζουν διαφορετικά αναλυτικά πρότυπα και μεθοδολογίες. Το ζητούμενο κατά συνέπεια δεν είναι τόσο η πρόκριση ενός άριστου προτύπου ούτε η επιλογή μιας μέσης θεώρησης, αλλά η υιοθέτηση μια ενδεχομενικής στρατηγικής, σύμφωνα με την οποία η αναλυτική μεθοδολογία οφείλει να προσαρμόζεται στο αντικείμενο, τις συνθήκες και τους όρους της ανάλυσης. Το κρίσιμο ερώτημα, επομένως, είναι «τι είδους πρόβλημα» είναι αυτό που αντιμετωπίζεται και από την απάντησή του θα εξαρτηθεί ή θα προσδιοριστεί η καταλληλότητα της αναλυτικής μεθόδου, των κριτηρίων και των τεχνικών που χρησιμοποιούνται. Στον προσδιορισμό του είδους και του χαρακτήρα του προβλήματος οι οργανωτικές συνθήκες και ορισμένοι ψυχολογικοί και γνωστικοί όροι διαδραματίζουν έναν ιδιαίτερα κρίσιμο ρόλο.

Καταλήγοντας, θα πρέπει να τονίσουμε ότι το κλασικό πλέον μοντέλο του Simon (1960), που έχει επικρατήσει και που αναφέρεται στη διαδικασία λήψης μιας απόφασης από έναν αποφασίζοντα, χωρίζει τελικά τη διαδικασία λήψης μιας απόφασης στις ακόλουθες τρεις φάσεις:

- i. **Νοητική φάση (intelligent phase):** Αρχικά αναζητούνται καταστάσεις για τις οποίες μπορεί να ληφθούν αποφάσεις.
- ii. **Σχεδιασμός (design):** Κατά τη φάση αυτή, γίνεται έρευνα, ανάλυση και ανάπτυξη όλων των δυνατών εναλλακτικών τρόπων δράσης (αποφάσεων).

- iii. **Επιλογή (choice):** Στη τελική φάση, γίνεται η επιλογή της καταλληλότερης απόφασης μέσα από το σύνολο των εναλλακτικών τρόπων δράσης (αποφάσεων).

Όταν κατά τη διάρκεια λήψης μιας απόφασης και οι τρεις ανωτέρω φάσεις είναι δομημένες, τότε πρόκειται για πλήρως δομημένο πρόβλημα. Αν καμιά από τις τρεις φάσεις δεν είναι δομημένη, τότε το πρόβλημα θεωρείται πλήρως αδόμητο. Τέλος, αν κάποιες από τις φάσεις είναι δομημένες και κάποιες αδόμητες, τότε το πρόβλημα θεωρείται ημιδομημένο. Στη τρίτη φάση, περιέχεται και η εργασία της ολοκλήρωσης της λύσης (implementation), η οποία λόγω της σημαντικότητάς της αντιμετωπίζεται από πολλούς συγγραφείς ως μια τέταρτη ανεξάρτητη φάση

Τέλος, ο Anthony (1965) ορίζει τις ακόλουθες τρεις γενικές κατηγορίες αποφάσεων που εμπεριέχουν όλες τις δυνατές διοικητικές δραστηριότητες:

- **Λειτουργικός έλεγχος** (operational control), που αναφέρεται στην αποδοτικότητα και αποτελεσματικότητα εκτέλεσης ειδικών εργασιών.
- **Διοικητικός έλεγχος** (managerial control), που αφορά την απόκτηση και αποδοτική χρησιμοποίηση των απαραίτητων μέσων για την επίτευξη των στόχων.
- **Στρατηγικός σχεδιασμός** (strategic planning), που αναφέρεται στους μακροπρόθεσμους στόχους και τις ακολουθούμενες πολιτικές κατανομής πόρων.

Στον πίνακα 3.5.1 παρουσιάζεται (Turban, 1993) το αποτέλεσμα του συνδυασμού των εργασιών των Simon (1960) και Anthony (1965) από τους Gorry και Scott-Morton (1971).

Αποφάσεις	Λειτουργικός Έλεγχος	Διοικητικός Έλεγχος	Στρατηγικός Σχεδιασμός	Υποστηρίζεται από
Δομημένες	Έλεγχος αποθεμάτων, Χρεώστες	Βραχυπρόθεσμες προβλέψεις, Ανάλυση προϋπολογισμού, Αναφορές προσωπικού	Τόπος εγκατάστασης εργοστασίων και αποθηκών	MIS Μοντέλα OR
Ημιδομημένες	Πρόγραμμα παραγωγής, Έλεγχος απογραφής	Προετοιμασία προϋπολογισμού	Σχεδίαση νέων προϊόντων, Νέα σχέδια	DSS
Αδόμητες	Διαχείριση χρηματικών πόρων, Αγορά software.	Διαπραγματεύσεις, Αγορά hardware.	Προγραμματισμός ερευνών και ανάπτυξης, Ανάπτυξη νέων τεχνολογιών.	DSS, Expert Systems

Πίνακας 3.5.1. Πλαίσιο υποστήριξης αποφάσεων (Πηγή: Turban, 1993)

3.6 Πολυκριτήρια Ανάλυση Αποφάσεων

Η υπόθεση όμως ενός και μοναδικού κριτηρίου καθιστά δύσχρηστη την εφαρμογή των παραπάνω μοντέλων στα προβλήματα της πραγματικής ζωής. Η πραγματικότητα ποτέ δεν είναι μονοδιάστατη, αλλά αντίθετα πολυδιάστατη. Μια ολοκληρωμένη προσέγγιση της πραγματικότητας απαιτεί τη συγκέντρωση του συνόλου της πληροφορίας, που αντανακλά τις κατευθύνσεις της. Αυτή η βασική παρατήρηση αποκτά ιδιαίτερη βαρύτητα στο χώρο των αποφάσεων, καθώς αυτές αναφέρονται σε προβλήματα της πραγματικότητας. Κατ' αυτό τον τρόπο, η δόμηση ενός προβλήματος απόφασης απαιτεί την παρουσία όλων των κριτηρίων που το χαρακτηρίζουν. Τα κριτήρια αυτά, συνήθως είναι ανεξάρτητα και ανταγωνίζονται μεταξύ τους για τις προτιμήσεις του συμμετέχοντος στην απόφαση.

Συμπεραίνεται λοιπόν, όπως αναφέρει και ο Ματσατσίνης (2004), ότι οι λαμβανόμενες αποφάσεις ανήκουν στη κατηγορία των ημιδομημένων αποφάσεων, γεγονός που δημιουργεί την ανάγκη για υποστήριξη του αποφασίζοντος μέσω της ανάπτυξης κατάλληλων πολυκριτήριων μοντέλων. Παρόλο που οι βασικές έννοιες της πολυκριτήριας ανάλυσης ήταν ήδη γνωστές από το 18ο αιώνα (Bernoulli and Cramer), εν τούτοις η θεωρητική αντιμετώπιση και η αξιωματική θεμελίωσή της έγινε γνωστή από τις εργασίες των von Neumann και Morgenstern (1944) και Savage (1954). Το 1944 οι Von Neuman και Morgenstern δημοσίευσαν ένα βιβλίο σχετικά με τη θεωρία των παιγνίων και την οικονομική συμπεριφορά, το οποίο οδήγησε την θεωρία των παιγνίων στη διαδικασία λήψης αποφάσεων. Στα τέλη της δεκαετίας του 1950 στη Γαλλία, ο Maurice Allais ανέπτυξε τη δική του θεωρία, η οποία ονομάστηκε Multiattribute Utility Theory (M.A.U.T.).

Με αφετηρία το πρώτο συνέδριο με αντικείμενο την Πολυκριτήρια Ανάλυση Αποφάσεων (Multiple Criteria Decision Making), που πραγματοποιήθηκε το 1972 στο Πανεπιστήμιο της Νότιας Καρολίνας, αναπτύχθηκε με γρήγορους ρυθμούς αυτός ο νέος κλάδος της επιχειρησιακής έρευνας (Cohrane and Zeleny, 1973; Fishburn, 1970; 1972; Keeney and Raiffa, 1976; Evans and Steuer, 1973; Zeleny, 1974; Roy, 1968; Jacquet-Lagrange and Siskos, 1982). Στο Social Sciences Research Institute, το οποίο εδρεύει στο University of Southern California, ο Ward Edwards, πιστεύοντας πολύ στην απλοποίηση των προβλημάτων, λανσάρισε τη δεκαετία του '80 τη Simple Multiattribute Rating Technique (S.M.A.R.T.). Στις δεκαετίες του '60 και '70, οι Keeney και Raiffa, και άλλοι συνέλαβαν την ιδέα ότι οι άνθρωποι εκφράζουν τους εαυτούς τους μέσω συστημάτων αξιών, δηλαδή η πραγματική αξία από κάποιο πράγμα δεν εξαρτάται ακριβώς από αριθμούς, αλλά περισσότερο από το τι σημαίνει η πραγματική αξία για κάποιον. Το να κερδίσει κάποιος, ένα εκατομμύριο ευρώ, δεν έχει την ίδια αξία για όλους. Την αξία αυτή την ονόμασαν Subjective Equivalent Utility (S.E.U.).

Το 1973 ο Lofty Zadeh παρουσίασε στον κόσμο τη θεωρία του Fuzzy logic και γεννήθηκε η ασαφής διαδικασία λήψης απόφασης. Στην πραγματικότητα, η θεωρία του Fuzzy logic προήλθε από τη Theory of Vagueness, η οποία παρουσιάστηκε από το Bertrand Russell στις αρχές του αιώνα. Η διαφορετική

προσέγγιση της διαδικασίας απόφασης δημιούργησε δύο σχολές, την αμερικανική και την ευρωπαϊκή: Η πολυκριτήρια λήψη αποφάσεων (MultiCriteria Decision Making – MCDM), δέχεται την ύπαρξη ενός συστήματος αξιών που αντιπροσωπεύει τις προτιμήσεις των αποφασιζόντων σε ένα σύνολο εναλλακτικών ενεργειών και έχει ως κύριους εκφραστές, τους Fishburn (1970), Keeney και Raifa (1976), Saaty (1990), Keeney (1992), French (1993), von Winterfeldt and Edwards (1993). Το σύστημα αξιών αφορά τη διαμόρφωση μιας συνάρτησης χρησιμότητας και των σχετικών βαρών των προτεραιοτήτων. Σύμφωνα με αυτή, οι προτιμήσεις του αποφασίζοντα σε ένα σύνολο εναλλακτικών ενεργειών λαμβάνονται υπόψη στη διαμόρφωση ενός συστήματος αξιών, το οποίο ικανοποιεί ένα σύνολο συνθηκών, μέσα από το οποίο ο αποφασίζων θα οδηγηθεί στην επιλογή της σωστότερης λύσης.

Οι βασικοί στόχοι της πολυκριτήριας λήψης αποφάσεων είναι:

- Να καθορισθούν οι συνθήκες που πρέπει να ικανοποιούνται ώστε να υφίσταται το σύστημα αξιών.
- Να υποστηρίξει τον αποφασίζοντα ώστε να ανακαλύπτει μέσα από μια διαδικασία ένα σύστημα αξιών και να παίρνει τη σωστή απόφαση.

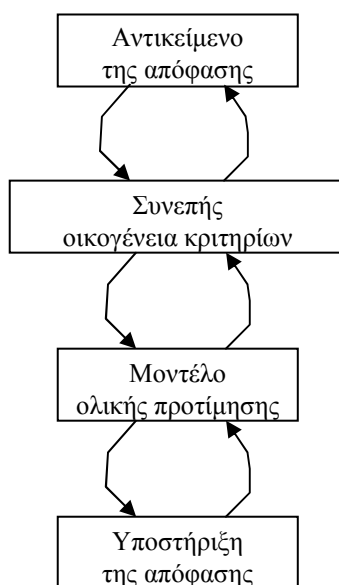
Η πολυκριτήρια υποβοήθηση της λήψης αποφάσεων (MultiCriteria Decision Aid –MCDA), έχει ως κύριους εκφραστές, τους Roy (1976; 1985; 1990), Bouyssou (1984), Jacquet-Lagrèze and Siskos (1982), Vincke (1992). Σύμφωνα με τον Roy (1990):

- Στη διαδικασία λήψης μιας απόφασης εμπλέκονται κατά κανόνα περισσότεροι του ενός αποφασίζοντα με διαφορετικές απόψεις και πιστεύω.
- Στη περίπτωση που έχουμε ένα μόνο αποφασίζοντα, οι προτιμήσεις του έχουν κατά κανόνα ένα μεγάλο βαθμό ασάφειας με ασυνέπειες και συγκρούσεις.
- Οι εναλλακτικές επιλογές, τα κριτήρια, η αξιολόγηση των εναλλακτικών επιλογών αλλά και οι προτιμήσεις των αποφασιζόντων καθορίζονται με αυθαίρετες ως επί το πλείστον διαδικασίες.
- Δεν είναι εύκολο να ειπωθεί ότι μια απόφαση που λαμβάνεται με την εφαρμογή ενός μαθηματικού μοντέλου, είναι καλή, ενώ μια άλλη είναι κακή κυρίως διότι δεν είναι δυνατή η αξιολόγηση των αποτελεσμάτων της εφαρμογής όλων των εναλλακτικών επιλογών, επειδή:
 - ο Η επιλογή και υλοποίηση μιας απόφασης αποκλείει αυτόματα την επιλογή και υλοποίηση των υπολοίπων.
 - ο Οι συνέπειες από την υλοποίηση μιας απόφασης γίνονται γνωστές κατά κανόνα μετά από μεγάλο χρονικό διάστημα.
 - ο Κατά το χρονικό διάστημα υλοποίησης μιας απόφασης μεταβάλλονται οι συνθήκες που ίσχυαν τη χρονική στιγμή που ελήφθη η απόφαση.

Ο ρόλος της πολυκριτήριας ανάλυσης υποστήριξης αποφάσεων δεν είναι να υποδείξει στον αποφασίζοντα αποφάσεις καλύτερες από αυτές που αντιλαμβάνεται, αλλά να τον οδηγήσει στην επιλογή της, μέσα από τη σταδιακή κατανόηση και βελτίωση των ικανοτήτων και γνώσεών του.

3.6.1 Μεθοδολογία μοντελοποίησης στην Πολυκριτήρια Ανάλυση

Το γενικό πλαίσιο μοντελοποίησης προβλημάτων απόφασης στην πολυκριτήρια ανάλυση οριοθετείται από τέσσερα διαδοχικά αλλά και αλληλεπιδρώντα στάδια, όπως φαίνεται συνοπτικά στο Σχήμα 3.6.1.1.



Σχήμα 3.6.1.1 Διαδοχικά στάδια μοντελοποίησης

Ας εξετάσουμε όμως αναλυτικότερα κάθε στάδιο χωριστά:

Στάδιο Ι : Αντικείμενο της απόφασης.

Στο στάδιο αυτό είναι απαραίτητο να ολοκληρωθούν οι εξής δύο βασικές εργασίες:

- Αυστηρός ορισμός του συνόλου Α των δραστηριοτήτων
- Καθορισμός μιας προβληματικής

Το αντικείμενο της απόφασης οφείλει να αναλυθεί σε ένα πεπερασμένο ή συνεχές σύνολο δραστηριοτήτων (σύνολο Α). Ο ορισμός μιας προβληματικής πάνω στο σύνολο Α αποσκοπεί στο να δώσει επιχειρησιακό ρόλο στο έργο της υποστήριξης της απόφασης. Σχετίζεται άμεσα με το ερώτημα: «πώς θα διαχειριστούμε τις δραστηριότητες;» ή πιο γενικά «τι θέλουμε να κάνουμε;».

Μπορούμε να διακρίνουμε τέσσερις προβληματικές αναφορές:

Προβληματική α: επιλογή μιας και μόνης δραστηριότητας από το σύνολο Α.

Προβληματική β: ταξινόμηση των δραστηριοτήτων σε ομογενείς προκαθορισμένες κατηγορίες

Προβληματική γ: κατάταξη των δραστηριοτήτων του συνόλου Α από την καλύτερη μέχρι τη χειρότερη, και

Προβληματική δ: περιγραφή των δραστηριοτήτων και των συνεπειών τους στη γλώσσα των αποφασιζόντων.

Παράδειγμα προβληματικής α αποτελεί η εκλογή ενός διευθυντή ή υπαλλήλου μέσα από ένα σύνολο πολλών υποψηφίων που αποτελούν το σύνολο Α. Παράδειγμα προβληματικής β αποτελεί το πρόβλημα παροχής πίστης από μια τράπεζα, όπου οι αιτήσεις δανειοδότησης (σύνολο Α) πρέπει να ταξινομηθούν σε αυτές που θα εγκριθούν, σε αυτές που θα απορριφθούν και σε εκείνες για τις οποίες απαιτούνται πρόσθετα στοιχεία. Η κατάταξη των υποψηφίων μαθητών για τα Α.Ε.Ι. (σύνολο Α) μέσω των πανελληνίων εξετάσεων είναι ένα κλασικό παράδειγμα προβληματικής γ. Τέλος, σε προβλήματα όπου πρώτητη ανάγκη είναι η γνωριμία με ένα σύνολο δραστηριοτήτων (αγορά μικροϋπολογιστών, στρατηγικές επένδυσης, αναπτυξιακά προγράμματα κ.λπ.), η προβληματική δ συνίσταται στην περιγραφή / αξιολόγηση των δραστηριοτήτων μέσω των επιπτώσεων τους πάνω στα κριτήρια που ενδιαφέρουν τον αποφασίζοντα. Μια και μόνη προβληματική δεν παραμένει αναγκαστικά σταθερή καθ' όλη τη διάρκεια της διαδικασίας της απόφασης, αλλά μπορεί να μεταβάλλεται με την πολυπλοκότητα του προβλήματος.

Στάδιο II: Συνεπής οικογένεια κριτηρίων.

Κάθε δυνατή δραστηριότητα από το σύνολο Α αντανakλά ένα «νέφος στοιχειωδών επιπτώσεων» (χαρακτηριστικά, ιδιότητες, πλεονεκτήματα, μειονεκτήματα κ.λπ.) μέσω των οποίων είναι δυνατή η εκτίμησή της από τον αποφασίζοντα. Για έναν επιβάτη του μετρό, για παράδειγμα, αυτό το «νέφος» θα περιλαμβάνει διαστάσεις: οικονομικές (τιμή εισιτηρίου), χρονικές (απαραίτητος χρόνος για την άφιξη στο σταθμό, αναμονή στην αποβάθρα, διάρκεια ταξιδιού κ.λπ.), άνεσης (αριθμός προβλεπόμενων ή αναγκαστικών στάσεων, αισθητικό περιβάλλον, κλιματισμός, θόρυβος κ.λπ.), ασφάλειας, πληροφόρησης, κ.λπ.

Ο ρόλος του μοντελοποιού συνίσταται στη διασάφηση των επιπτώσεων των διαφόρων δραστηριοτήτων του συνόλου Α και, εν συνεχεία, στην επινόηση και μοντελοποίηση των κριτηρίων βάσει των οποίων θα παρθεί η απόφαση.

Ορίζουμε ως κριτήριο, κάθε μονότονη μεταβλητή δηλωτική των προτιμήσεων ενός αποφασίζοντα. Ένα κριτήριο μπορεί να είναι είτε ποσοτικό και να εκφράζεται από μια συνεχή κλίμακα (χρόνου, κόστους, κ.λπ.) είτε ποιοτικό, για την μοντελοποίηση του οποίου οριοθετείται μια συμβατική κλίμακα διακεκριμένων τιμών, του τύπου: Άριστος (βαθμός 5), Πολύ Καλός (βαθμός 4), κ.λπ.

Στη μαθηματική γλώσσα, ένα κριτήριο αντιπροσωπεύεται από μια πραγματική συνάρτηση:

$$g: A \rightarrow R / a \rightarrow g(a)$$

όπου $g(a)$ είναι η εκτίμηση της δραστηριότητας $a \in A$ πάνω στο κριτήριο g . Η συνάρτηση αυτή οφείλει να πληρεί την ιδιότητα μονοτονίας. Εάν δηλαδή a και b είναι στοιχεία του Α, ισχύει:

$$g(a) > g(b) \Leftrightarrow a \text{ προτιμάται από την } b$$

$$g(a) = g(b) \Leftrightarrow a \text{ ισοδύναμη της } b$$

Τα κριτήρια αποτελούν μοντέλα σύγκρισης των δραστηριοτήτων του προβλήματος και οφείλουν να πληρούν τρεις θεμελιώδεις συνθήκες: μονοτονία, επάρκεια, μη

πλεονασμός. Ένα τέτοιο σύστημα κριτηρίων ονομάζεται **συνεπής οικογένεια κριτηρίων**.

Ας θεωρήσουμε το σύνολο των κριτηρίων $(g_1, g_2, \dots, g_n)$. Οι τρεις συνθήκες ορίζονται τότε ως εξής:

1. Μονοτονία (monotonicity)

Εάν για ένα ζεύγος δραστηριοτήτων (a, b) ισχύει: $g_i(a) = g_i(b)$ για κάθε $i \neq j$ και $g_j(a) > g_j(b)$, τότε η δραστηριότητα a προτιμάται από τη b .

2. Επάρκεια (exhaustivity)

Εάν για ένα ζεύγος δραστηριοτήτων (a, b) ισχύει: $g_i(a) = g_i(b)$ για κάθε $i = 1, 2, \dots, n$, τότε η δραστηριότητα a είναι ισοδύναμη της b , δηλαδή δεν απουσιάζει κανένα κριτήριο απόφασης από το σύνολο των n κριτηρίων.

3. Πλεονασμός (non redundancy)

Η διαγραφή ενός κριτηρίου g , από το σύνολο των κριτηρίων είναι ικανή να αναιρέσει μια από τις προηγούμενες δύο συνθήκες για κάποιο ζεύγος δραστηριοτήτων.

Τα κριτήρια κατασκευάζονται με τη βοήθεια ποσοτικών ή ποιοτικών διαστάσεων – κλιμάκων που επιτρέπουν την εκτίμηση όλων των δραστηριοτήτων του συνόλου A . Χρησιμοποιούνται για το σκοπό αυτό διαστάσεις οικονομικές, περιβαλλοντικές, κοινωνικές, τεχνολογικές, πολιτικές, αναπτυξιακές κ.λπ.

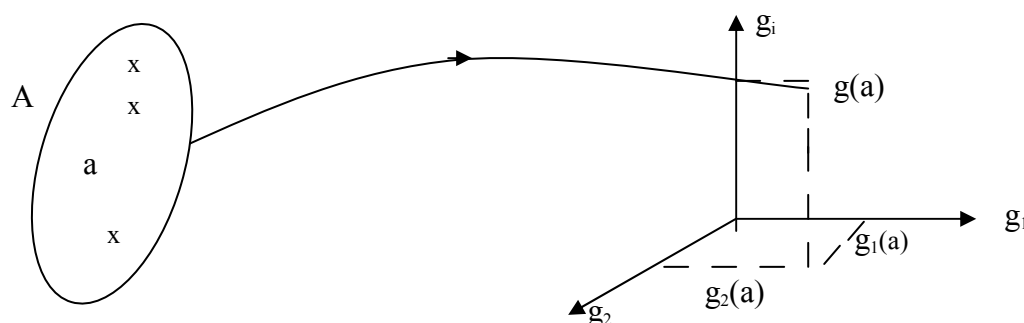
Συμβαίνει συχνά, ορισμένες στοιχειώδεις επιπτώσεις δραστηριοτήτων να εκφράζονται από κάποια διάσταση, η οποία, όμως, δεν εκφράζει και τις προτιμήσεις του αποφασίζοντα κατά τρόπο μονότονο. Για τη θερμοκρασία ενός χώρου, για παράδειγμα, δεν μπορούμε να πούμε, ούτε «όσο χαμηλότερη θερμοκρασία τόσο καλύτερη» ούτε το αντίστροφο. Σε τέτοιες περιπτώσεις, επιβάλλεται η μετατροπή της κλίμακας σε μια άλλη, πολλές φορές ποιοτική, η οποία υπακούει στη συνθήκη της μονοτονίας. Έτσι, για το συγκεκριμένο παράδειγμα της θερμοκρασίας θα μπορούσαμε να δεχτούμε τη μοντελοποίηση που υποδεικνύει ο επόμενος πίνακας (πίνακας 3.6.1.1).

Θερμοκρασία C	Ποιότητα	Κλίμακα
22-33	Άριστη	5
19-21 και 24-27	Πολύ καλή	4
15-18 και 28-30	Καλή	3
11-14 και 31-35	Μέτρια	2
0-10 και 36-40	Κακή	1
κάτω του 0 και πάνω από 40	Ανυπόφορη	0

Πίνακας 3.6.1.1 Κλίμακα ποιοτικών κριτηρίων

Μια συνεπής οικογένεια κριτηρίων απεικονίζει το σύνολο των δραστηριοτήτων A μέσα στον n -διάστατο πραγματικό χώρο R^n , όπως φαίνεται και στο Σχήμα 3.6.1.2.

Τέλος, με $g(a) = (g_1(a), g_2(a), \dots, g_n(a))$ συμβολίζουμε το διάνυσμα των εκτιμήσεων της δραστηριότητας $a \in A$ πάνω στα n κριτήρια, το οποίο ονομάζουμε πολυκριτήρια εκτίμηση.



Σχήμα 3.6.1.2. Πολυκριτήρια απεικόνιση του συνόλου A

Στάδιο III: Μοντέλο ολικής προτίμησης.

Πρόκειται για τον κανόνα σύνθεσης κριτηρίων (μερικές προτιμήσεις) μέσα από ένα μοντέλο ολικής προτίμησης. Οι δραστηριότητες του συνόλου A συγκρίνονται συνολικά με βάση το μοντέλο και τον τύπο προβληματικής που έχει οριστεί στο στάδιο I.

Στάδιο IV: Υποστήριξη της απόφασης.

Στο στάδιο αυτό, ο μοντελοποιός αναζητεί και οργανώνει τα στοιχεία απάντησης σε συγκεκριμένα ερωτήματα που θέτει το ίδιο το πρόβλημα καθώς επίσης και ο λήπτης της απόφασης. Πρόκειται για συμπληρωματικό στάδιο του προηγούμενου, με λόγο ύπαρξης το γεγονός ότι μια λύση που δίνει ένα μοντέλο δεν είναι άμεσα εκμεταλλεύσιμη στα πεδία λήψης αποφάσεων και / ή διαπραγματεύσεων.

3.6.2 Η Ανάλυση των Δεδομένων

Στην πρώτη εμφάνιση της, η Στατιστική χρησιμοποιήθηκε μόνο για την καταμέτρηση και την παρουσίαση στοιχείων. Με την επιστημονική της όμως καθιέρωση, η Στατιστική προχώρησε πέρα από τα στάδια της παρουσίασης και της συλλογής και καθιερώθηκε στους τομείς της ανάλυσης και εξαγωγής συμπερασμάτων ή προβλέψεων. Η πολυδιάστατη θεώρηση των χαρακτηριστικών ενός φαινομένου, οδήγησε με τη σειρά της στη δημιουργία της οικογένειας των μεθόδων της Πολυδιάστατης Στατιστικής Ανάλυσης. Οι μέθοδοι αυτές επιτρέπουν την ανάλυση μεγάλου όγκου δεδομένων και επικεντρώνονται στην ταυτόχρονη παρουσίαση και αναζήτηση σχέσεων μεταξύ πολλών (ομογενών ή ετερογενών) μεταβλητών.

Η Ανάλυση Δεδομένων υπάγεται στις μεθόδους της Πολυδιάστατης Στατιστικής Ανάλυσης και προσφέρει σημαντικά εργαλεία για την ανάδειξη των κυρίαρχων τάσεων που χαρακτηρίζουν ένα αρχικό σύνολο δεδομένων καθώς και των νόμων στους οποίους τα δεδομένα αυτά υπακούουν. Οι μέθοδοι της Ανάλυσης

Δεδομένων προέκυψαν από το έλλειμμα της πληροφορίας που δημιουργεί μια απλή στατιστική ανάλυση ενός μεγάλου συνόλου δεδομένων. Οι απαιτήσεις για μια επαρκή ανάλυση και επεξεργασία πλήθους ερωτηματολογίων, ερευνών και συνεντεύξεων επέβαλε την ανάγκη μιας εποπτικής παρουσίασης των δεδομένων σε συνθήκες μέγιστης πληροφορίας. Σε γενικές γραμμές, η Ανάλυση Δεδομένων χαρακτηρίζεται ως σύνολο μεθόδων, των οποίων η βασική αρχή συνίσταται στην ανάδειξη των επικρατέστερων σχέσεων μεταξύ των αντικειμένων (παρατηρήσεων), μεταξύ των μεταβλητών που τα χαρακτηρίζουν καθώς και μεταξύ των αντικειμένων και των μεταβλητών. Με άλλα λόγια, οι μέθοδοι αυτές μας επιτρέπουν να μελετήσουμε τις σχέσεις μεταξύ των αντικειμένων και να μετρήσουμε την σημαντικότητα καθεμιάς μεταβλητής στις σχέσεις αυτές.

Η Ανάλυση Δεδομένων βασίστηκε στις εργασίες των στατιστικών της δεκαετίας του 1930. Η μεγάλη ανάπτυξη, όμως, στον τομέα αυτό, εκδηλώθηκε στην Γαλλία τη δεκαετία του 1960. Η ανάπτυξη αυτή απέφερε εντυπωσιακά αποτελέσματα και οδήγησε στην καθιέρωση της γαλλικής σχολής της Ανάλυσης Δεδομένων. Η σχολή αυτή διέπεται από την αρχή να μείνουμε όσο γίνεται πλησιέστερα στα αρχικά δεδομένα (χωρίς τροποποιήσεις και *a priori* υποθέσεις), μετασχηματίζοντας τα το λιγότερο δυνατό. Χαρακτηριστική είναι η φράση του Benzecri (1973), ενός εκ των πρώτων και σημαντικότερων ερευνητών της Ανάλυσης Δεδομένων : « Το μοντέλο οφείλει να προσαρμόζεται στα δεδομένα και όχι το αντίστροφο». Κατ' αυτό τον τρόπο, οι στατιστικοί επιστήμονες της σχολής αυτής, αρνούμενοι να κάνουν αμφίβολες υποθέσεις, παρουσιάζουν τα αρχικά δεδομένα και τις τάσεις που εξάγονται από αυτά. Η απουσία υποθέσεων διευκολύνει έναν λιγότερο εξειδικευμένο στατιστικά επιστήμονα να οδηγηθεί σε ερμηνείες των γεγονότων και των συνεπειών τους.

Δύο είναι οι βασικές προσεγγίσεις στο χώρο της γαλλικής σχολής της Ανάλυσης Δεδομένων, η Παραγοντική Ανάλυση και η Ταξινόμηση:

- Η παραγοντική ανάλυση έχει σαν σκοπό την αναγνώριση των αξόνων, γύρω από τους οποίους οργανώνονται τα δεδομένα, και την παρουσίαση μιας «γεωμετρικής εικόνας τους». Κυριότερος εκπρόσωπος της προσέγγισης αυτής είναι η μέθοδος της Παραγοντικής Ανάλυσης των Αντιστοιχιών, με την οποία αναλύονται πίνακες συχνοτήτων ποιοτικών, κυρίως, μεταβλητών.
- Η Ταξινόμηση, διαχωρίζει σε κλάσεις ένα σύνολο δεδομένων ανάλογα με τις ιδιότητες που το χαρακτηρίζουν και δομεί τις σχέσεις μεταξύ των παρατηρήσεων.

Η Στατιστική επιστήμη θεωρείται το λίκνο της θεωρίας των Αποφάσεων. Οι στατιστικές μετρήσεις των δεδομένων παρείχαν, εξ αρχής, ελκυστικές εκτιμήσεις των προτιμήσεων ενός συνόλου ατόμων. Η ραγδαία εξέλιξη της Στατιστικής επέτρεψε επίσης, την εξαγωγή συμπερασμάτων, με τη μορφή των προβλέψεων, που αφορούσαν αποφάσεις.

Σε σχέση με το σύνολο των ρευμάτων που χαρακτηρίζουν τα δύο αυτά επιστημονικά πεδία, τη μεγαλύτερη συνάφεια παρουσιάζουν μεταξύ τους, τα δύο

ρεύματα της Ανάλυσης Δεδομένων και της Πολυκριτήριας Ανάλυσης. Αν και με μια πρώτη ματιά οι δύο αυτοί χώροι φαίνεται να παρουσιάζουν σημαντικές διαφορές, παρόλα αυτά εμφανίζουν λίγα, αλλά σημαντικά, κοινά χαρακτηριστικά. Τα χαρακτηριστικά αυτά εντοπίζονται στον:

1. Κοινωνικό χαρακτήρα των μεθόδων
2. Πολυδιάστατη αντιμετώπιση ενός φαινομένου
3. Απουσία αυθεντίας.

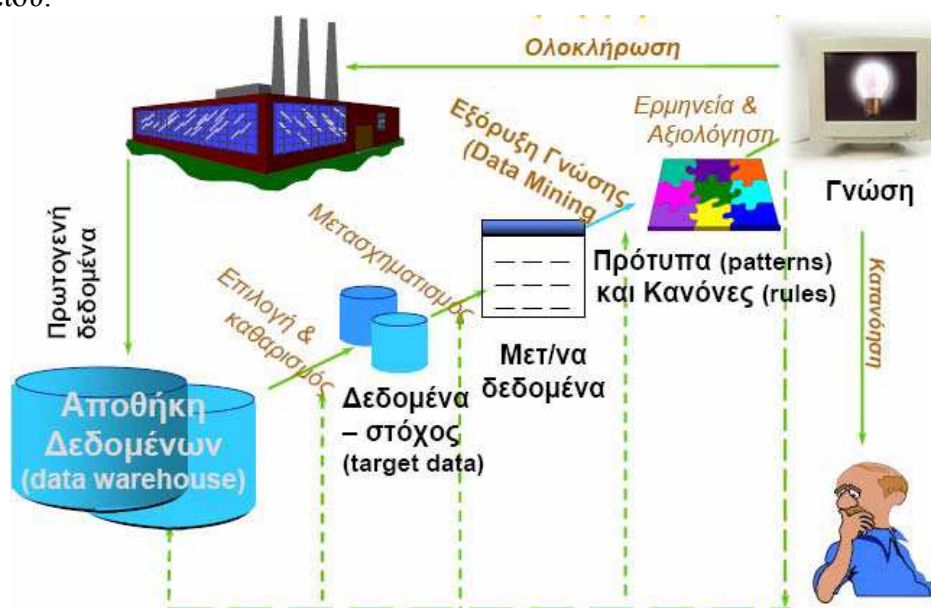
Τα δύο πρώτα χαρακτηριστικά, αναφέρθηκαν επανειλημμένα ως βασικά χαρακτηριστικά των δύο χώρων. Το τρίτο χαρακτηριστικό εδραιώνεται στη δημοκρατική θεώρηση των μεθόδων της Πολυκριτήριας Ανάλυσης καθώς και στην απουσία αρχικών υποθέσεων που παρατηρείται στην Ανάλυση Δεδομένων.

Από το 1972, έτος πραγματοποίησης του πρώτου επίσημου συνεδρίου στο πανεπιστήμιο της Νότιας Καρολίνας των Η.Π.Α. με θέμα την Πολυκριτήρια Ανάλυση Αποφάσεων, αναπτύχθηκε με ταχύ ρυθμό μία νέα θεωρία της επιχειρησιακής έρευνας, με βασικό αντικείμενο τη λήψη των αποφάσεων υπό το καθεστώς πολλαπλών κριτηρίων απόφασης. Η νέα αυτή αντίληψη των προβλημάτων απόφασης βοήθησε στο να αναθεωρηθούν/ γενικευθούν πολλές από τις κλασικές μεθόδους βελτιστοποίησης της επιχειρησιακής έρευνας. Παράλληλα, ο ρεαλισμός που διέπει τα μοντέλα της Πολυκριτήριας Ανάλυσης οδήγησε σε πληθώρα εφαρμογών σε πραγματικά προβλήματα οργάνωσης και διοίκησης. Τα κυριότερα θεωρητικά ρεύματα της Πολυκριτήριας Ανάλυσης είναι τέσσερα:

- Ο πολυκριτήριος μαθηματικός προγραμματισμός
- Η πολυκριτήρια θεωρία χρησιμότητας
- Η θεωρία των σχέσεων υπεροχής
- Η θεωρία μονότονης παλινδρόμησης

3.6.3 Από την Πολυκριτήρια Ανάλυση στην Εξόρυξη Δεδομένων (Data Mining)

Η μοντελοποίηση του ερωτηματολογίου (βλέπε Παράρτημα) έγινε με βάση τη μεθοδολογία της Πολυκριτήριας Ανάλυσης, όπου τα κριτήρια κατασκευάζονται με τη βοήθεια ποσοτικών ή ποιοτικών διαστάσεων – κλιμάκων που επιτρέπουν την εκτίμηση όλων των δραστηριοτήτων του συνόλου. Ο όγκος των δεδομένων που προκύπτει, η επεξεργασία του συνόλου των απαντήσεων / προτιμήσεων καθώς και η εξαγωγή συμπερασμάτων για τα ερωτήματα που τέθηκαν στην αρχή αυτής της διπλωματικής εργασίας απαιτούν τη χρησιμοποίηση ενός σύγχρονου μεθοδολογικού εργαλείου.



Σχήμα 3.6.3.1 Η πορεία προς τη γνώση

Αφού λοιπόν έχουμε δημιουργήσει αυτόν τον, αρκετά μεγάλο, όγκο δεδομένων, πώς θα εξερευνήσουμε όλες αυτές τις πληροφορίες; Πώς θα ξέρουμε ποια είναι η πραγματική σημασία αυτών των δεδομένων; Τι συμπεράσματα προκύπτουν από όλα αυτά τα δεδομένα; Είναι πραγματικά χρήσιμα όλα αυτά τα δεδομένα; Με ποιόν τρόπο θα μπορέσουμε να ανακαλύψουμε τη γνώση που περιέχεται μέσα σε αυτό τον όγκο των δεδομένων; (Σχήμα 3.6.3.1.) Σε όλα αυτά τα ερωτήματα, μπορούμε να απαντήσουμε χρησιμοποιώντας την Εξόρυξη Δεδομένων (Data Mining) κι αν θέλαμε να δώσουμε έναν σύντομο ορισμό της έννοιας Data Mining, θα καταλήγαμε στον ακόλουθο:

«Είναι η αυτοματοποιημένη εξαγωγή γνώσης και χρήσιμων πληροφοριών που δεν μας ήταν γνωστές, μέσα από σύνολα δεδομένων, τα οποία από μόνα τους και στη μορφή στην οποία βρίσκονται, δεν θα ήταν δυνατόν να μας τις προσφέρουν. Περιλαμβάνει ένα σύνολο μεθόδων και τεχνικών που έχουν σκοπό την ανακάλυψη των πληροφοριών αυτών και την μετατροπή τους σε χρήσιμα και εκμεταλλεύσιμα δεδομένα ή συμπεράσματα».

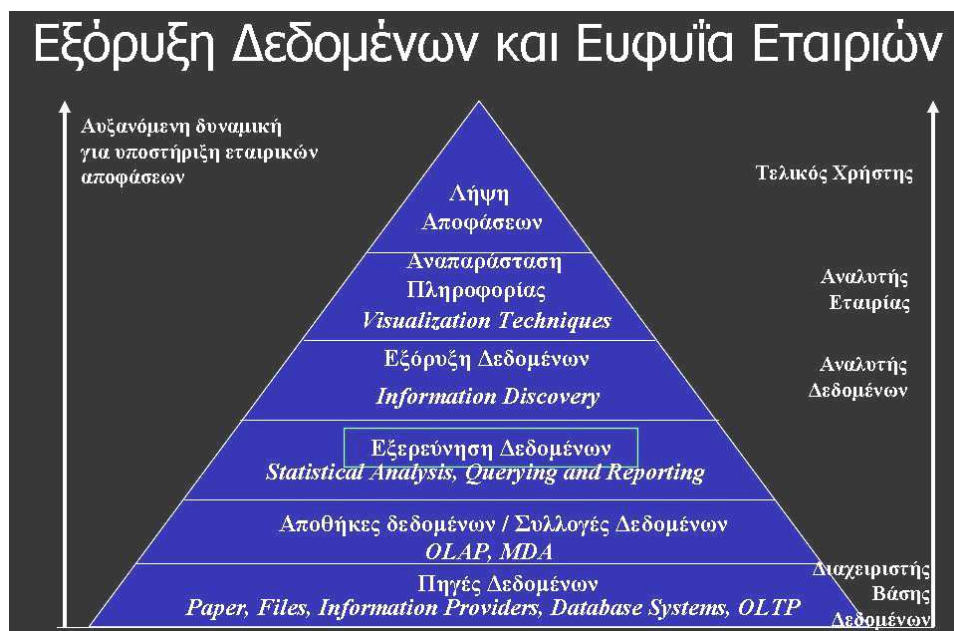
Η Εξόρυξη Γνώσης και Δεδομένων:

(Προτεινόμενη μεθοδολογία ποιοτικής ανάλυσης της συμπεριφοράς των ψηφοφόρων)

3.7 Data Mining (Εξόρυξη Δεδομένων)

Οι Βάσεις Δεδομένων σήμερα μπορούν να κυμαίνονται σε μέγεθος ίσο με Terabytes - δηλαδή πάνω από 1.000.000.000.000 bytes δεδομένων. Μέσα σε αυτό το πλήθος των στοιχείων βρίσκονται κρυμμένες πληροφορίες στρατηγικής σημασίας. Αλλά όταν υπάρχουν τόσα πολλά και διαφορετικά “δέντρα”, πώς μπορείς να αντλήσεις σημαντικά συμπεράσματα σε σχέση με το “δάσος”;

Η νεότερη απάντηση είναι η Εξόρυξη Δεδομένων (Σχήμα 3.1.1). Τα επιδιωκόμενα οφέλη είναι τεράστια. Καινοτόμοι οργανισμοί σε όλο τον κόσμο χρησιμοποιούν ήδη τη μέθοδο Εξόρυξη Δεδομένων για να εντοπίσουν και να απευθυνθούν σε υψηλότερης αξίας πελάτες, να αναμορφώσουν το προϊόν τους, έτσι ώστε να αυξήσουν τις πωλήσεις και να ελαχιστοποιήσουν τις απώλειες λόγω απάτης ή λάθους. Η Εξόρυξη Δεδομένων είναι μια διαδικασία, η οποία χρησιμοποιεί μια σειρά από εργαλεία ανάλυσης δεδομένων για να «αποκαλύψει» συνήθειες και σχέσεις σε δεδομένα, που μπορούν να χρησιμοποιηθούν για να γίνουν αξιόπιστες προβλέψεις.



Σχήμα 3.7.1 Data Mining (Εξόρυξη Δεδομένων) και Λήψη Απόφασης
(πηγή: Καραναστάση Α, 2008)

Το πρώτο και απλούστερο αναλυτικό βήμα προς την Εξόρυξη Δεδομένων είναι η περιγραφή των δεδομένων - η γραφική αποτύπωση των στατιστικών χαρακτηριστικών (καθώς και των τυπικών αποκλίσεων), καθώς και η ανεύρεση

δυναμικά σημαντικών σχέσεων μεταξύ μεταβλητών (όπως π.χ. τιμές - απαντήσεις που συχνά συνυπάρχουν). Όπως υπογραμμίζεται στην ενότητα σχετικά με τη διαδικασία Εξόρυξης Δεδομένων, η συλλογή, η αναζήτηση και η επιλογή των σωστών στοιχείων είναι εξαιρετικά σημαντική. Αλλά η περιγραφή των δεδομένων δεν μπορεί από μόνη της να παρουσιάσει ένα σχέδιο δράσης. Πρέπει να οικοδομήσουμε ένα μοντέλο πρόβλεψης που να βασίζεται σε πρότυπα καθορισμένα από γνωστά αποτελέσματα και στη συνέχεια, να ελέγξουμε αυτό το μοντέλο σε αποτελέσματα πέραν του αρχικού δείγματος. Ένα καλό πρότυπο μπορεί να μην επεξηγεί ακριβώς την πραγματικότητα (γνωρίζετε έναν οδικό χάρτη που να είναι πιστή αναπαράσταση των πραγματικών δρόμων;), αλλά μπορεί να αποτελέσει έναν χρήσιμο οδηγό για την κατανόηση των δεδομένων που επεξεργαζόμαστε. Το τελικό βήμα είναι να επαληθευτεί εμπειρικά το μοντέλο. Για παράδειγμα, από μια βάση δεδομένων των πελατών που έχουν ήδη ανταποκριθεί σε μια συγκεκριμένη προσφορά της επιχείρησης, έχουμε χτίσει ένα μοντέλο πρόβλεψης του οποίου οι προοπτικές είναι πιθανότερο να επαληθευτούν σε μια ίδια προσφορά. Μπορούμε να βασιστούμε σε αυτήν την πρόβλεψη; Αρκεί να ρωτήσουμε – εξετάσουμε μια λίστα νέων πελατών και να δούμε τι αποτελέσματα θα έχουμε.

Συνοψίζοντας, τα βασικά βήματα της Εξόρυξης Δεδομένων για την ανακάλυψη της γνώσης είναι:

1. Ορισμός επιχειρηματικού προβλήματος
2. Κατασκευή βάσης εξόρυξης δεδομένων
3. Εξερεύνηση δεδομένων
4. Προετοιμασία δεδομένων για την μοντελοποίηση
5. Κατασκευή του μοντέλου / προτύπου
6. Αξιολόγηση του μοντέλου
7. Ανάπτυξη του μοντέλου και των αποτελεσμάτων

Και τα καθήκοντα στην οικοδόμηση μιας βάσης δεδομένων της εξόρυξης δεδομένων είναι:

- α. Η συλλογή δεδομένων
- β. Η περιγραφή των δεδομένων
- γ. Επιλογή
- δ. Ποιότητα των δεδομένων και αξιολόγησή τους
- ε. Η ενοποίηση και ολοκλήρωση
- στ. Κατασκευή μετα-δεδομένων
- ζ. Διατήρηση / ενημέρωση της βάσης εξόρυξης δεδομένων

Όπως επίσης θα πρέπει να θυμόμαστε ότι οι παρακάτω διαδικασίες δεν εκτελούνται με αυστηρή σειρά, αλλά με βάση τις ανάγκες:

Κατάταξη / Ταξινόμηση: Χάρτες δεδομένων σε προκαθορισμένες κατηγορίες

Παλινδρόμηση: Χάρτες δεδομένων σε μια συνάρτηση

Πρόβλεψη: Πρόβλεψη μελλοντικών δεδομένων (τμήμα αυτών)

Ανάλυση χρονοσειρών: Αναλύει τα δεδομένα στο χρόνο

Ομαδοποίηση: Εύρεση ομάδων παρόμοιων χαρακτηριστικών
Κανόνες σύνδεσης: Εύρεση σχέσεων μεταξύ των αντικειμένων
Χαρακτηριστικά: Αντληση αντιπροσωπευτικών πληροφοριών
Ακολουθία Ανακάλυψης (Discovery): Εύρεση διαδοχικών μοντέλων

3.8 Εξόρυξη δεδομένων: Τι δεν μπορεί να κάνει!

Ένα από τα πιο συνηθισμένα ερωτήματα που γεννιούνται σχετικά με το Data Mining, είναι ποιος είναι ο λόγος να το χρησιμοποιήσουμε, αφού ήδη υπάρχουν πολλές μέθοδοι, για να καταλήξουμε σε συμπεράσματα, χρησιμοποιώντας τα δεδομένα που διαθέτουμε. Κυριότερος «αντίπαλος» του Data Mining είναι οι στατιστικές μέθοδοι. Όμως, υπάρχει αρκετά μεγάλη διαφορά μεταξύ τους παρά τη φαινομενική ομοιότητα των μεθόδων και των συμπερασμάτων που προκύπτουν και από τους δύο τρόπους επεξεργασίας των δεδομένων (Σχήμα 3.8.1).

Κατ' αρχήν, θα πρέπει να τονιστεί η διαφορά στη στόχευση της κάθε μεθόδου. Το Data Mining έχει δημιουργηθεί για χρήση των αποτελεσμάτων του από χρήστες που δεν είναι ειδικοί στο Data Mining ή για την εισαγωγή τους σε συστήματα που θα τα χρησιμοποιήσουν για ευρύτερες και πολύπλοκες αναλύσεις επιχειρησιακών δεδομένων και όχι από κάποιον πολύ εξειδικευμένο επιστήμονα, όπως μαθηματικό, στατιστικολόγο κ.λπ. Επίσης, το Data Mining οδηγεί σε συμπεράσματα, ακόμα και σε δημιουργία κανόνων, σε αντίθεση με τη στατιστική, όπου θα πρέπει τις περισσότερες φορές ο χρήστης των στατιστικών πακέτων να ερμηνεύσει και να εξηγήσει τα αποτελέσματα των στατιστικών μεθόδων και να αξιολογήσει τη χρησιμότητά τους.



Σχήμα 3.8.1 Η χρησιμότητα και σημασία της Εξόρυξης Δεδομένων
(πηγή: Καραναστάση Α, 2008)

Βέβαια στην αναλυτική περιγραφή του Data Mining φαίνεται η χρήση στατιστικών μεθόδων κατά τη διαδικασία Data Mining, αλλά χρησιμοποιούνται ως

ένα από τα εργαλεία και όχι ως οι μέθοδοι που θα οδηγήσουν στα τελικά συμπεράσματα. Έτσι, λοιπόν, θα ήταν σωστότερο να πούμε ότι η στατιστική προσφέρει τις χρήσιμες μεθόδους της, οι οποίες χρησιμοποιούνται σε συνδυασμό με τις υπόλοιπες μεθόδους στη διαδικασία του Data Mining.

Η Εξόρυξη Δεδομένων είναι λοιπόν ένα εργαλείο, δεν είναι κάποιο μαγικό ραβδί. Δεν θα εγκατασταθεί στη βάση δεδομένων σας, παρακολουθώντας τι συμβαίνει για να σας στέλνει ενημερωτικά e-mail, όταν ανακαλύπτει ένα ενδιαφέρον μοτίβο. Αυτό, βέβαια, δεν καταργεί την ανάγκη να γνωρίζουμε για παράδειγμα καλύτερα την επιχείρησή μας, να κατανοούμε τα δεδομένα μας, ή να κατανοούμε τις μεθόδους ανάλυσης.

Θα πρέπει να θυμόμαστε ότι οι σχέσεις πρόβλεψης (predictive relationships) που βρέθηκαν μέσω της Εξόρυξης Δεδομένων δεν είναι απαραίτητως αίτια μιας ενέργειας ή συμπεριφοράς. Για παράδειγμα, η Εξόρυξη Δεδομένων μπορεί να καθορίζει ότι οι άντρες των οποίων τα εισοδήματα είναι μεταξύ 50.000 € και 65.000 € και που είναι συνδρομητές σε συγκεκριμένα περιοδικά, είναι πιθανοί αγοραστές του προϊόντος που θέλετε να πουλήσετε. Αν και μπορείτε να εκμεταλλευτείτε αυτό το πρότυπο, στοχεύοντας στην προώθηση των προϊόντων σας σε ανθρώπους που ταιριάζουν σε αυτό (το πρότυπο), δεν θα πρέπει να υποθέσουμε ότι κάποια από τα παραπάνω στοιχεία (εισόδημα, φύλο, αγορά περιοδικών) προκάλεσε την αγορά του προϊόντος σας. Για τη διασφάλιση ουσιαστικών αποτελεσμάτων, είναι ζωτικής σημασίας να καταλαβαίνουμε τα δεδομένα μας.

3.9 Εξόρυξη δεδομένων και αποθήκευση δεδομένων

Συχνά, τα δεδομένα που εξορύσσουμε είναι τα πρώτα που εξάγουμε από μια επιχειρησιακή αποθήκη (βάση) δεδομένων σε μια βάση εξόρυξης δεδομένων. Υπάρχει πραγματικό όφελος, εάν τα δεδομένα μας είναι ήδη μέρος μιας βάσης δεδομένων. Όπως θα δούμε παρακάτω, το πρόβλημα της εκκαθάρισης στοιχείων από μια αποθήκη δεδομένων και από μια βάση εξόρυξης δεδομένων είναι σχεδόν παρόμοια. Εάν τα στοιχεία έχουν ήδη εκκαθαριστεί από μια αποθήκη δεδομένων, τότε είναι πολύ πιθανόν να μην απαιτηθεί περαιτέρω εκκαθάριση, ώστε να τα εξορύξουμε. Επιπλέον, θα πρέπει να έχουν ήδη αντιμετωπιστεί πολλά από τα προβλήματα της ενοποίησης δεδομένων και να τεθούν σε εφαρμογή διαδικασίες συντήρησης.

Η βάση δεδομένων της εξόρυξης δεδομένων μπορεί να είναι ένα λογικό και όχι ένα φυσικό υποσύνολο των αποθηκευμένων δεδομένων σας, υπό την προϋπόθεση ότι η αποθήκη δεδομένων DBMS μπορεί να υποστηρίξει τις πρόσθετες απαιτήσεις των πόρων της εξόρυξης δεδομένων. Εάν δεν μπορεί, τότε θα είναι καλύτερα αν δημιουργήσουμε μια ξεχωριστή βάση εξόρυξης δεδομένων. Μια αποθήκη δεδομένων δεν αποτελεί απαραίτητη προϋπόθεση για την εξόρυξη δεδομένων. Η δημιουργία μιας μεγάλης αποθήκης δεδομένων που ενοποιεί δεδομένα από πολλαπλές πηγές, επιλύει προβλήματα αξιοπιστίας των δεδομένων.

3.10 Εξόρυξη δεδομένων, Μηχανική μάθησης και Στατιστική

Η Εξόρυξη Δεδομένων επωφελείται των πλεονεκτημάτων στους τομείς της τεχνητής νοημοσύνης (AI) και της στατιστικής. Και οι δύο κλάδοι ασχολούνται με τα προβλήματα της αναγνώρισης προτύπων και ταξινόμησης. Αμφότεροι έχουν συνεισφέρει στην κατανόηση και εφαρμογή των νευρωνικών δικτύων και των δέντρων αποφάσεων.

Η Εξόρυξη Δεδομένων δεν αντικαθιστά τις παραδοσιακές στατιστικές τεχνικές. Απεναντίας, είναι μια επέκταση των στατιστικών μεθόδων. Η ανάπτυξη των περισσότερων στατιστικών μεθόδων ήταν, μέχρι πρόσφατα, με βάση τη θεωρία και τις μεθόδους ανάλυσης που λειτούργησε αρκετά καλά για μικρό πλήθος ανάλυσης δεδομένων. Η αύξηση της ισχύος των ηλεκτρονικών υπολογιστών και το χαμηλότερο κόστος τους, σε συνδυασμό με την ανάγκη για ανάλυση τεράστιων αρχείων δεδομένων, επέτρεψαν την ανάπτυξη νέων τεχνικών, βασιζόμενοι στη μεγάλη ισχύ εξερεύνησης των πιθανών λύσεων.

Οι νέες τεχνικές περιλαμβάνουν σχετικά πρόσφατους αλγόριθμους, όπως τα νευρωνικά δίκτυα και τα δέντρα αποφάσεων, καθώς και νέες προσεγγίσεις παλαιότερων αλγορίθμων, όπως της διαφορικής ανάλυσης. Οι παραδοσιακές στατιστικές τεχνικές εξαρτώνται από τον σχεδιαστή (modeler) που καθορίζει τη λειτουργική μορφή και τις αλληλεπιδράσεις. Το σημείο κλειδί είναι ότι η Εξόρυξη Δεδομένων είναι η εφαρμογή αυτών αλλά και άλλων A.I (Τεχνητή Νοημοσύνη) και στατιστικών τεχνικών. Η Εξόρυξη Δεδομένων είναι ένα εργαλείο για την αύξηση της παραγωγικότητας των ανθρώπων που προσπαθούν να οικοδομήσουν την έξυπνη δημιουργία μοντέλων.

3.11 Εφαρμογές εξόρυξης δεδομένων

Η Εξόρυξη Δεδομένων γίνεται ολοένα και πιο δημοφιλής λόγω της σημαντικής συμβολής που μπορεί να έχει. Για παράδειγμα, μπορεί να χρησιμοποιηθεί για τον έλεγχο των δαπανών, καθώς και να συμβάλλει στην αύξηση των εσόδων. Πολλοί οργανισμοί χρησιμοποιούν την Εξόρυξη Δεδομένων για να διαχειριστούν όλες τις φάσεις του κύκλου ζωής του πελάτη, συμπεριλαμβανομένης της απόκτησης νέων πελατών, την αύξηση των εσόδων από τους υπάρχοντες πελάτες και τη διατήρηση καλών πελατών. Με τη διαπίστωση των χαρακτηριστικών των καλών πελατών (profiling), μια εταιρεία μπορεί να σχεδιάσει έρευνες, με τα ίδια χαρακτηριστικά. Με την κατανόηση και την κατηγοριοποίηση των πελατών που έχουν αγοράσει ένα συγκεκριμένο προϊόν θα μπορούσε να εστιάσει την προσοχή της σε παρόμοιους πελάτες που δεν έχουν αγοράσει το προϊόν αυτό (διασταυρούμενες πωλήσεις). Με την κατανόηση και την κατηγοριοποίηση των πελατών που έχει χάσει μια εταιρεία, μπορεί να ενεργήσει για να διατηρήσει τους πελάτες που βρίσκονται σε κίνδυνο να τους χάσει (μείωση απωλειών ή φθοράς), διότι είναι συνήθως πολύ λιγότερο δαπανηρό να διατηρήσεις έναν πελάτη από το να αποκτήσεις έναν νέο.

Η Εξόρυξη Δεδομένων προσφέρει αξία σε ένα ευρύ φάσμα βιομηχανιών. Τηλεπικοινωνίες και τράπεζες είναι δύο από τους ηγέτες στο πλαίσιο της εφαρμογής Εξόρυξη Δεδομένων για την ανίχνευση λανθάνουσας χρήσης των υπηρεσιών τους. Οι ασφαλιστικές εταιρείες και τα χρηματιστήρια επίσης ενδιαφέρονται για την εφαρμογή αυτής της τεχνολογίας για τη μείωση της απάτης. Οι Ιατρικές εφαρμογές είναι άλλος ένας γόνιμος τομέας: η Εξόρυξη Δεδομένων μπορεί να χρησιμοποιηθεί για να προβλεφθεί η αποτελεσματικότητα στις χειρουργικές επεμβάσεις, στα φάρμακα ή στις ιατρικές εξετάσεις. Επίσης, οι λιανοπωλητές χρησιμοποιούν αρκετά την Εξόρυξη Δεδομένων για να αποφασίσουν ποια προϊόντα θα αποθεματοποιήσουν (stock) σε συγκεκριμένα καταστήματα (ακόμη και πώς να τα τοποθετήσουν στην αποθήκη), καθώς και να αξιολογήσουν την αποτελεσματικότητα της προώθησης και των εκπτώσεων.

Η έρευνά μας έρχεται να προσθέσει έναν ακόμη τομέα, ένα ακόμη πεδίο εφαρμογής της Εξόρυξη Δεδομένων. Αυτό της ποιοτικής ανάλυσης της συμπεριφοράς των πολιτών και της λήψης πολιτικών αποφάσεων. Διότι η κατανόηση και κατηγοριοποίηση του εκάστοτε ψηφοφόρου καθώς και της ενδεχόμενης συμπεριφοράς του μπορεί να οδηγήσει στην εξαγωγή κρίσιμων, χρήσιμων και σημαντικών συμπερασμάτων προκειμένου να ληφθούν πολιτικές αποφάσεις.

3.12 Βασικές Εργασίες του Data Mining

Οι δύο βασικοί στόχοι του Data Mining πρακτικά είναι η πρόβλεψη (prediction) και η περιγραφή (description). Η πρόβλεψη περιλαμβάνει τη χρήση κάποιων μεταβλητών ή πεδίων στις βάσεις δεδομένων για να προβλέψουμε άγνωστες ή μελλοντικές τιμές άλλων μεταβλητών που έχουν ενδιαφέρον. Η περιγραφή επικεντρώνεται στην εύρεση προτύπων που περιγράφουν τα δεδομένα και τα οποία μπορούν να ερμηνευτούν από τον άνθρωπο. Η σχετική σημαντικότητα της πρόβλεψης και περιγραφής για συγκεκριμένες Data Mining εφαρμογές μπορούν να διαφέρουν σημαντικά. Ωστόσο, σε ότι αφορά την ανακάλυψη γνώσης, η περιγραφή τείνει να είναι περισσότερο σημαντική σε σχέση με την πρόβλεψη σε αντίθεση με τις εφαρμογές αναγνώρισης προτύπων και μηχανικής μάθησης που βασικός σκοπός είναι η πρόβλεψη.

Οι μέθοδοι Data Mining προκειμένου να επιτύχουν τους στόχους για την εξαγωγή και περιγραφή γνώσης από ένα σύνολο δεδομένων, χρησιμοποιούν ή εκτελούν κατά την εφαρμογή τους ένα σύνολο από εργασίες (tasks). Οι βασικότερες από αυτές τις εργασίες περιγράφονται στη συνέχεια.

3.12.1 Ομαδοποίηση (Description Data Mining)

Η ομαδοποίηση χωρίζει μια βάση δεδομένων σε διάφορες ομάδες. Ο στόχος της ομαδοποίησης είναι να βρούμε τις ομάδες που είναι διαφορετικές μεταξύ τους, τα μέλη των οποίων όμως έχουν κοινά χαρακτηριστικά. Σε αντίθεση με την ταξινόμηση (βλέπε Predictive Data Mining, παρακάτω), δεν ξέρουμε ποιες θα είναι οι ομαδοποιήσεις όταν ξεκινάμε την επεξεργασία, ή ποια από τα χαρακτηριστικά στοιχεία θα ομαδοποιηθούν. Συνεπώς, οι συσπειρώσεις θα πρέπει να ερμηνεύουν κατάλληλα. Συχνά είναι απαραίτητο να τροποποιήσουμε την ομαδοποίηση με τον αποκλεισμό εκείνων των μεταβλητών που κρίνονται ως άνευ σημασίας ή που δεν έχουν νόημα.

Αφού καταλήξουμε σε ομαδοποιήσεις που βρίσκονται σε αντιστοιχία με τη βάση δεδομένων μας, μπορούμε στη συνέχεια να τις χρησιμοποιήσουμε για την ταξινόμηση των νέων δεδομένων. Ορισμένοι από τους κοινούς αλγόριθμους που χρησιμοποιούνται για την εκτέλεση των συσπειρώσεων / ομαδοποιήσεων, περιλαμβάνουν Kohonen και K- means χάρτες χαρακτηριστικών.

Δεν θα πρέπει να συγχέουμε την ομαδοποίηση με την τμηματοποίηση. Η τμηματοποίηση αναφέρεται στο γενικό πρόβλημα του προσδιορισμού των ομάδων που έχουν κοινά χαρακτηριστικά. Η ομαδοποίηση είναι ένας τρόπος για τον διαχωρισμό / κατάτμηση των δεδομένων σε ομάδες που δεν είχαν οριστεί παλαιότερα, ενώ η ταξινόμηση είναι ένας τρόπος για την κατάτμηση των δεδομένων από ομάδες που έχουν ήδη οριστεί.

3.12.2 Ταξινόμηση (Predictive Data Mining)

Τα προβλήματα ταξινόμησης έχουν ως στόχο τον προσδιορισμό των χαρακτηριστικών που δείχνουν την ομάδα στην οποία κατά περίπτωση ανήκουν. Αυτό το πρότυπο μπορεί να χρησιμοποιηθεί τόσο για την κατανόηση των υφιστάμενων δεδομένων όσο και για την πρόβλεψη συμπεριφοράς νέων περιπτώσεων. Για παράδειγμα, μπορεί να θέλετε να προβλέψετε κατά πόσο τα άτομα γίνεται να ταξινομηθούν, σε αυτούς που μπορεί να ανταποκριθούν σε ένα e - mail άμεσης προσέγγισης πελατών, ή σε αυτούς που είναι ευάλωτοι στο να αλλάξουν πάροχο σταθερής / κινητής τηλεφωνίας ή σε αυτούς που, όπως εξετάζει και η έρευνά μας, ψηφίζουν με βάση το πρόγραμμα των κομμάτων.

Η Εξόρυξη Δεδομένων δημιουργεί μοντέλα κατάταξης μέσω της εξέτασης των δεδομένων που έχουν ήδη ταξινομηθεί (περιπτώσεις) και εφευρίσκει επαγωγικά ένα προγνωστικό πρότυπο (μοντέλο). Αυτές οι υπάρχουσες περιπτώσεις μπορεί να προέρχονται από μια παλαιότερη (ιστορική) βάση δεδομένων, όπως είναι τα άτομα που έχουν ήδη υποβληθεί σε μια συγκεκριμένη ιατρική αγωγή ή έχουν μετακινηθεί σε μια νέα εταιρία τηλεφωνικών υπηρεσιών. Είναι επίσης πιθανόν να προέρχονται από μια δοκιμή κατά την οποία ένα δείγμα του συνόλου της βάσης δεδομένων δοκιμάζεται (ελέγχεται) στον πραγματικό κόσμο, σε πραγματικές συνθήκες και τα

αποτελέσματα χρησιμοποιούνται για να δημιουργηθεί ένας ταξινομητής. Για παράδειγμα, από ένα δείγμα του καταλόγου του ηλεκτρονικού ταχυδρομείου που θα σταλεί μια προσφορά και τα αποτελέσματα της αλληλογραφίας που θα προκύψουν, μπορεί να αναπτυχθεί ένα ταξινομητικό πρότυπο που να εφαρμόζεται σε όλη τη βάση δεδομένων. Μερικές φορές, ο μελετητής / εμπειρογνώμονας κατατάσσει ένα δείγμα της βάσης δεδομένων, και αυτή η κατάταξη στη συνέχεια χρησιμοποιείται για να δημιουργηθεί το μοντέλο που θα εφαρμοστεί σε όλη τη βάση δεδομένων.

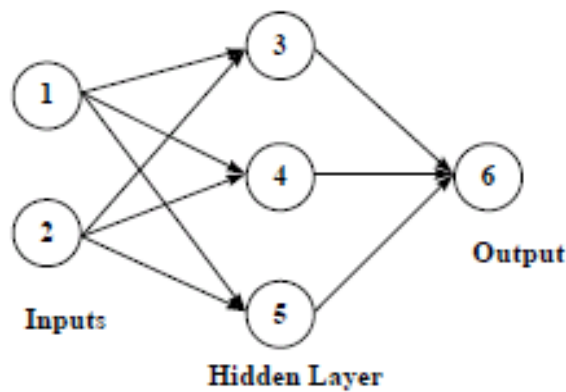
3.12.3 Παλινδρόμηση

Η παλινδρόμηση χρησιμοποιεί υπάρχουσες τιμές για να προβλέψει άλλες. Στην απλούστερη περίπτωση, η παλινδρόμηση χρησιμοποιεί τυποποιημένες στατιστικές τεχνικές, όπως η γραμμική παλινδρόμηση. Δυστυχώς, πολλά προβλήματα της πραγματικότητας (του πραγματικού κόσμου) δεν είναι απλώς γραμμικές προβλέψεις προηγούμενων τιμών. Για παράδειγμα, οι όγκοι των πωλήσεων, οι τιμές των αποθεμάτων, τα ποσοστά αποτυχίας παραγόμενων προϊόντων είναι πολύ δύσκολο να προβλεφθούν, διότι μπορεί να εξαρτώνται από πολύπλοκες αλληλεπιδράσεις πολλαπλών μεταβλητών πρόβλεψης. Ως εκ τούτου, πιο πολύπλοκες τεχνικές (π.χ., παλινδρόμηση εφοδιαστικής αλυσίδας (logistic regression), δέντρα αποφάσεων ή νευρωνικά δίκτυα) μπορεί να είναι αναγκαία για την πρόβλεψη μελλοντικών τιμών.

3.13 Μοντέλα και Αλγόριθμοι Εξόρυξης Δεδομένων

3.13.1 Νευρωνικά Δίκτυα

Τα Νευρωνικά δίκτυα παρουσιάζουν ιδιαίτερο ενδιαφέρον λόγω του ότι προσφέρουν ένα αποτελεσματικό μέσο για την μοντελοποίηση μεγάλων και πολύπλοκων προβλημάτων στα οποία μπορεί να υπάρχουν εκατοντάδες μεταβλητές πρόβλεψης με πολλές αλληλεπιδράσεις (πραγματικά βιολογικά νευρωνικά δίκτυα είναι ασύγκριτα πιο περίπλοκα). Τα νευρωνικά δίκτυα μπορούν να χρησιμοποιηθούν σε προβλήματα ταξινόμησης (όπου η «έξοδος» είναι μια ρητή μεταβλητή) ή για παλινδρομήσεις (όπου η «έξοδος» είναι μια συνεχής μεταβλητή). Ένα νευρωνικό δίκτυο (Σχήμα 3.13.1.1) ξεκινά με ένα επίπεδο εισόδου, όπου κάθε κόμβος αντιστοιχεί σε μεταβλητή πρόβλεψης. Αυτοί οι κόμβοι εισόδου είναι συνδεδεμένοι με μια σειρά από κόμβους σε ένα κρυμμένο επίπεδο. Κάθε κόμβος εισόδου είναι συνδεδεμένος με κάθε κόμβο στο κρυμμένο επίπεδο. Οι κόμβοι στο κρυμμένο επίπεδο μπορούν να συνδεθούν με τους κόμβους σε ένα άλλο κρυμμένο επίπεδο ή στο επίπεδο εξόδου. Το επίπεδο εξόδου αντιστοιχεί σε μία ή περισσότερες μεταβλητές απάντησης / απόφασης.

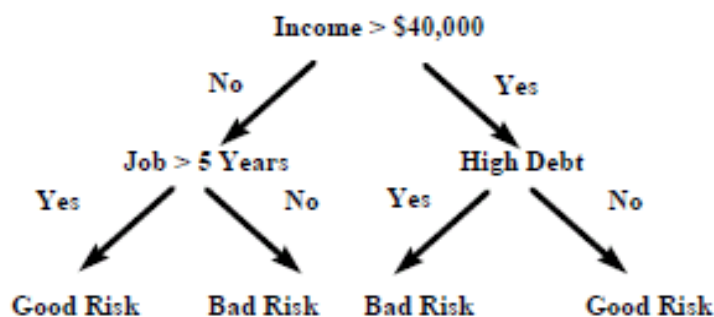


Σχήμα 3.13.1.1. Νευρωνικό δίκτυο με ένα κρυμμένο επίπεδο
 (πηγή: “Introduction to Data Mining and Knowledge Discovery”, 2005)

Η αρχιτεκτονική (ή τοπολογία) ενός νευρωνικού δικτύου είναι ο αριθμός των κόμβων και των κρυμμένων επιπέδων, και δείχνει πώς αυτά συνδέονται μεταξύ τους. Στο σχεδιασμό ενός νευρωνικού δικτύου, είτε ο χρήστης είτε το λογισμικό θα πρέπει να επιλέξουν τον αριθμό των κρυφών, των κρυμμένων επιπέδων, τη λειτουργική συνάρτηση (συνάρτηση ενεργοποίησης), και τα όρια σχετικά με τα βάρη. Τα Νευρωνικά δίκτυα διαφέρουν, λοιπόν, στη φιλοσοφία από πολλές στατιστικές μεθόδους, κι αυτό γιατί συνήθως έχουν περισσότερες παραμέτρους από ό, τι ένα τυπικό στατιστικό μοντέλο.

3.13.2 Δέντρα Αποφάσεων

Τα δέντρα αποφάσεων είναι ένας τρόπος αναπαράστασης μιας σειράς κανόνων που οδηγούν σε μια ταξινόμηση ή σε κάποια τιμή. Για παράδειγμα, μπορεί να θέλετε να ταξινομήσετε τις αιτήσεις δανείου σύμφωνα με τον πιστωτικό κίνδυνο (μεγάλο ή μικρό). Το Σχήμα 3.13.2.1 δείχνει ένα απλό δέντρο αποφάσεων που λύνει αυτό το πρόβλημα, απεικονίζοντας όλα τα βασικά στοιχεία ενός δέντρου απόφασης: κόμβοι απόφασης, «κλαδιά» και «φύλλα».



Σχήμα 3.13.2.1 Ένα απλό δέντρο ταξινόμησης
 (Πηγή: “Introduction to Data Mining and Knowledge Discovery”, 2005)

Τα μοντέλα των δέντρων απόφασης χρησιμοποιούνται συνήθως στην Εξόρυξη Δεδομένων για να εξετάσουμε τα δεδομένα και να δημιουργήσουμε το δέντρο και τους κανόνες του, που θα χρησιμοποιηθούν για να γίνουν προβλέψεις. Τα δένδρα αποφάσεων αναπτύσσονται μέσω ενός επαναληπτικού διαχωρισμού των δεδομένων σε διακριτές ομάδες, όπου ο στόχος είναι να μεγιστοποιήσουμε την «απόσταση» μεταξύ των ομάδων και των διαιρέσεών τους.

Τα δένδρα αποφάσεων που χρησιμοποιούνται για την πρόβλεψη κατηγοριοποιημένων (ασυνεχών;) μεταβλητών ονομάζονται δένδρα κατάταξης ή ταξινόμησης διότι κατηγοριοποιούν ή ταξινομούν τις αποστάσεις (μεταξύ των κριτηρίων / μεταβλητών). Τα δένδρα αποφάσεων που χρησιμοποιούνται για την πρόβλεψη συνεχών μεταβλητών ονομάζονται δένδρα παλινδρόμησης.

3.13.3 Διαφορική ανάλυση

Η Διαφορική Ανάλυση είναι η παλαιότερη μέθοδος (τεχνική) μαθηματικής ταξινόμησης, έχοντας πρώτα δημοσιευθεί από τον R. A. Fisher το 1936 με την κατηγοριοποίηση του περίφημου Iris botanical data σε τρεις κατηγορίες. Ανακαλύπτει υπέρ-επίπεδα (hyper planes) (π.χ. γραμμές σε δύο διαστάσεις, επίπεδα σε τρεις κ.λπ.) που διαχωρίζουν τις κατηγορίες. Το προκύπτον μοντέλο είναι πολύ εύκολο να ερμηνευτεί διότι ο χρήστης το μόνο που έχει να κάνει είναι να καθορίσει σε ποια μεριά της γραμμής (ή του υπερ-επίπεδου) καταγράφεται ένα σημείο (ένα δεδομένο). Η μέθοδος είναι αρκετά ευαίσθητη σε πρότυπα που ενυπάρχουν στα δεδομένα και χρησιμοποιείται πολύ συχνά σε σημαντικούς επιστημονικούς κλάδους, όπως η ιατρική, οι κοινωνικές επιστήμες και η βιολογία.

Παρόλα αυτά η Διαφορική Ανάλυση δεν είναι διαδεδομένη στην Εξόρυξη Δεδομένων. Κι αυτό για τρεις κυρίως λόγους. Πρώτον, συμπεραίνει ότι όλες οι μεταβλητές πρόβλεψης είναι κανονικές κατανομές (η καμπύλη στα ιστογράμματα έχει την μορφή «καμπάνας», όπως συνηθίζεται να αναφέρεται), κάτι που μπορεί να μην είναι αληθές. Δεύτερον, αταξινόμητες κατηγορικές μεταβλητές πρόβλεψης (π.χ. κόκκινο/μπλε/πράσινο) δεν μπορούν να χρησιμοποιηθούν καθόλου. Τρίτον, τα όρια που διαχωρίζουν τις κατηγορίες είναι γραμμικής μορφής, μόνο που μερικές φορές τα δεδομένα δεν γίνεται να διαχωριστούν με αυτόν τον τρόπο.

Πιο πρόσφατες εκδοχές της Διαφορική Ανάλυση απαντούν σε μερικά από αυτά τα προβλήματα, χρησιμοποιώντας τόσο τετραγωνικά όρια όσο και γραμμικά, κάτι που αυξάνει σημαντικά τον βαθμό ευαισθησίας σε συγκεκριμένες περιπτώσεις. Υπάρχουν επίσης τεχνικές/μέθοδοι που επιτρέπουν την αντικατάσταση της κανονικής κατανομής από μία εκτίμηση πραγματικής κατανομής (π.χ. αντικατάσταση της θεωρητικής καμπύλης σχήματος «καμπάνα», με ιστογράμματα από τις μεταβλητές πρόβλεψης). Ταξινομημένα κατηγορικά δεδομένα μπορούν να μοντελοποιηθούν σχηματίζοντας το ιστόγραμμα με τιμές που καθορίζονται από τις κατηγορικές μεταβλητές.

3.14 Αξιολόγηση εργαλείων Data Mining

Εφ' όσον μια επιχείρηση έχει πάρει την στρατηγική απόφαση να προχωρήσει στη χρησιμοποίηση των διαδικασιών του Data Mining, θα πρέπει στη συνέχεια να γίνει η επιλογή του κατάλληλου εργαλείου. Υπάρχουν τρία βασικά κριτήρια, τα οποία θα πρέπει κανείς να χρησιμοποιήσει οπωσδήποτε για να μπορέσει να αξιολογήσει ένα εργαλείο Data Mining.

Εγκυρότητα. Το εργαλείο για το Data Mining είναι απαραίτητο να έχει την ικανότητα να παράγει όσο το δυνατόν πιο έγκυρα μοντέλα. Η εγκυρότητα των μοντέλων θα πρέπει να μην επηρεάζεται από τις εξωτερικές συνθήκες ή τον «θόρυβο» των δεδομένων.

Ερμηνεία. Θα πρέπει το σύστημα, εκτός από την παραγωγή ορθών μοντέλων, να μπορεί να τα κάνει κατανοητά στον τελικό χρήστη. Θα πρέπει ο χρήστης να αισθάνεται ότι καταλαβαίνει πλήρως τη μοντελοποίηση που έχει γίνει, ώστε να μπορεί να κατανοήσει και τα συμπεράσματα που προκύπτουν από την περαιτέρω επεξεργασία.

Διασύνδεση. Όπως αναφέρθηκε και σε προηγούμενη παράγραφο, είναι απαραίτητο να υπάρχει διασύνδεση της διαδικασίας του Data Mining με την επιχειρησιακή λειτουργία της εταιρίας. Έτσι λοιπόν θα πρέπει και το σύστημα για το Data Mining να έχει τη δυνατότητα να συνδέεται όσο το δυνατόν με περισσότερα στάδια της λειτουργίας της επιχείρησης και να συλλέγει δεδομένα από διάφορα σημεία της ροής των πληροφοριών.

Η ικανοποίηση αυτών των κριτηρίων από ένα σύστημα Data Mining, είναι πολύ σημαντική για την παραγωγή αποτελεσματικών και ρεαλιστικών μοντέλων, τα οποία θα έχουν διαχρονικότητα και τη δυνατότητα να προσαρμόζονται στις νέες συνθήκες και στις μεταβολές των δεδομένων.

3.15 Το μέλλον του Data Mining

Η εμφάνιση του Data Mining στην αγορά προκάλεσε ανάμεικτα συναισθήματα. Οι πολέμοί του αμφισβήτησαν ότι μη συστηματικά ερωτήματα (π.χ. στηριζόμενα στη στατιστική) είναι δυνατό να οδηγήσουν σε χρήσιμα συμπεράσματα και χαρακτήρισαν αυτού του τύπου το λογισμικό "εφεύρεση γνώσης", όμως πολλοί άλλοι έσπευσαν να το υιοθετήσουν. Σε κάθε περίπτωση, η αξία του Data Mining δεν πρέπει να υπερεκτιμηθεί, γιατί κάτι τέτοιο συχνά οδηγεί στην απογοήτευση, όπως, για παράδειγμα, συνέβη πριν από κάποια χρόνια με τη τεχνητή νοημοσύνη. Το Data Mining είναι ένας νέος τομέας που κερδίζει ολοένα περισσότερο έδαφος, αλλά χρειάζεται ακόμα ανάπτυξη και εξέλιξη.

Τα βασικά προβλήματα που αφορούν στη διαχείριση μεγάλου όγκου δεδομένων παραμένουν, γι' αυτό θα πρέπει οι προσδοκίες από το Data Mining να περιορίζονται σε λογικά επίπεδα. Μερικά από τα προβλήματα που δεν έχουν ακόμα επιλυθεί πλήρως είναι και τα ακόλουθα:

- Ανάπτυξη αλγορίθμων για ταξινόμηση, ομαδοποίηση, ανάλυση συνάφειας, αναγνώριση τάσεων, ώστε να μπορούν να χρησιμοποιηθούν σε μεγάλες βάσεις δεδομένων.
- Ανάπτυξη σχημάτων κωδικοποίησης metadata (δεδομένων τα οποία χαρακτηρίζουν το περιεχόμενο και τη σημασία άλλων δεδομένων) μέσα σε πίνακες, ώστε να μπορεί να γίνει ανάκληση της αντίστοιχης πληροφορίας από το χρήστη.
- Εύρεση ασφαλιστικών δικλίδων, ώστε να αποφεύγονται περιπτώσεις όπου δεδομένα προσομοιώνονται με κάποιο μοντέλο κατά τύχη. Το πρόβλημα αυτό γίνεται εμφανέστερο καθώς το πρόγραμμα εξετάζει ένα μεγάλο πλήθος παραμέτρων για τον υπολογισμό της συνάφειας με διαφορετικά μοντέλα. Ως αποτέλεσμα, οι διαφορές είναι μικρές και κατά συνέπεια η πιθανότητα σύγχυσης μεγάλη.
- Ανάπτυξη αποτελεσματικών τρόπων μείωσης των παραμέτρων και του δείγματος που χρησιμοποιούνται. Παρ' ότι όσο μεγαλύτερο δείγμα χρησιμοποιούμε τόσο περισσότερες παραμέτρους μπορούμε να εξετάσουμε, κινδυνεύουμε να χάσουμε τον έλεγχο.
- Ανάπτυξη σχημάτων αναζήτησης σε μη ομογενή δεδομένα (π.χ. μείγμα δεδομένων video, φωνής και κειμένου).
- Ανάπτυξη νέων μεθόδων Data Mining, οι οποίες επιτρέπουν χρήση γνώσης για τα δεδομένα, ώστε να περιοριστεί το πεδίο της αναζήτησης, αντιμετωπίζοντας έτσι όσα προβλήματα δημιουργούνται από έλλειψη ή ασάφεια στα δεδομένα.
- Επέκταση των συστημάτων διαχείρισης βάσεων δεδομένων (όπως π.χ. η SQL), ώστε να υποστηρίζουν νέες δομές, οι οποίες επιτρέπουν άμεση εξαγωγή στατιστικών στοιχείων.
- Εφαρμογή των ανωτέρω μεθόδων και σε παράλληλες βάσεις δεδομένων με εκατοντάδες πίνακες, χιλιάδες πεδία και terabytes από δεδομένα. Τεχνικές βελτιστοποίησης είναι απαραίτητες.
- Δημιουργία κατανοητών μοντέλων, συμβιβάζοντας την πολυπλοκότητα με την ευκολία κατανόησης και χρησιμοποιώντας εργαλεία απεικόνισης, με παράλληλη χρήση διαλογικής αναζήτησης από τους αλγορίθμους του Data Mining, ώστε ο χρήστης να είναι σε θέση να καθοδηγεί την αναζήτηση.
- Ανάπτυξη του θεωρητικού υπόβαθρου αλλά και των αλγορίθμων, οι οποίοι απαιτούνται για τη μοντελοποίηση της αλλαγής στα δεδομένα. Οι μεγάλες βάσεις δεδομένων εξελίσσονται σε σχετικά μεγάλο χρονικό διάστημα, με άμεση συνέπεια να μην μπορούν να περιγραφούν με κάποια από τις συνηθισμένες στατιστικές κατανομές.

Τα ερωτήματα που καλείται να απαντήσει σήμερα το Data Mining είναι του τύπου "Τι συμβαίνει και γιατί;". Αύριο, όμως, θα πρέπει να ανταποκριθεί σε ακόμη μεγαλύτερες προκλήσεις, απαντώντας σε ερωτήματα του τύπου "Πώς μπορώ να διορθώσω το συγκεκριμένο πρόβλημα;".

Παρ' ότι δεν πρέπει να υπερεκτιμούμε τις σημερινές δυνατότητες του Data Mining, θέτοντας υπερβολικούς στόχους και έχοντας ουτοπικές προσδοκίες, είναι γεγονός ότι αυτός ο νέος τομέας εφαρμογής της τεχνητής νοημοσύνης άνοιξε νέους δρόμους στην επιχειρηματική πρακτική. Οι σωστές πληροφορίες αποτελούν το ισχυρότερο στρατηγικό όπλο κάθε επιχείρησης - και το Data Mining υπόσχεται να τις προσφέρει.

ΚΕΦΑΛΑΙΟ 4:

Υλοποίηση πρακτικής εφαρμογής (case study)

4.1 Γενικά

Στο παρόν κεφάλαιο θα παρουσιάσουμε την υλοποίηση της πρακτικής εφαρμογής (έρευνας κοινής γνώμης), που πραγματοποιήσαμε σε δείγμα 400 πολιτών / ψηφοφόρων του Δήμου Σερρών στο διάστημα 15 Ιουνίου με 15 Ιουλίου 2007. Χρησιμοποιήθηκε η μέθοδος της στρωματοποιημένης δειγματοληψίας και της χιονοστιβάδας. Είναι σημαντικό, επίσης, να αναφέρουμε ότι στην προσπάθεια να συγκεντρωθεί το απαραίτητο δείγμα των 400 πολιτών / ψηφοφόρων, μοιράστηκαν 520 ερωτηματολόγια εκ των οποίων και απαντήθηκαν τα 400 (δηλαδή ποσοστό αποχής: 23,1%). Αν και συγκρινόμενο με το ποσοστό αποχής των εθνικών εκλογών του 2004 στο Δήμο Σερρών (26,64% και 32,83% το 2007) προκύπτει μια παρόμοια στάση / συμπεριφορά των ψηφοφόρων, εν τούτοις οι λόγοι που οδήγησαν τους 120 συγκεκριμένους πολίτες να μην απαντήσουν στην έρευνα θα ήταν ίσως αντικείμενο μιας άλλης μελέτης, προκειμένου να εξαχθούν χρήσιμα συμπεράσματα για τους ψηφοφόρους που τελικά απέχουν από την εκλογική διαδικασία.

Αρχικά πραγματοποιήσαμε μια προκαταρκτική ανάλυση της συμπεριφοράς του ψηφοφόρου. Ο καθορισμός της συνεπούς οικογένειας κριτηρίων είναι το πιο κρίσιμο σημείο στην εφαρμογή του μοντέλου. Επιπρόσθετα, θα πρέπει να δοθεί ιδιαίτερη προσοχή στην επιλογή των βαθμίδων της κλίμακας μέτρησης των κριτηρίων. Αυτές οι ποσοτικές και ποιοτικές κλίμακες έκφρασης της επιλογής των ψηφοφόρων θα πρέπει να πληρούν τις τρεις θεμελιώδεις ιδιότητες της συνεπούς οικογένειας κριτηρίων (μονοτονία, πληρότητα, μη πλεονασμός).

Με βάση, λοιπόν, τα αποτελέσματα της προκαταρκτικής μας ανάλυσης συντάξαμε το ερωτηματολόγιο. Το ερωτηματολόγιο πρέπει να είναι πλήρες και για την ευκολότερη επεξεργασία χωρίστηκε σε κατηγορίες παραγόντων επηρεασμού της ψήφου (γενικοί παράγοντες επηρεασμού, παράγοντες που αφορούν τα κόμματα, παράγοντες που αφορούν τους αρχηγούς). Οι ερωτήσεις είναι κλειστού τύπου και η κλίμακα αξιολόγησης που υιοθετήθηκε είναι εξαβάθμια με τις εξής επιλογές απάντησης: Πολύ, Αρκετά, Μέτρια, Λίγο, Καθόλου, Δεν ξέρω/Δεν απαντώ. Στη συνέχεια του ερωτηματολογίου ακολουθούν οι ερωτήσεις που αφορούν σε θέματα ψήφου (πρόθεση, βεβαιότητα, προηγούμενη ψήφος, παράσταση νίκης). Στο τέλος του ερωτηματολογίου υπάρχουν ερωτήσεις, προκειμένου να συλλεγούν γενικές πληροφορίες που αφορούν το δείγμα της έρευνάς μας και όσον αφορά τα δημογραφικά χαρακτηριστικά του, αλλά και τις πηγές ενημέρωσης και πληροφόρησής του σχετικά με τις πολιτικές εξελίξεις.

Με την πραγματοποίηση της έρευνάς μας, τη συλλογή των απαντημένων ερωτηματολογίων, δημιουργήσαμε μία βάση δεδομένων με την κωδικοποίηση των αποτελεσμάτων, ώστε να μπορέσουμε να προχωρήσουμε στην περεταίρω ανάλυση

των απαντήσεων / αποτελεσμάτων, τόσο με τη χρήση του Excel (στατιστική ανάλυση) όσο και μετέπειτα με τη χρήση του υπολογιστικού συστήματος Weka (Εξόρυξη Δεδομένων / Data Mining).

4.2 ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ

4.2.1 Παρουσίαση δημογραφικών στοιχείων

Στη συγκεκριμένη ενότητα του κεφαλαίου θα παρουσιάσουμε τα δημογραφικά στοιχεία του δείγματος, με μορφή σχημάτων, καθώς και σχήματα στατιστικής ανάλυσης εκείνων των κριτηρίων επηρεασμού της ψήφου που θα χρησιμοποιήσουμε και στην Εξόρυξη Δεδομένων (Data Mining), συνοδευόμενα από ένα σύντομο σχολιασμό.

Το δείγμα μας, όπως αναφέρθηκε και στην περίληψη είναι 400 άτομα, εκ των οποίων 221 είναι άντρες και 179 γυναίκες ο αριθμός και το ποσοστό των οποίων απεικονίζονται στο Σχήμα 4.2.1.1.



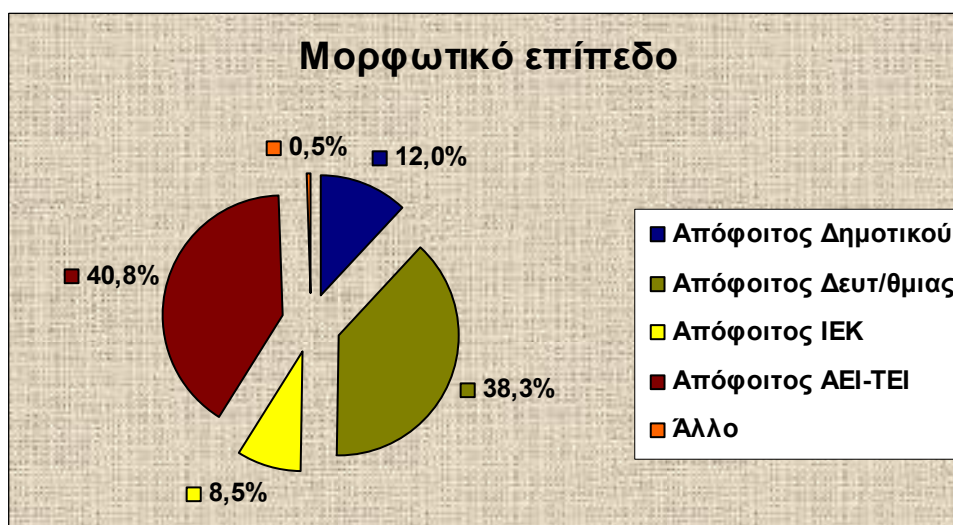
Σχήμα 4.2.1.1 Ποσοστά ανδρών-γυναικών δείγματος

Στη συνέχεια, παρουσιάζονται στο Σχήμα 4.2.1.2 τα ποσοστά των ηλικιακών ομάδων των ερωτηθέντων. Ο διαχωρισμός των ηλικιών έγινε λαμβάνοντας υπόψη τις διαφορετικές απαιτήσεις που μπορεί να έχουν από το πολιτικό σύστημα και τα κόμματα τα άτομα διαφορετικών ηλικιών. Παρατηρούμε ότι οι ηλικίες 26 έως 41 αποτελούν την πλειοψηφία του δείγματός μας με το ποσοστό τους να φθάνει το 72%



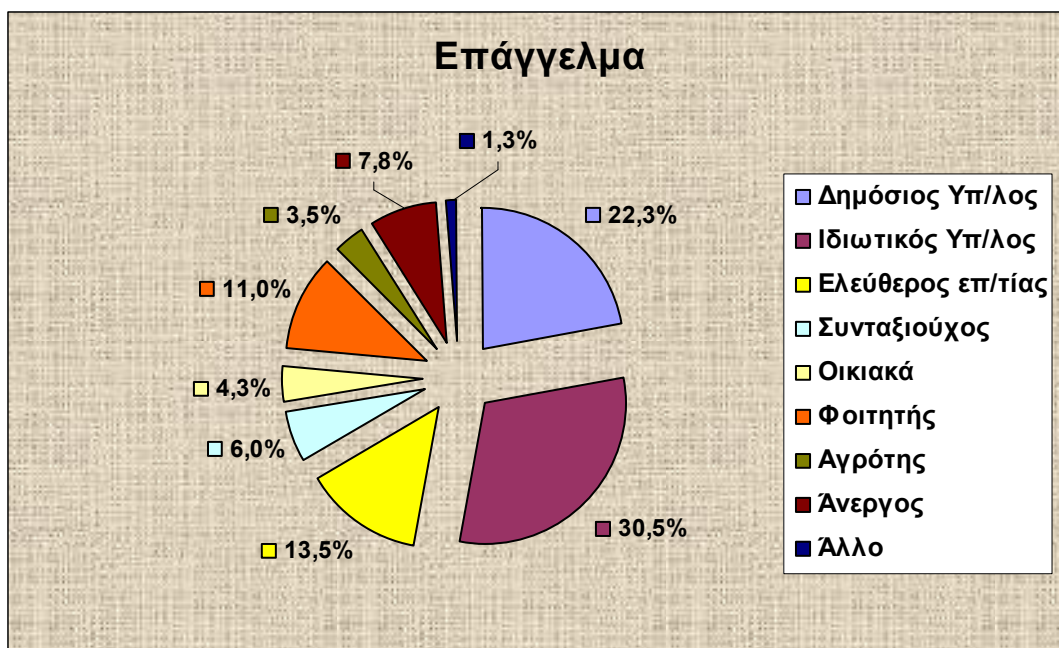
Σχήμα 4.2.1.2. Ποσοστά ερωτηθέντων ανά ηλικιακή ομάδα

Όσον αφορά το επίπεδο εκπαίδευσης / μόρφωσης των ερωτηθέντων στο Σχήμα 4.2.1.3 που ακολουθεί, παρατηρούμε δύο σημαντικές ομαδοποιήσεις: μία είναι αυτή που αφορά τους απόφοιτους Δευτεροβάθμιας εκπαίδευσης, ποσοστό 38,3%, και μία είναι αυτή που αφορά τους απόφοιτους ΑΕΙ – ΤΕΙ, ποσοστό 40,8%.



Σχήμα 4.2.1.3 Επίπεδο εκπαίδευσης ερωτηθέντων

Σε σχέση με το επάγγελμα των ερωτηθέντων, η πλειονηφία του δείγματός μας δήλωσε ότι είναι ιδιωτικοί και δημόσιοι υπάλληλοι (ποσοστό 30,5% και 22,3% αντίστοιχα). Ακολουθούν αυτοί που δήλωσαν ότι είναι ελεύθεροι επαγγελματίες (ποσοστό 13,5%) και φοιτητές (ποσοστό 11%), ενώ τα ποσοστά όλων των ατόμων που δήλωσαν και τα υπόλοιπα επαγγέλματα, φαίνονται στο Σχήμα 4.2.1.4.



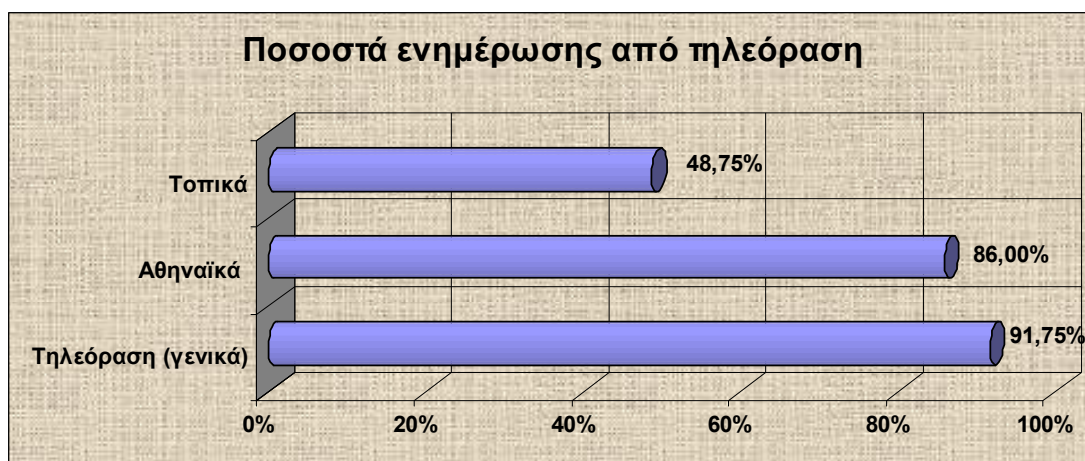
Σχήμα 4.2.1.4 Τα επαγγέλματα των ερωτηθέντων

Αναφορικά με τη συμμετοχή ή μη των ερωτώμενων σε οργανωμένες κομματικές δραστηριότητες προέκυψε ότι η συντριπτική πλειοψηφία των πολιτών απάντησε σε ποσοστό 89% πως δεν συμμετέχει, ενώ το 11%, όπως φαίνεται και στο Σχήμα 4.2.1.5, απάντησε πως είναι οργανωμένο σε κάποια κομματική οργάνωση.



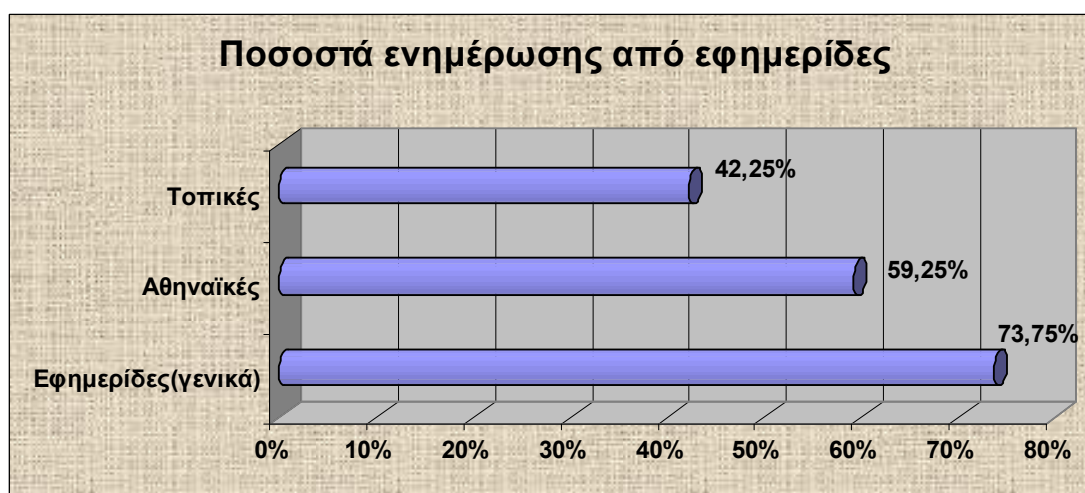
Σχήμα 4.2.1.5 Ποσοστά συμμετοχής σε κομματική οργάνωση

Στο Σχήμα 4.2.1.6 που ακολουθεί, αποτυπώνονται οι απαντήσεις των ψηφοφόρων του δείγματός μας σε σχέση με τα ποσοστά ενημέρωσής τους από την τηλεόραση, ενώ στο Σχήμα 4.2.1.7 τα ποσοστά ενημέρωσής τους από τις εφημερίδες.



Σχήμα 4.2.1.6 Ποσοστά ενημέρωσης από τηλεοπτικά δίκτυα

Εντύπωση προκαλούν τα υψηλά ποσοστά ενημέρωσης είτε από τα τοπικά τηλεοπτικά δίκτυα (48,75%) είτε από τις τοπικές εφημερίδες (42,25%).



Σχήμα 4.2.1.7 Ποσοστά ενημέρωσης από εφημερίδες

Τέλος, όσον αφορά την ενημέρωση από άλλες πηγές, από τις απαντήσεις του δείγματός μας προέκυψε, όπως αποτυπώνεται και στο Σχήμα 4.2.1.8, ότι οι συζητήσεις με φίλους για θέματα πολιτικής εξακολουθούν ακόμη και σήμερα να αποτελούν πηγή ενημέρωσης σε ποσοστό 44%, ενώ το διαδίκτυο προτιμά μόλις το 14,5%.



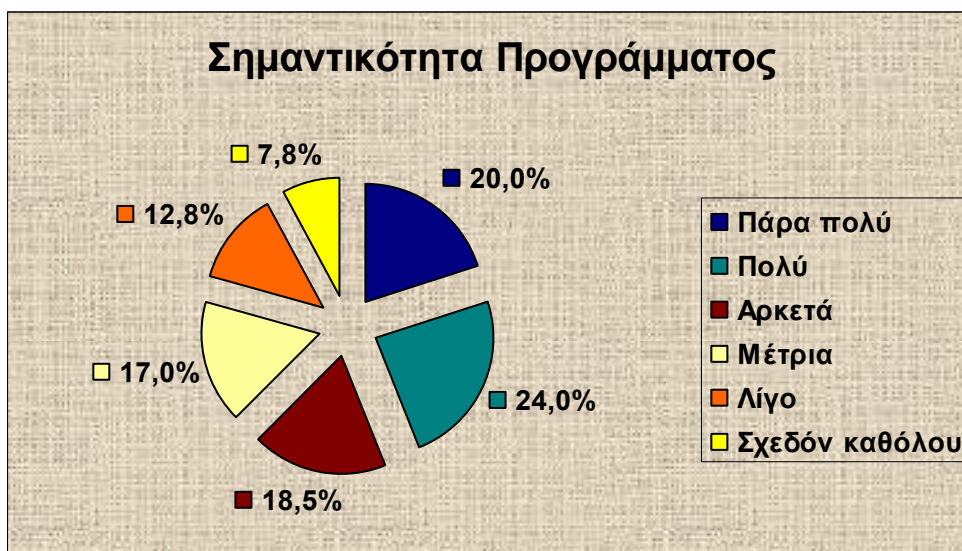
Σχήμα 4.2.1.8 Ποσοστά ενημέρωσης από άλλες πηγές

4.2.2 Παρουσίαση αποτελεσμάτων στατιστικής ανάλυσης

Σε αυτήν την ενότητα θα αναπτύξουμε εκείνα τα κριτήρια που θα χρησιμοποιήσουμε και μετέπειτα στην προτεινόμενη μεθοδολογία μας για την ποιοτική ανάλυση της συμπεριφοράς των ψηφοφόρων. Οι απαντήσεις των 400 πολιτών / ψηφοφόρων θα παρουσιαστούν σχηματικά με βάση το βαθμό επηρεασμού (Πολύ, Αρκετά, Μέτρια, Λίγο, Καθόλου, Δεν γνωρίζω) και με βάση το βαθμό σημαντικότητας (Πάρα πολύ, Πολύ, Αρκετά, Μέτρια, Λίγο, Σχεδόν καθόλου) που έδωσαν στα κριτήρια:

- Πρόγραμμα
- Υποψήφιοι βουλευτές
- Ιδεολογία
- Αρχηγός κόμματος
- Εμπιστοσύνη
- Προεκλογική εκστρατεία

Τα Σχήματα που θα ακολουθήσουν θα αποτελούν ενότητες σε σχέση με τη σημαντικότητα και τον επηρεασμό από τα παραπάνω κριτήρια, όπως προέκυψαν από τις απαντήσεις των ερωτηθέντων του δείγματός μας. Στο Σχήμα 4.2.2.1_α λοιπόν, που αφορά τη σημαντικότητα, παρατηρούμε ότι θεωρούν το πρόγραμμα ως κριτήριο ψήφου αρκετά έως πάρα πολύ σημαντικό το 62,5% των ερωτηθέντων και στο Σχήμα 4.2.2.1_β που αφορά το βαθμό επηρεασμού από το πρόγραμμα των κομμάτων, θεωρούν ότι επηρεάζονται από το κριτήριο αυτό το 70,8% του δείγματός μας.

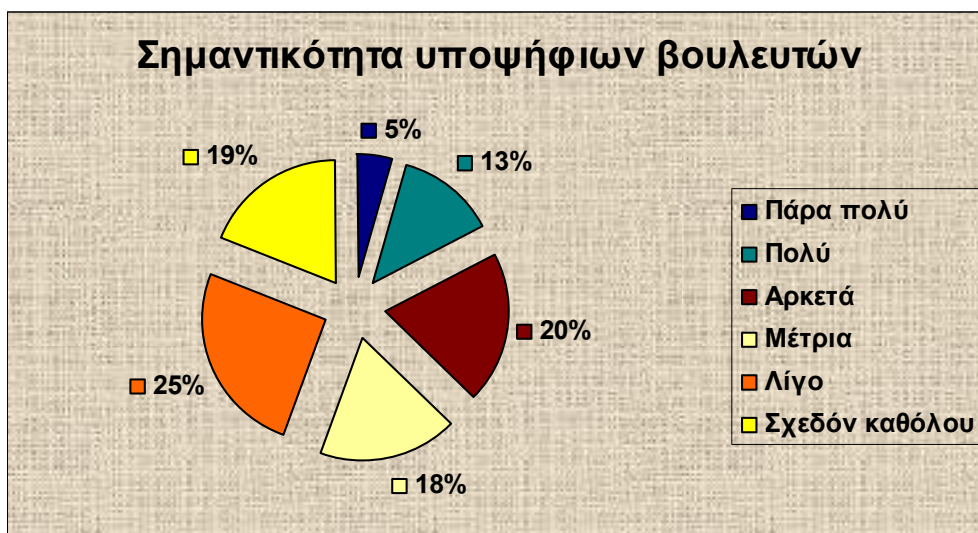


Σχήμα 4.2.2.1α. Πόσο σημαντικό είναι το πρόγραμμα των κομμάτων προκειμένου να διαλέξετε τι θα ψηφίσετε

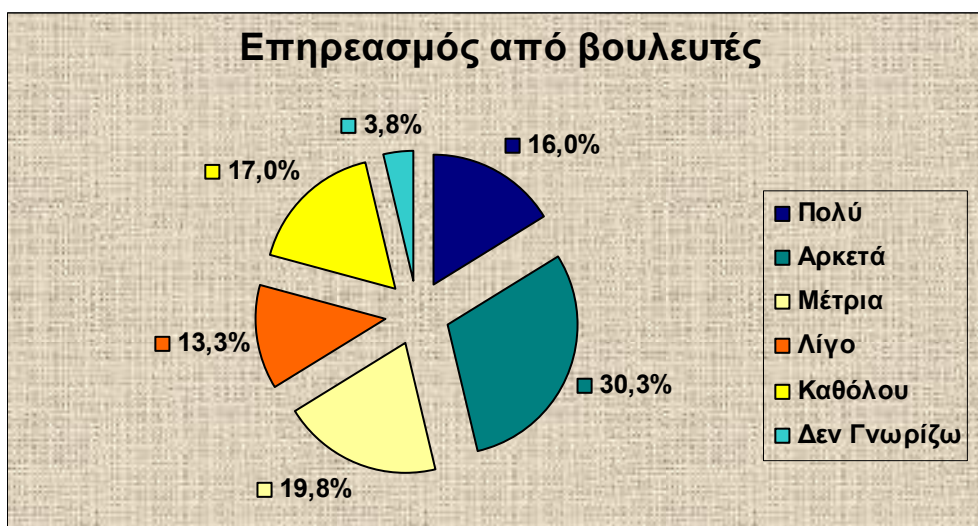


Σχήμα 4.2.2.1β. Πόσο σας επηρεάζει το πρόγραμμα των κομμάτων προκειμένου να διαλέξετε τι θα ψηφίσετε

Όσον αφορά τη σημαντικότητα και το βαθμό επηρεασμού από τη βουλευτική εκπροσώπηση των κομμάτων, παρατηρούμε στο Σχήμα 4.2.2.2_α ότι το θεωρεί αρκετά έως πάρα πολύ σημαντικό το 38%, ενώ λίγο και σχεδόν καθόλου σημαντικό, ως προς το τι θα ψηφίσουν, το θεωρεί το 44% (25% και 19% αντίστοιχα) των ερωτηθέντων. Εντούτοις, μας απάντησαν ότι επηρεάζονται πολύ και αρκετά, ως προς την επιλογή τους, το 46,3%, ενώ το 13,3% και 17% θεωρεί ότι επηρεάζονται λίγο και σχεδόν καθόλου αντίστοιχα, όπως φαίνεται και στο Σχήμα 4.2.2.2_β, από τη βουλευτική εκπροσώπηση / ομάδα των κομμάτων.

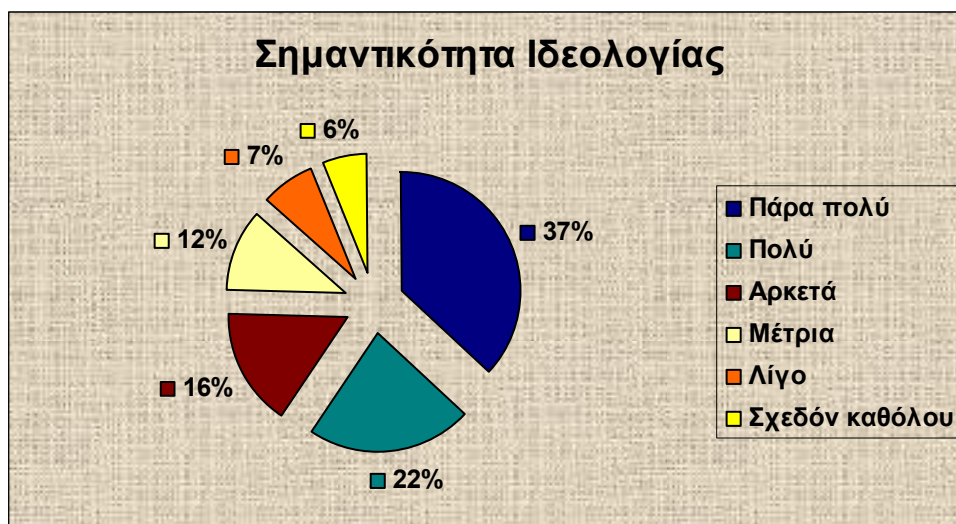


Σχήμα 4.2.2.2_α. Πόσο σημαντική είναι η βουλευτική ομάδα ενός κόμματος προκειμένου να διαλέξετε τι θα ψηφίσετε

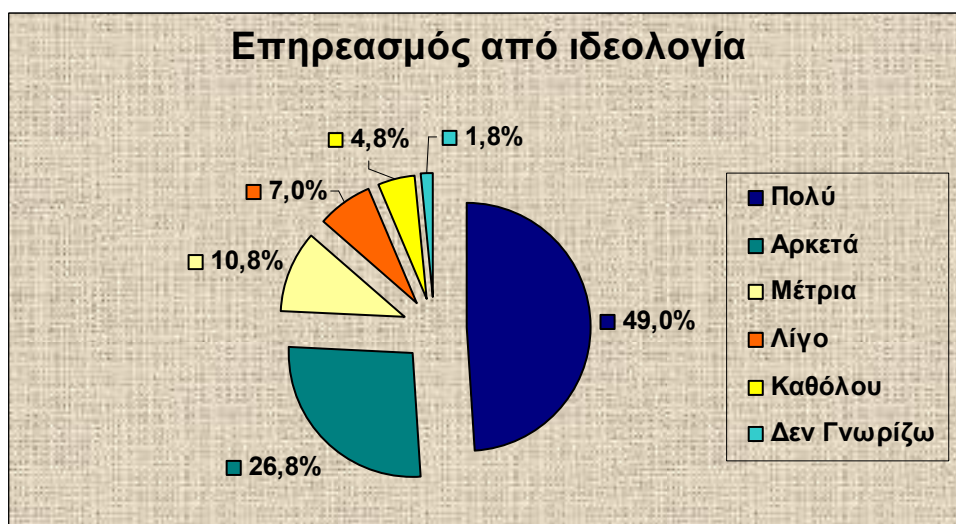


Σχήμα 4.2.2.2_β. Πόσο σας επηρεάζει η βουλευτική ομάδα ενός κόμματος προκειμένου να διαλέξετε τι θα ψηφίσετε

Στη συνέχεια, στο Σχήμα 4.2.2.3_α και 4.2.2.3_β θα δούμε αντίστοιχα πως απεικονίζεται η σημαντικότητα και ο βαθμός επηρεασμού από την ιδεολογική ταύτιση των ερωτηθέντων, προκειμένου να αποφασίσουν ποιο κόμμα θα ψηφίσουν. Αποτυπώνεται, λοιπόν, στο Σχήμα 4.2.2.3_α ότι πάρα πολύ, πολύ και αρκετά σημαντικό κριτήριο ψήφου θεωρούν την ιδεολογία το 75% (37%, 22% και 16% αντίστοιχα) των ερωτηθέντων. Καθώς επίσης, στο Σχήμα 4.2.2.3_β παρατηρούμε ότι η πλειοψηφία των ερωτηθέντων, ποσοστό 75,8%, απάντησε ότι επηρεάζονται πολύ και αρκετά (49% και 26,8% αντίστοιχα) από το προαναφερθέν κριτήριο.

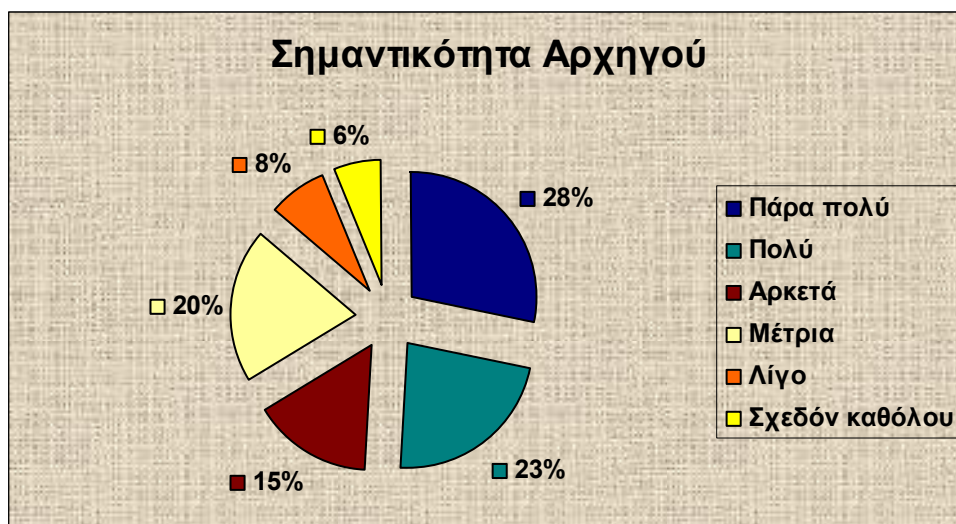


Σχήμα 4.2.2.3_α. Πόσο σημαντική είναι η ιδεολογία ενός κόμματος προκειμένου να διαλέξετε τι θα ψηφίσετε

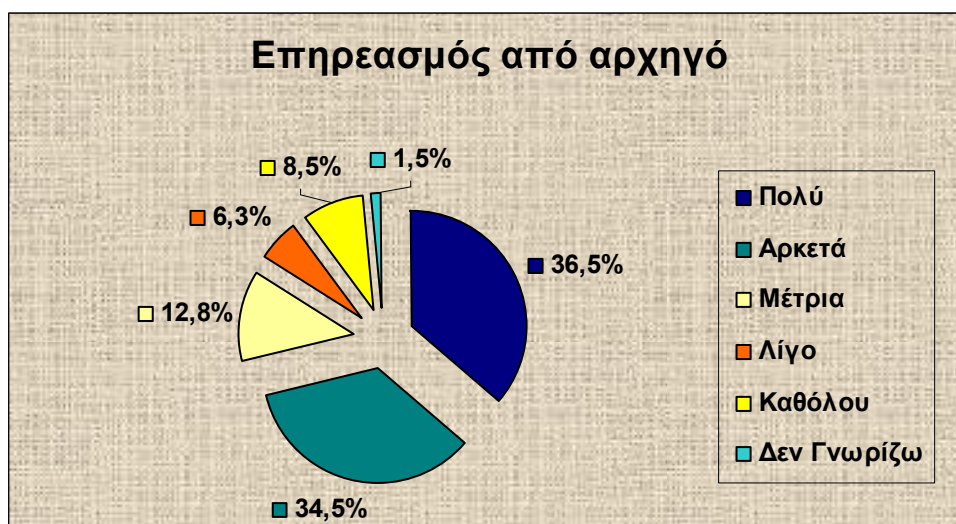


Σχήμα 4.2.2.3_β. Πόσο σας επηρεάζει η ιδεολογία ενός κόμματος προκειμένου να διαλέξετε τι θα ψηφίσετε

Αναφορικά με το κριτήριο του αρχηγού, το κατά πόσο σημαντικός είναι δηλαδή σε σχέση με την επιλογή της ψήφου των ερωτηθέντων, στο Σχήμα 4.2.2.4_α παρατηρούμε ότι το θεωρεί από αρκετά έως πάρα πολύ σημαντικό κριτήριο ψήφου το 66% του δείγματός μας. Επίσης η πλειοψηφία των ερωτηθέντων, ποσοστό 71%, δήλωσε, όπως απεικονίζεται και στο Σχήμα 4.2.2.4_β, ότι επηρεάζονται πολύ και αρκετά (36,5% και 34,5% αντίστοιχα), από το ποιος καθοδηγεί, ποιος είναι ο αρχηγός σε ένα κόμμα.

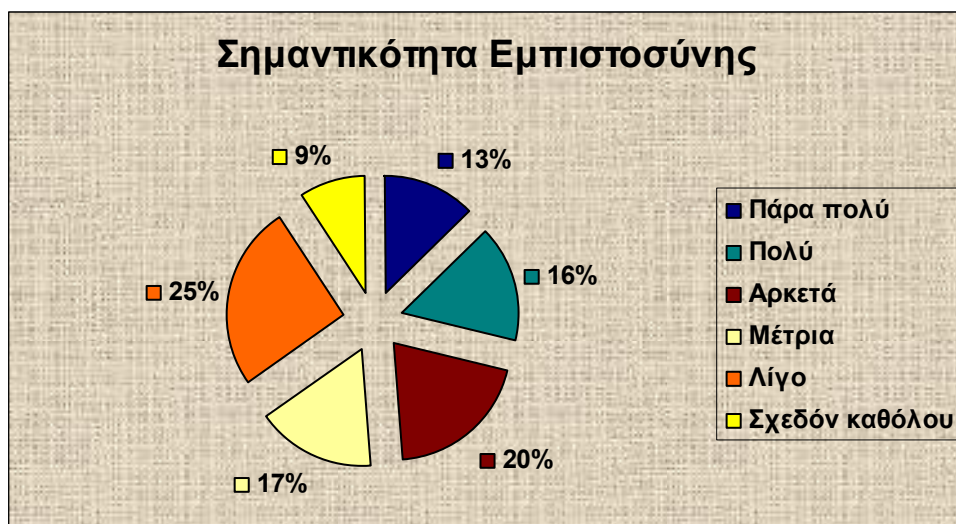


Σχήμα 4.2.2.4_α. Πόσο σημαντικός είναι ο αρχηγός ενός κόμματος προκειμένου να διαλέξετε τι θα ψηφίσετε

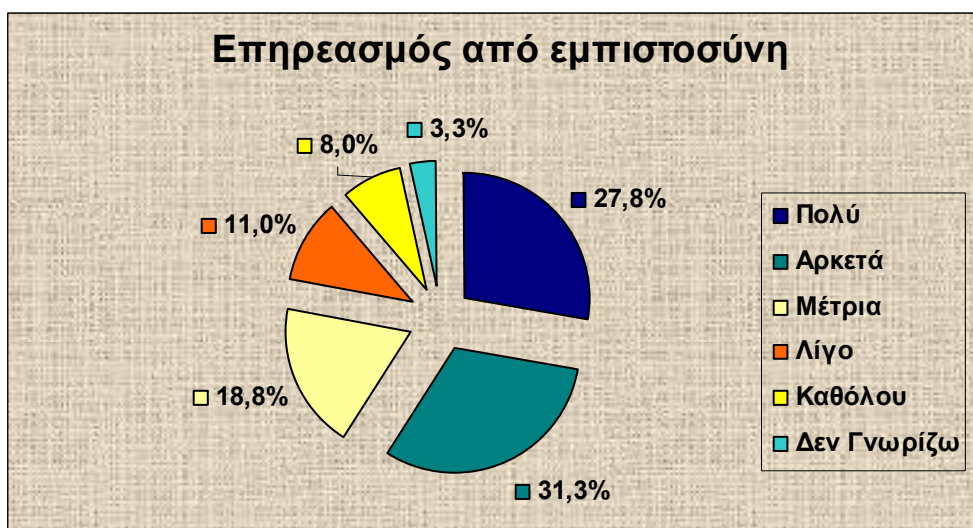


Σχήμα 4.2.2.4_β. Πόσο σας επηρεάζει ο αρχηγός ενός κόμματος προκειμένου να διαλέξετε τι θα ψηφίσετε

Σε σχέση με το κατά πόσο σημαντική είναι η εμπιστοσύνη σε ένα κόμμα, στο Σχήμα 4.2.2.5_α παρατηρούμε ότι τη θεωρεί μέτρια και λίγο σημαντική το 17% και το 25% του δείγματός μας, αντίστοιχα, ενώ πάρα πολύ και πολύ τη θεωρεί το 13% και το 16% των ερωτηθέντων αντίστοιχα. Αντίθετα, οι απαντήσεις που έδωσαν στο κατά πόσο επηρεάζονται, σε σχέση με το τι θα ψηφίσουν, από την εμπιστοσύνη προς ένα κόμμα, όπως αποτυπώνεται και στο Σχήμα 4.2.2.5_β, μας δείχνουν ότι σε ποσοστό 27,8% και 31,3% απαντούν πολύ και αρκετά αντίστοιχα.



Σχήμα 4.2.2.5α. Πόσο σημαντική είναι εμπιστοσύνη σας, προς ένα κόμμα, προκειμένου να διαλέξετε τι θα ψηφίσετε



Σχήμα 4.2.2.5β. Πόσο σας επηρεάζει η εμπιστοσύνη σας, προς ένα κόμμα, προκειμένου να διαλέξετε τι θα ψηφίσετε

Τέλος όσον αφορά τη σημαντικότητα από τη προεκλογική εκστρατεία των κομμάτων, παρατηρούμε στο Σχήμα 4.2.2.6α ότι η μεγάλη πλειοψηφία του δείγματός μας τη θεωρεί σχεδόν καθόλου σημαντική (ποσοστό 61%). Αντίστοιχα, στο Σχήμα 4.2.2.6β, όπου οι ερωτηθέντες μας δηλώνουν το κατά πόσο επηρεάζονται από την προεκλογική εκστρατεία των κομμάτων, το 26,5% μας απάντησε λίγο και το 34,8% μας απάντησε καθόλου (άθροισμα 61,3%), ενώ αν και μόλις το 5% μας δήλωσε ότι επηρεάζεται πολύ από την προεκλογική εκστρατεία των κομμάτων, εν τούτοις σε μία ρευστή εκλογική αναμέτρηση μπορεί να κρίνει και την τελική έκβαση της εκλογής – επιλογής κυβέρνησης.



Σχήμα 4.2.2.6α. Πόσο σημαντική είναι η προεκλογική εκστρατεία ενός κόμματος, προκειμένου να διαλέξετε τι θα ψηφίσετε



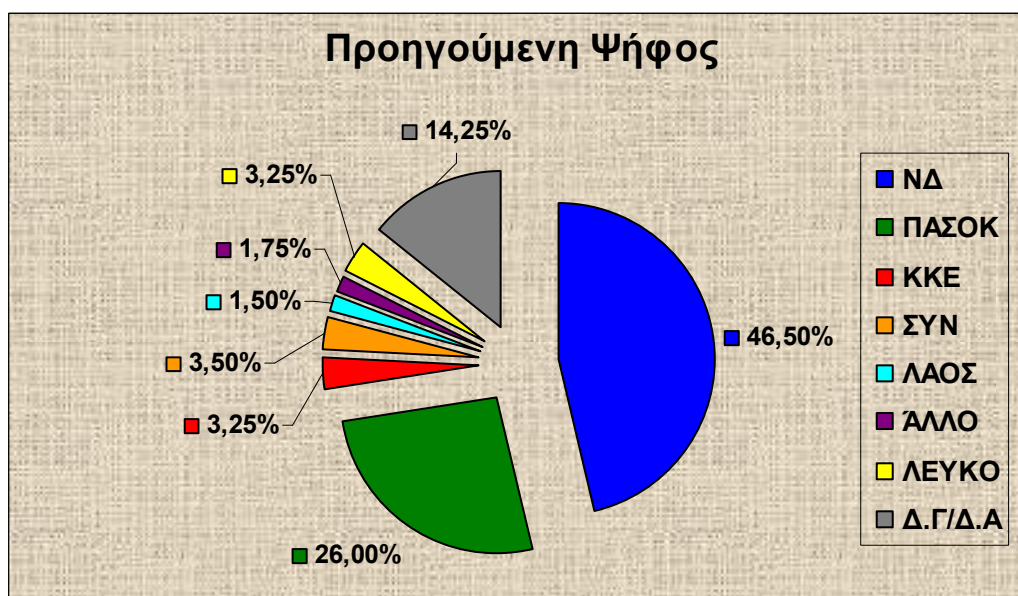
Σχήμα 4.2.2.6β. Πόσο σας επηρεάζει η προεκλογική καμπάνια/ επικοινωνία ενός κόμματος προκειμένου να διαλέξετε τι θα ψηφίσετε

Στη συνέχεια θα ασχοληθούμε με τα θέματα της ψήφου, δηλαδή την πρόθεση, την πρόβλεψη νίκης, τη βεβαιότητα καθώς και την ψήφο στις προηγούμενες εκλογές (του 2004) για να μπορέσουμε να έχουμε μία ολοκληρωμένη εικόνα της άποψης του δείγματός μας σχετικά με την εκλογική του τοποθέτηση, αλλά και για να ολοκληρώσουμε τη στατιστική μας ανάλυση και να μπορέσουμε να εντυπώσουμε στην “εξόρυξη γνώσης” (data mining) και να καταλήξουμε στα συμπεράσματά μας. Στο Σχήμα 4.2.2.7, λοιπόν, αποτυπώνεται η πρόθεση ψήφου των 400 ερωτηθέντων του δείγματός μας, από όπου παρατηρούμε ότι η Νέα Δημοκρατία (ποσοστό 34,75%) υπερτερεί του ΠΑ.ΣΟ.Κ (24,75%) κατά 10 ποσοστιαίες μονάδες, το Κ.Κ.Ε και το ΛΑ.Ο.Σ συγκεντρώνουν το ίδιο ποσοστό (6,25%) και το Λευκό/άκυρο καθώς και η αδιευκρίνιστη ψήφος (Δεν γνωρίζω/δεν απαντώ) συγκεντρώνουν διψήφια ποσοστά (11,75% και 12,5% αντίστοιχα).



Σχήμα 4.2.2.7. Αν την Κυριακή είχαμε εκλογές πιο κόμμα θα ψηφίζατε

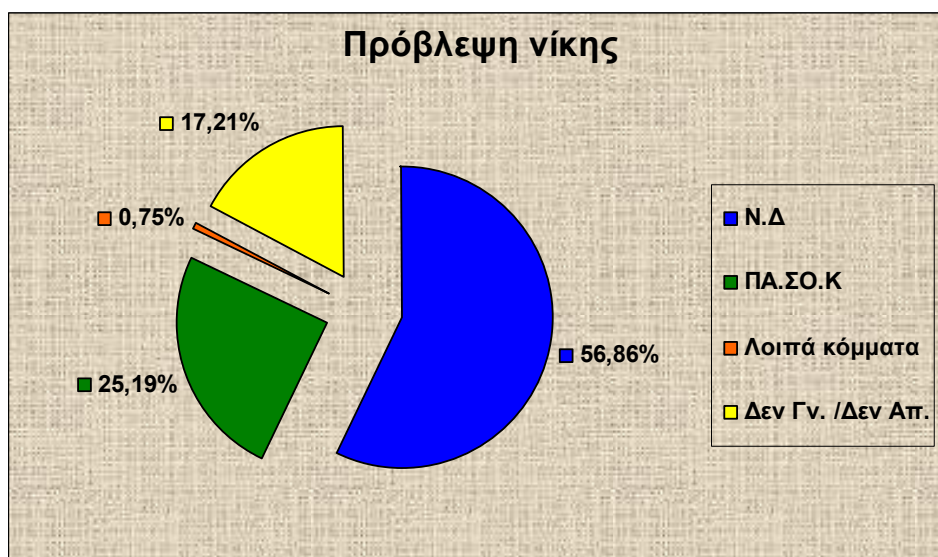
Οι διαφορές που προκύπτουν όσον αφορά τη ψήφο στις προηγούμενες βουλευτικές εκλογές (Σχήμα 4.2.2.8) με την προαναφερθείσα πρόθεση ψήφου είναι αξιοσημείωτες. Παρατηρούμε, λοιπόν, ότι η Νέα Δημοκρατία συγκέντρωσε 46,5%, δηλαδή σχεδόν 12 ποσοστιαίες μονάδες υψηλότερα σε σχέση με τη τωρινή πρόθεση ψήφου. Επίσης, παρατηρούμε ότι αν και η αδιευκρίνιστη ψήφος συγκεντρώνει εξίσου, με την πρόθεση ψήφου, διψήφια ποσοστά (14,25%), εν τούτοις η επιλογή Λευκό/άκυρο συγκεντρώνει μόλις 3,25%, 8,5 ποσοστιαίες μονάδες (8,5%) λιγότερο απ' ό,τι στην πρόθεση ψήφου.



Σχήμα 4.2.2.8. Τι ψηφίσατε στις προηγούμενες βουλευτικές εκλογές

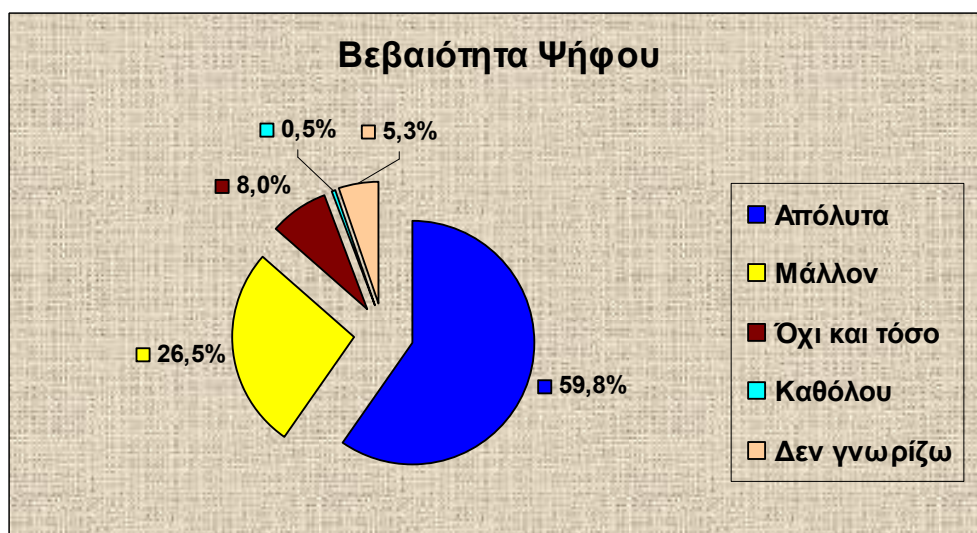
Σχετικά με την πρόβλεψη νίκης, δηλαδή σχετικά με το ποιο κόμμα πιστεύουν οι ερωτηθέντες ότι θα κέρδιζε τις εκλογές, ανεξάρτητα από το τι θα ψηφίζαν, οι

διαφορές είναι ξεκάθαρες. Έτσι, όπως αποτυπώνεται και στο Σχήμα 4.2.2.9 η Νέα Δημοκρατία συγκεντρώνει την πλειοψηφία των ερωτηθέντων (ποσοστό 56,86%), ενώ το ΠΑ.ΣΟ.Κ συγκεντρώνει το 25,19% δείγματός μας.



Σχήμα 4.2.2.9. Ανεξάρτητα από το κόμμα που θα ψηφίζατε αν την Κυριακή είχαμε εκλογές, ποιο κόμμα πιστεύετε ότι θα τις κέρδιζε

Αξιοσημείωτο είναι επίσης και το πόσο βέβαιοι δήλωσαν οι ερωτηθέντες του δείγματός μας σχετικά με την επιλογή της ψήφου τους. Παρατηρούμε, λοιπόν, στο Σχήμα 4.2.2.10 ότι αν και η πλειοψηφία, μας απάντησε ότι είναι απόλυτα βέβαιοι (ποσοστό 59,8%) σχετικά με την επιλογή τους, εντούτοις το 26,5% δήλωσε μάλλον βέβαιοι και το 8% δήλωσε όχι και τόσο βέβαιοι, κάτι που, όπως προαναφέραμε, σε μία ρευστή εκλογική αναμέτρηση μπορεί να κρίνει και το τελικό αποτέλεσμα.



Σχήμα 4.2.2.10. Πόσο βέβαιος είστε για την επιλογή της ψήφου σας

4.3 ΠΟΙΟΤΙΚΗ ΑΝΑΛΥΣΗ:

Παρουσίαση αποτελεσμάτων ποιοτικής ανάλυσης με τη χρήση του υπολογιστικού συστήματος Weka

4.3.1 Μια περιγραφική ανάλυση του υπολογιστικού συστήματος Weka: Wekato Environment for knowledge Analysis

Το Weka είναι ένα λογισμικό (software) Εξόρυξης Δεδομένων, γραμμένο σε γλώσσα προγραμματισμού JAVA, το οποίο περιέχει υλοποιημένες μεθόδους για:

- Προ-επεξεργασία Δεδομένων (Pre-process)
- Ταξινόμηση (Classify)
- Συσταδοποίηση (Clustering)
- Εύρεση Κανόνων Συσχέτισης (Associate)

Παρατίθενται εν συντομία τα γενικά χαρακτηριστικά αυτού του υπολογιστικού συστήματος:

- Έχει ένα φιλικό γραφικό περιβάλλον εργασίας
- Όλα τα εργαλεία του Weka: φίλτρα, κατηγοριοποιητές, ταξινομητές, συσχετιστές και επιλογείς χαρακτηριστικών μπορούν να παρουσιαστούν γραφικά με δυσδιάστατα γραφήματα (Visualization)
- Με βάση τις γραφικές αναπαραστάσεις των επιλεγμένων δεδομένων δοκιμάζονται και προβλέπεται η απόδοση των εναλλακτικών μοντέλων που θα επιλεγούν να χρησιμοποιηθούν τελικά για την διαμόρφωση της «θαμμένης» γνώσης.
- Είναι επιλέξιμα μέσα από αναδυόμενα μενού

Επιλέγοντας ένα σύνολο δεδομένων (αρχείο .arff), εμφανίζονται γραφικά τα δεδομένα για καθένα από τα γνωρίσματα ξεχωριστά καθώς και στατιστικές πληροφορίες για αυτά. Εάν στο σύνολο των δεδομένων δίνεται και κάποια κλάση στην οποία ταξινομούνται, τα δεδομένα που ανήκουν στην ίδια κλάση, εμφανίζονται με το ίδιο χρώμα

- Τα αρχεία που περιέχουν το σύνολο δεδομένων πρέπει να έχουν συγκεκριμένο format και να αποθηκεύονται με την επέκταση .arff
- Δεδομένα μπορούν επίσης να δοθούν από ένα URL ή από μία SQL βάση.

Εμείς, μετά τη συλλογή των 400 ερωτηματολογίων, κωδικοποιήσαμε τις ερωτήσεις και τις απαντήσεις σ' ένα υπολογιστικό φύλλο επεξεργασίας Excel, όπου όπως φαίνεται στο Σχήμα 4.3.1.1 και 4.3.1.2, δημιουργήσαμε έναν πίνακα 227 (στήλες) x 400 (σειρές). Στις στήλες κωδικογραφήσαμε τις ερωτήσεις – κριτήρια και στις γραμμές / σειρές, τις κλίμακες των απαντήσεων για κάθε μία από τις ερωτήσεις και για κάθε ένα ερωτηματολόγιο (σύνολο 400 ερωτηματολόγια στο Δήμο Σερρών).

Microsoft Excel - epeksergasia_teliko.xls

File Edit View Insert Format Tools Data Window Help

Type a question for help

100%

Arial 10

B I U

Reply with Changes... End Review...

L102 4

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1															
2	A/A	A1	A2	A3	A4	επηρεασμός από Βουλευτές		A7	B1	B2	B3	B4	B5	B6	B7
3	1	4	4	3	3	3	2	99	2	4	3	1	5	6	99
4	2	5	2	1	1	3	3	99	5	6	4	1	2	3	99
5	3	4	5	5	4	5	3	5	6	4	6	5	6	3	99
6	4	5	5	5	5	4	3	99	2	4	1	3	5	6	99
7	5	3	4	4	3	3	2	99	2	4	1	5	3	6	99
8	6	3	4	4	4	4	2	99	1	3	2	4	5	6	99
9	7	5	5	4	4	4	4	99	3	5	2	1	4	6	99
10	8	4	3	4	3	3	3	99	2	6	5	2	5	2	99
11	9	5	5	5	1	5	1	99	2	6	1	4	3	5	99
12	10	5	5	5	3	5	0	99	1	4	3	2	5	6	99
13	11	5	5	5	5	5	5	99	2	2	1	1	2	2	99
14	12	5	5	5	5	5	5	99	2	1	2	1	1	1	99
15	13	5	5	4	3	4	2	99	4	5	1	2	3	6	99
16	14	5	2	1	2	3	3	99	4	5	6	1	2	3	99
17	15	5	4	4	5	5	3	99	4	6	5	6	6	4	99
18	16	5	5	4	1	4	2	99	4	3	1	2	5	6	99
19	17	4	5	1	3	2	2	99	3	4	1	2	6	5	99
20	18	5	5	5	5	5	5	99	5	3	2	1	4	6	99
21	19	5	5	4	5	5	3	99	5	3	2	1	4	6	99
22	20	4	5	4	3	4	3	99	4	5	1	2	3	6	99
23	21	4	5	4	5	5	1	99	5	4	1	3	2	6	99
24	22	4	4	5	4	3	1	99	1	4	2	3	5	6	99
25	23	5	5	4	4	5	5	99	6	3	1	2	4	5	99
26	24	4	2	2	4	3	2	99	3	2	4	1	6	5	99
27	25	4	5	5	5	5	1	99	1	1	1	2	1	6	99
28	26	5	4	5	5	5	1	99	2	3	5	1	4	6	99
29	27	5	4	5	5	4	1	99	3	4	5	1	2	6	99
30	28	5	4	5	5	4	2	99	2	4	3	1	5	6	99
31	29	3	5	3	3	3	3	99	6	3	6	5	4	1	99
32	30	5	3	5	5	4	3	99	2	3	4	1	5	6	99
33	31	5	5	5	5	5	5	99	5	4	1	2	3	6	99

Geniko / nd / pasok / kke / syn / laos / allo / leykoakyro / denapantw / Sheet1 /

Cell E2 commented by anesiadis

NUM

Σχήμα 4.3.1.1 Κωδικογράφηση ερωτήσεων / κριτηρίων και των απαντήσεων του ερωτηματολογίου της έρευνάς μας

Microsoft Excel - epeksergasia_teliko.xls																
File Edit View Insert Format Tools Data Window Help																
Type a question for help																
100% Arial																
Reply with Changes... End Review...																
HQ2	Rc1															
	HF	HG	HH	HI	HJ	HK	HL	HM	HN	HO	HP	HQ	HR	HS	HT	
1																
2	Q1	Q2	Q3	Q4	Q5	Ra1	Ra2	Ra3	Rb1	Rb2	Rb3	Rc1	ΕΝΗΜΕΡΩΣΗ από Διαδίκτυο (Internet)		Rf1	
372	2	25	4	1	1	2	2	2	2	2	2	1	1	2	99	
373	1	30	4	7	1	2	2	1	2	2	1	2	1	2	99	
374	2	25	4	1	1	2	2	1	2	2	2	1	1	2	99	
375	2	31	2	2	1	2	2	2	2	2	2	1	1	1	99	
376	2	25	4	1	1	2	2	2	2	2	2	1	1	2	99	
377	1	34	1	1	1	2	2	1	1	1	1	1	1	2	99	
378	1	43	4	8	1	2	2	2	2	2	2	1	2	2	99	
379	2	24	4	1	1	2	2	2	1	1	1	1	1	2	99	
380	2	35	2	7	1	2	2	2	1	1	1	1	1	2	99	
381	2	31	4	1	1	2	2	1	1	1	1	2	1	2	99	
382	1	33	2	7	1	2	2	2	2	2	2	1	1	1	99	
383	2	57	2	4	1	2	2	1	1	1	1	1	1	2	99	
384	2	40	2	4	1	2	2	1	2	2	1	1	1	1	99	
385	2	41	2	7	1	2	2	1	2	2	1	1	1	1	99	
386	1	38	3	7	1	2	2	1	2	2	1	1	1	1	99	
387	2	30	3	7	1	2	2	2	2	2	2	2	1	2	99	
388	2	28	4	1	1	2	2	2	2	2	2	2	2	2	99	
389	2	50	1	7	1	2	2	2	2	2	2	1	1	2	99	
390	1	60	2	5	1	2	2	1	2	2	1	1	1	1	99	
391	2	35	2	1	1	2	2	2	2	2	2	1	1	1	99	
392	2	27	4	1	1	2	2	2	2	2	2	1	1	2	99	
393	2	60	2	2	1	2	2	2	1	1	1	1	1	1	99	
394	2	48	4	8	1	2	2	2	2	2	2	1	1	2	99	
395	2	24	4	7	1	2	2	2	2	2	2	1	1	1	99	
396	2	55	2	8	1	2	2	2	2	2	2	1	1	2	99	
397	2	18	4	3	1	2	2	1	2	2	1	1	1	2	99	
398	1	51	2	7	2	2	2	1	2	2	1	1	1	2	99	
399	2	25	4	1	1	2	2	1	2	2	1	1	1	2	99	
400	2	41	2	7	1	2	2	2	1	1	1	1	1	2	99	
401	2	24	4	7	1	2	2	2	2	1	2	1	1	2	99	
402	2	22	4	6	1	1	1	1	1	1	1	1	1	2	99	
Geniko / nd / pasok / kke / syn / laos / allo / leykoakryo / denapantw / Sheet1 /																
Cell HQ2 commented by anesiadis																
NUM																

Σχήμα 4.3.1.2 Κωδικογράφηση ερωτήσεων / κριτηρίων και των απαντήσεων του ερωτηματολογίου της έρευνάς μας

Microsoft Excel - ekpaideftiko.xls

File Edit View Insert Format Tools Data Window Help

Type a question for help

100% Arial 10 B I U

Reply with Changes... End Review...

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Programme	Vouleftes	Ideologia	Leader	Trust	Communication	komma							
2	2	5	3	1	6	4	Nd							
3	6	4	5	1	2	3	Nd							
4	4	5	4	6	4	3	Nd							
5	2	5	1	6	3	4	Nd							
6	1	3	2	5	6	4	Nd							
7	3	6	2	1	5	4	Nd							
8	2	4	1	5	3	6	Nd							
9	1	5	3	2	6	4	Nd							
10	2	2	1	1	2	2	Nd							
11	5	6	1	2	3	4	Nd							
12	5	6	4	1	2	3	Nd							
13	5	4	6	4	4	5	Nd							
14	3	5	1	2	4	6	Nd							
15	6	3	2	1	5	4	Nd							
16	6	3	2	1	5	4	Nd							
17	6	5	1	3	2	4	Nd							
18	1	5	2	3	6	4	Nd							
19	4	3	1	2	5	6	Nd							
20	1	1	1	2	1	4	Nd							
21	2	3	6	1	5	4	Nd							
22	3	5	6	1	2	4	Nd							
23	4	3	4	6	5	1	Nd							
24	2	3	5	1	6	4	Nd							
25	6	5	1	2	3	4	Nd							
26	2	3	1	1	1	5	Nd							
27	6	3	5	1	2	4	Nd							
28	2	5	3	1	6	4	Nd							
29	5	1	3	1	1	3	Nd							
30	6	5	1	2	3	4	Nd							
31	6	3	1	2	5	4	Nd							
32	2	3	5	1	6	4	Nd							
33	1	6	2	3	5	4	Nd							

Φύλλο1 / Φύλλο2 / Φύλλο3

Ready NUM

Σχήμα 4.3.1.3 Επιλογή κριτηρίων και των απαντήσεων για τη δημιουργία αρχείου επεξεργασίας Weka

Από την παραπάνω επεξεργασία προκρίναμε για την καλύτερη και πιστότερη ανάλυση του δείγματός μας τα κριτήρια που αφορούν το πρόγραμμα , τη βουλευτική ομάδα κάθε κόμματος, την ιδεολογία, την εμπιστοσύνη, την επικοινωνία καθώς και την τελική επιλογή κάθε ερωτώμενου σε σχέση με τη ψήφο του σε ενδεχόμενες βουλευτικές εκλογές (Σχήμα 4.3.1.3). Επίσης, είναι σημαντικό να αναφέρουμε ότι για την αναλογική κατανομή του δείγματος ακολουθήσαμε την εξής μέθοδο δειγματοληψίας – επιλογής των ερωτηματολογίων, για τη περαιτέρω ποιοτική ανάλυση με τη χρήση του Weka: Επιλέγαμε τρία (3) ερωτηματολόγια στη σειρά, για το «εκπαιδευτικό» μας δείγμα, αφήνοντας το τέταρτο (στη σειρά) για το «επαληθευτικό» μας δείγμα. Έτσι, επιλέξαμε τα 300 ερωτηματολόγια που θα μας χρησιμεύσουν στην εξαγωγή γνώσης και συμπερασμάτων ως το εκπαιδευτικό μας δείγμα και τα υπόλοιπα 100 σαν επαληθευτικό δείγμα, προκειμένου να ελέγξουμε την όλη διαδικασία Εξόρυξης Γνώσης. Το υπολογιστικό φύλλο excel, το μετατρέψαμε σε αρχείο *.csv (για να αποτελέσει αποδεκτή μορφή), προκειμένου να μπορέσουμε εν συνεχεία με τις απαραίτητες μετατροπές να καταλήξουμε σε ένα αρχείο .arff (μορφή εισόδου για το Weka) που είναι και το αρχείο που μπορεί να επεξεργαστεί το υπολογιστικό σύστημα Weka.

Ένα μέρος του αρχείου, από το σύνολο των 300 απαντήσεων (300 σειρές), στη μορφή .arff, παραθέτουμε παρακάτω:

@relation ekpaideftiko

@attribute Programme real

@attribute Vouleftes real

@attribute Ideologia real

@attribute Leader real

@attribute Trust real

@attribute Communication real

@attribute komma {Nd,Pasok,Kke,Syn,Laos,Allo,Denapantw}

@data

2,5,3,1,6,4,Nd

6,4,5,1,2,3,Nd

4,5,4,6,4,3,Nd

2,5,1,6,3,4,Nd

1,3,2,5,6,4,Nd

3,6,2,1,5,4,Nd

2,4,1,5,3,6,Nd

1,5,3,2,6,4,Nd

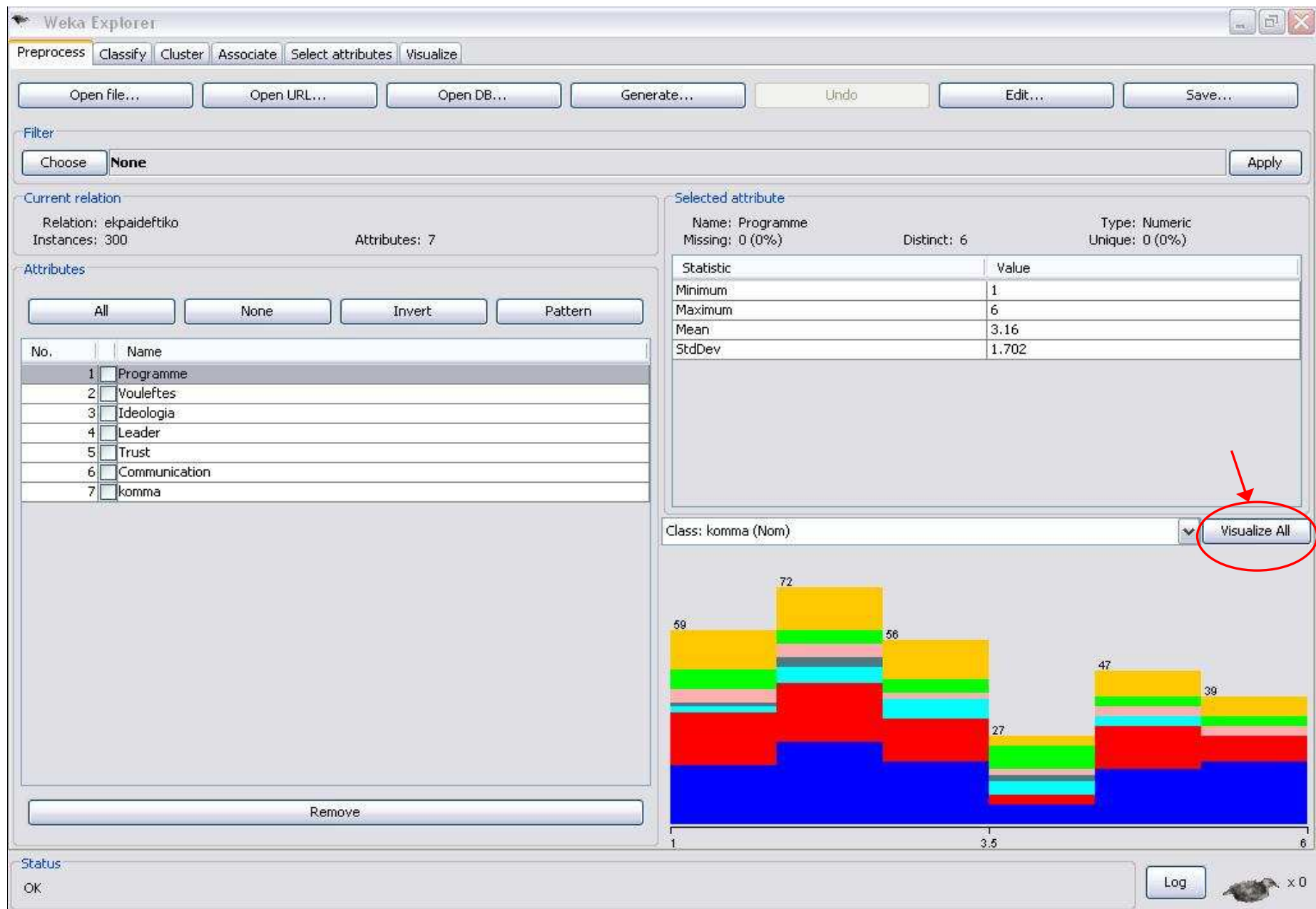
2,2,1,1,2,2,Nd
 5,6,1,2,3,4,Nd
 5,6,4,1,2,3,Nd

 6,5,1,3,2,4,Pasok
 1,3,2,1,2,3,Pasok
 5,4,3,6,1,4,Pasok
 1,5,1,2,6,3,Pasok
 1,4,3,5,2,6,Pasok
 2,2,1,1,2,4,Pasok
 5,6,3,1,2,4,Kke
 3,6,2,5,1,4,Kke
 3,6,2,5,1,4,Kke
 1,2,6,3,5,4,Syn
 3,3,2,1,2,5,Laos
 6,5,1,3,2,4,Laos
 2,3,1,5,6,4,Laos
 1,2,4,3,5,6,Allo
 4,2,5,3,6,1,Allo
 5,3,2,1,6,4,Allo
 2,6,1,5,3,4,Allo
 6,4,4,4,4,4,Denapantw
 3,6,1,2,5,4,Denapantw
 1,3,2,5,6,4,Denapantw
 6,4,4,4,4,5,Allo
 5,2,3,1,6,4,Nd
 1,4,4,4,4,4,Nd
 3,5,1,2,6,4,Allo
 2,6,1,3,5,4,Laos
 1,5,3,2,6,4,Nd

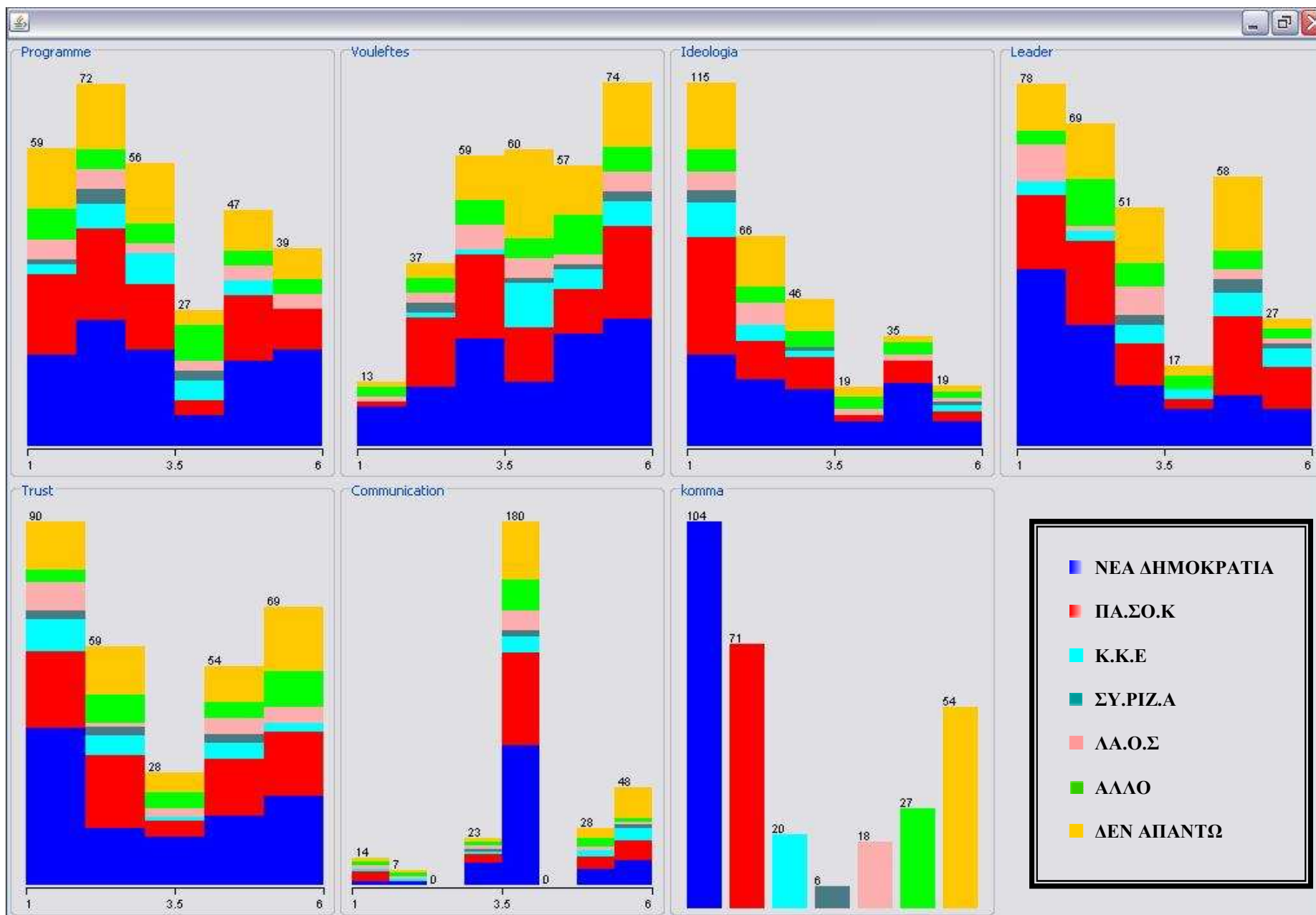
 1,4,2,3,5,6,Denapantw
 2,6,1,4,5,3,Nd
 3,5,1,2,6,4,Pasok
 4,2,4,6,6,1,Pasok
 2,6,1,5,3,4,Denapantw

4.3.2 Προ – επεξεργασία Δεδομένων

Είναι το σημείο εκκίνησης Εξόρυξης Γνώσης. Από εδώ μπορεί να «φορτώσει» κανείς δεδομένα, να εξερευνήσει χαρακτηριστικά των διαφόρων εγγραφών και να εφαρμόσει οποιοδήποτε συνδυασμό «φίλτρων» που διαθέτει το Weka στα δεδομένα εισόδου. Στο Σχήμα 4.3.2.1 βλέπουμε πως απεικονίζονται στο Weka τα δεδομένα που έχουμε εισάγει, δηλαδή τα κριτήρια με τις απαντήσεις των ερωτηθέντων του δείγματός μας. Επιλέγοντας το **“Visualize All”** προκύπτει το Σχήμα 4.3.2.2, όπου και απεικονίζονται πλήρως οι προτιμήσεις των ερωτηθέντων σε κάθε κριτήριο μας ξεχωριστά. Η κλίμακα ικανοποίησης είναι από το ένα (1) έως το έξι (6), όπου το 1 είναι η απόλυτα θετική γνώμη / απάντηση και 6 είναι η απόλυτα αρνητική γνώμη / απάντηση.



Σχήμα 4.3.2.1 Απεικόνιση των κριτηρίων (Weka Visualize) σύμφωνα με τις απαντήσεις των ερωτηθέντων



Σχήμα 4.3.2.2 Απεικόνιση των κριτηρίων (Weka Visualize) σύμφωνα με τις απαντήσεις των ερωτηθέντων

4.3.3 Συσταδοποίηση Δεδομένων

Αφού έχει επιλεγεί ένα σύνολο δεδομένων, είναι δυνατόν να γίνει συσταδοποίηση / ομαδοποίηση αυτών των δεδομένων (εύρεση ομάδων ‘όμοιων’ δεδομένων). Από την καρτέλα cluster μπορεί να επιλεγεί ένας αλγόριθμος με βάση τον οποίο θα γίνει η συσταδοποίηση και με το κουμπί start να αρχίσει η εκτέλεση του αλγόριθμου. Οι αλγόριθμοι συσταδοποίησης που ενυπάρχουν είναι:

- Cobweb (ιεραρχική συσταδοποίηση)
- DB Scan
- E.M
- Farthest First
- OPTICS
- Simple K-means (K-means)
- X-means

Με δεξί κλικ πάνω στο όνομα του αλγορίθμου που έχει ήδη εκτελεστεί στα δεδομένα (από το result list) εμφανίζεται ένα μενού με την βοήθεια του οποίου μπορεί να εμφανιστεί η γραφική αναπαράσταση των δεδομένων, με βάση τις ομάδες που έχουν προκύψει από το clustering. Εμείς, για την επεξεργασία του δείγματός μας, επιλέξαμε τον αλγόριθμο K-means για τον οποίο θα αναφερθούμε περιληπτικά και θα παρουσιάσουμε και τα αποτελέσματα του δικού μας case study.

4.3.3.1 Ο αλγόριθμος K-Means για τη μη εποπτευόμενη εκμάθηση (Unsupervised Learning)

Ο αλγόριθμος k-means παίρνει ένα σύνολο δεδομένων και το χωρίζει σε **k** συστάδες, με το χρήστη να καθορίζει πόσες συστάδες πρόκειται να δημιουργηθούν, χρησιμοποιώντας την ακόλουθη γενική διαδικασία:

- Οι περιπτώσεις ορίζονται αρχικά τυχαία σε **k** συστάδες. Φανταστείτε ότι χωρίζετε μια τράπουλα σε δύο μέρη ($k=2$).
- Έπειτα οι περιπτώσεις “κινούνται” γύρω και μεταξύ των συστάδων, χρησιμοποιώντας μια επαναληπτική μέθοδο, έτσι ώστε η ταξινόμηση να παράγεται με τέτοιο τρόπο που οι συστάδες θα πρέπει να είναι εσωτερικά παρόμοιες αλλά και εξωτερικά ανόμοιες από τις άλλες συστάδες.
- Σταματάμε όταν με τη “κίνηση” και άλλων περιπτώσεων μεταξύ των συστάδων παρατηρούμε ότι (οι συστάδες) αρχίζουν να γίνονται πιο μεταβαλλόμενες. Παραδείγματος χάριν, στο παράδειγμα της τράπουλας, σταματάμε όταν καταλήξουμε με ένα σύνολο κόκκινων καρτών και ένα σύνολο μαύρων καρτών.

Επεξήγηση της διαδικασίας k-means:

Τα αποτελέσματα διαβάζονται με τον ίδιο περίπου τρόπο που διαβάζονται και τα αποτελέσματα από τον Ε.Μ Αλγόριθμο. Ο Ε.Μ (Expectation-Maximization / Προσδοκία-Μεγιστοποίηση) Αλγόριθμος είναι μέρος του πακέτου συσταδοποίησης – ομαδοποίησης του Weka. Ο Ε.Μ αλγόριθμος είναι ένα στατιστικό πρότυπο που χρησιμοποιεί το πεπερασμένο Γκαουσσισιανό (Gaussian) πρότυπο ανάμιξης.

Ο αλγόριθμος είναι παρόμοιος με τη διαδικασία k-means δεδομένου ότι ένα σύνολο παραμέτρων επαν-υπολογίζεται έως ότου επιτευχθεί μια επιθυμητή αξία σύγκλισης. Το πεπερασμένο πρότυπο ανάμιξης υποθέτει ότι όλες οι ιδιότητες / κριτήρια είναι ανεξάρτητες τυχαίες μεταβλητές.

Ένα μίγμα (μία ανάμιξη) είναι ένα σύνολο N κατανομών πιθανότητας όπου κάθε κατανομή αντιπροσωπεύει μια συστάδα. Σε μια μεμονωμένη περίπτωση ορίζεται μια πιθανότητα που έχει ένα ορισμένο σύνολο ιδιοτήτων / κριτηρίων εκτιμώντας ότι ήταν μέλος μιας συγκεκριμένης συστάδας. Στην απλούστερη περίπτωση $N=2$, οι κατανομές πιθανοτήτων υποτίθεται ότι είναι κανονικές και οι περιπτώσεις στοιχείων αποτελούνται από ένα μοναδικό, αντιπροσωπευτικό (real-valued) κριτήριο. Χρησιμοποιώντας το σενάριο, η εργασία του αλγορίθμου είναι να καθορίσει την αξία πέντε (5) παραμέτρων, συγκεκριμένα:

- 1 & 2. Τη μέση και σταθερή απόκλιση για τη συστάδα 1
- 3 & 4. Τη μέση και σταθερή απόκλιση για τη συστάδα 2
5. Την πιθανότητα δειγματοληψίας P για τη συστάδα 1 (η πιθανότητα για τη συστάδα 2 είναι $1-P$)

Παρακάτω παρουσιάζουμε και τη γενική διαδικασία:

1. Υποθέτουμε τις αρχικές τιμές για τις πέντε παραμέτρους.
2. Χρησιμοποιούμε την εξίσωση πυκνότητας πιθανότητας για μια κανονική κατανομή για να υπολογίσουμε την πιθανότητα της συστάδας για κάθε περίπτωση. Στην περίπτωση μιας ενιαίας ανεξάρτητης μεταβλητής με μέση τιμή μ και σταθερή απόκλιση σ , ο τύπος είναι:

$$f(x) = \frac{1}{(\sqrt{2\pi}\sigma)e^{\frac{-(x-\mu)^2}{2\sigma^2}}}$$

Στην περίπτωση δύο-συστάδων, θα έχουμε δύο τύπους κατανομής πιθανότητας που θα έχουν διαφορετική μέση τιμή και διαφορετική τιμή σταθερής απόκλισης.

3. Χρησιμοποιούμε τα αποτελέσματα των πιθανοτήτων για την επανεκτίμηση των πέντε παραμέτρων.
4. Επιστρέφουμε στο βήμα 2

Ο αλγόριθμος ολοκληρώνεται, όταν ο τύπος που “μετράει” την ποιότητα των συστάδων, δεν παρουσιάζει πλέον σημαντικές αυξήσεις. Ένα μέτρο της ποιότητας συστάδων είναι η πιθανότητα τα στοιχεία να προήλθαν από το σύνολο των δεδομένων που καθορίστηκαν από τη συσταδοποίηση. Ο υπολογισμός πιθανότητας είναι απλά ο πολλαπλασιασμός του αθροίσματος των πιθανοτήτων για κάθε μια από τις περιπτώσεις. Με δύο συστάδες **A** και **B** που περιέχουν περιπτώσεις (instances) $x_1, x_2, \dots, x_n$ όπου $P_A = P_B = 0.5$ ο υπολογισμός γίνεται:

$$[.5P(x_1|A) + .5P(x_1|B)][.5P(x_2|A) + .5P(x_2|B)] \dots [.5P(x_n|A) + .5P(x_n|B)]$$

Έστω λοιπόν ότι δημιουργήθηκαν 3 συστάδες, όπως φαίνονται παρακάτω:

0	102 (21%)
1	179 (38%)
2	196 (41%)

Ο πρώτος αριθμός είναι ο αριθμός της συστάδας στον οποίον αναφέρονται τα στατιστικά αποτελέσματα. Ο δεύτερος αριθμός είναι ο αριθμός των περιπτώσεων που περιέχονται σε εκείνη την συστάδα. Ο τρίτος αριθμός είναι το ποσοστό των περιπτώσεων που περιέχονται στην συστάδα.

4.3.4 Παρουσίαση και περιγραφή αποτελεσμάτων της έρευνάς μας, για τη συσταδοποίηση των δεδομένων:

Όπως αναφέραμε και προηγουμένως, προκειμένου να εξετάσουμε τη συσταδοποίηση / ομαδοποίηση του δείγματός μας, χρησιμοποιήσαμε τελικά τον αλγόριθμο k-means. Καθότι την ποσότητα των συστάδων (clusters) την καθορίζει ο χρήστης, αφού “τρέξαμε” τον αλγόριθμο για δύο (2), τρεις (3) και έως επτά (7) συστάδες (όσα και τα κριτήριά μας), επιλέξαμε να παρουσιάσουμε αναλυτικά τα αποτελέσματα που προέκυψαν από τη δημιουργία 2 συστάδων, ουσιαστικά ψάχνοντας να βρούμε τις ομαδοποιήσεις που δημιουργούνται για τα δύο κόμματα που συγκεντρώνουν την πλειοψηφία των απαντήσεών μας. Αυτό δεν σημαίνει ότι δεν μπορούμε να επεκτείνουμε την έρευνα μας δημιουργώντας περισσότερες συστάδες, προκειμένου να βρούμε και να ερμηνεύσουμε τις κατάλληλες πληροφορίες. Τα αποτελέσματα λοιπόν που προέκυψαν από την εφαρμογή του αλγορίθμου είναι:

==== Run information ====

Scheme: weka.clusterers.SimpleKMeans -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10

Relation: ekpaideftiko

Instances: 300

Attributes: 7

Programme

Vouleftes

Ideologia

Leader

Trust

Communication

komma

Test mode: evaluate on training data

==== Model and evaluation on training set ====

kMeans

=====

Number of iterations: 6

Within cluster sum of squared errors: 291.89230354252845

Missing values globally replaced with mean/mode

Cluster centroids:

		Cluster#	
Attribute	Full Data	0	1
	(300)	(134)	(166)

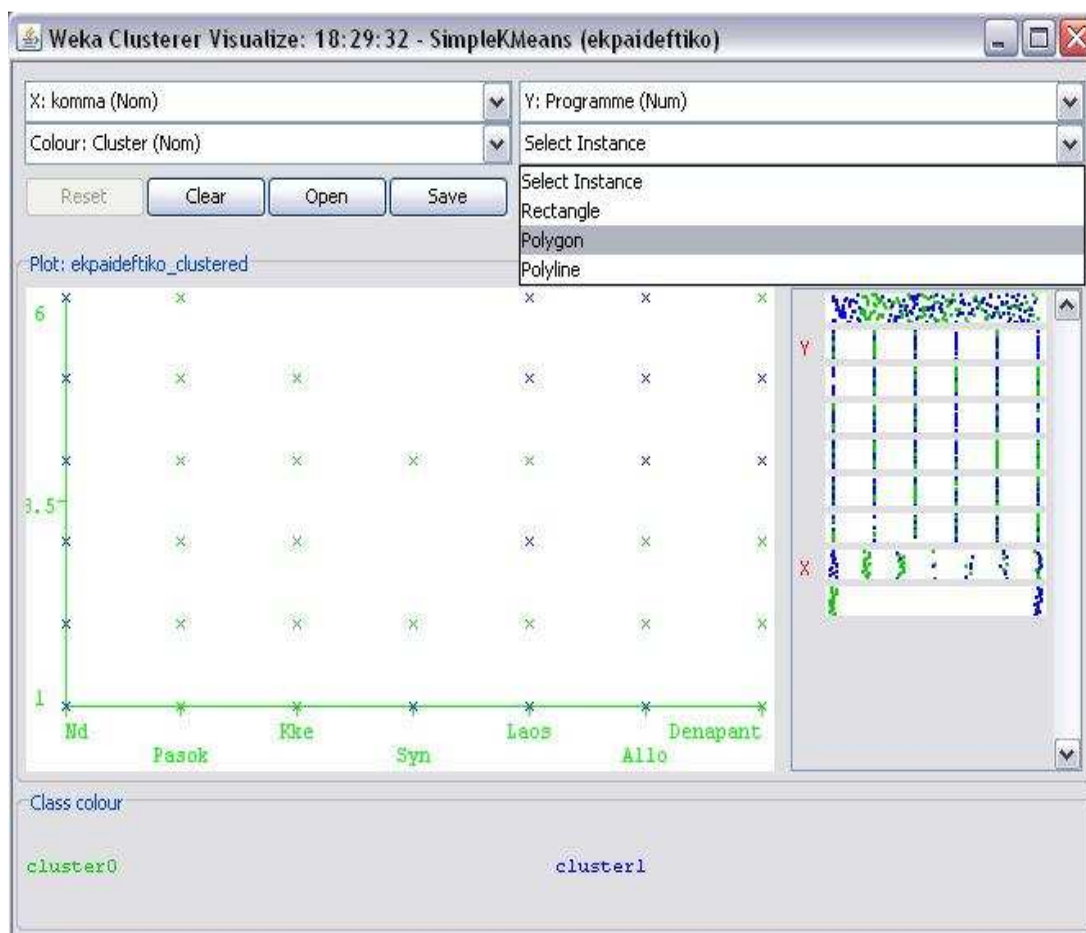
=====

Programme	3.16	2.6642	3.5602
Vouleftes	4.11	4.4328	3.8494
Ideologia	2.5	1.8806	3
Leader	2.9633	3.7313	2.3434
Trust	3.72	3.7836	3.6687
Communication	4.15	4.1791	4.1265
komma	Nd	Pasok	Nd

Clustered Instances

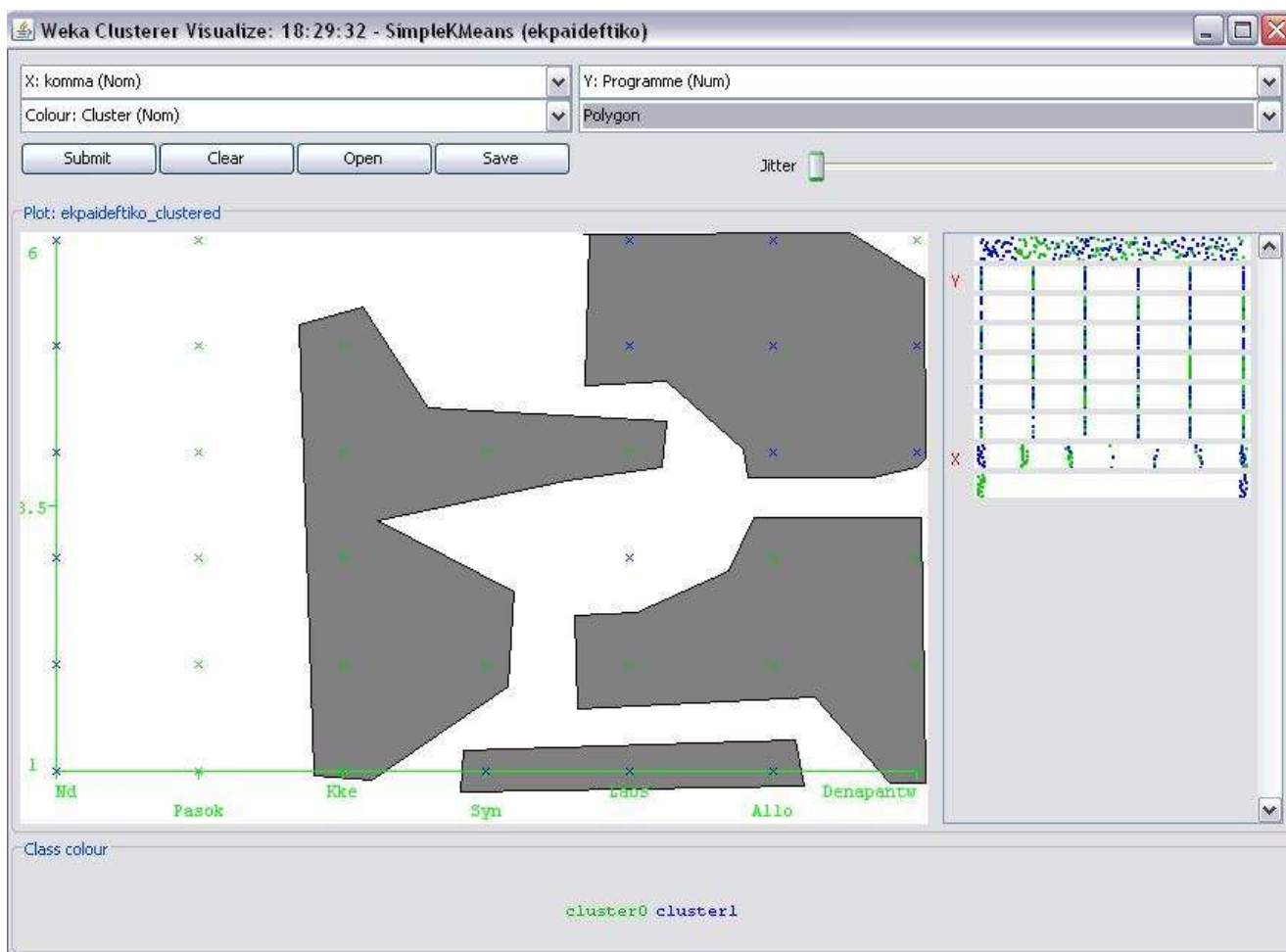
0	134 (45%)
1	166 (55%)

Παρατηρούμε λοιπόν ότι προκύπτουν δύο συστάδες / ομαδοποιήσεις, η **cluster 0**, όπου συγκεντρώνονται 134 από τις 300 περιπτώσεις, ή το 45%, του δείγματός μας και η **cluster 1**, όπου συγκεντρώνονται 166 από τις 300 περιπτώσεις, ή το 55%, του δείγματός μας. Η ομαδοποίηση **cluster 0** αναφέρεται στο ΠΑ.ΣΟ.Κ και η ομαδοποίηση **cluster 1** αναφέρεται στη Νέα Δημοκρατία. Στη στήλη “Result list” ανοίγουμε με δεξί κλικ στο αποτέλεσμα του “SimpleKMeans”, το αναδυόμενο μενού, και επιλέγουμε “Visualize cluster assignments”, όπου προκύπτει η ακόλουθη εικόνα / σχήμα (Σχήμα 4.3.4.1)



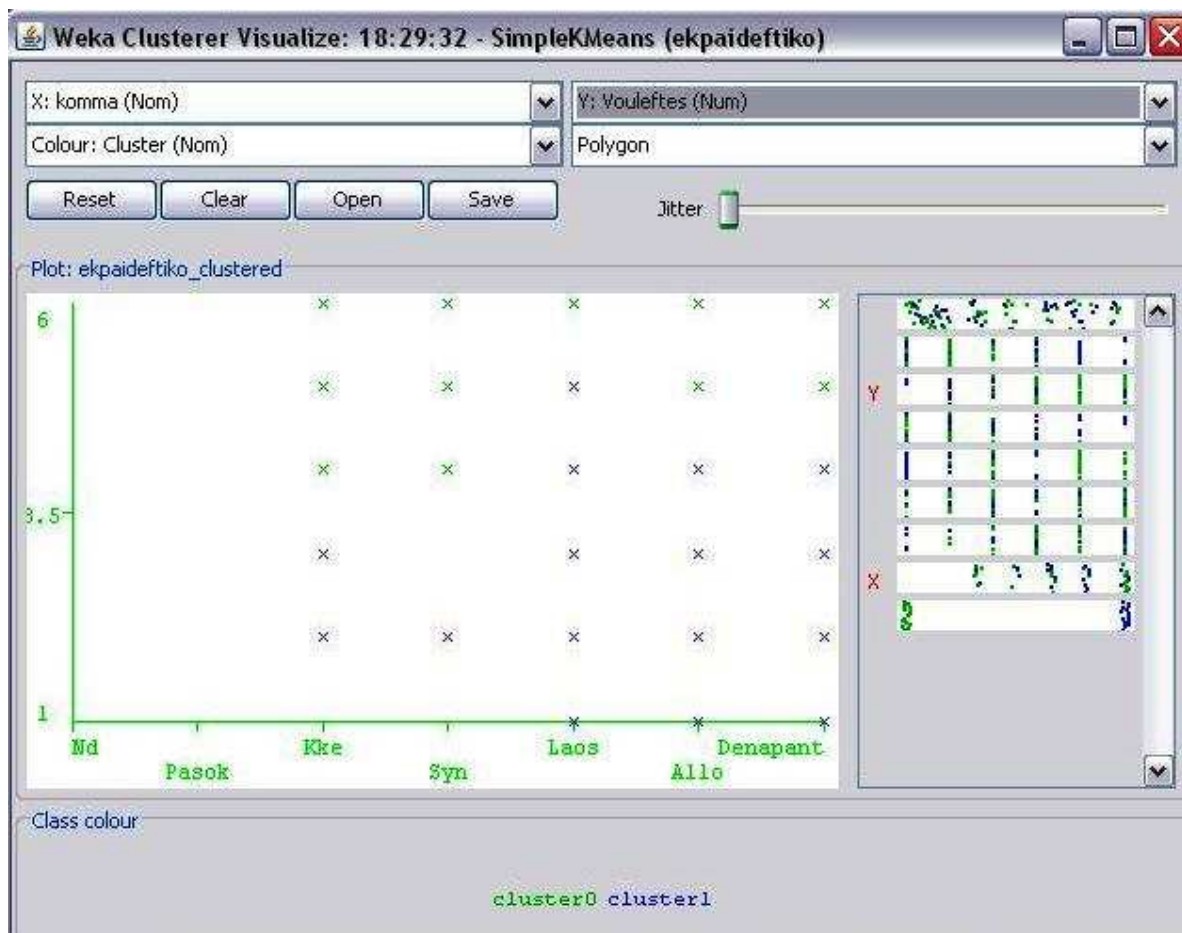
Σχήμα 4.3.4.1 Απεικόνιση ομαδοποιήσεων / συστάδων (clusters visualization)

Παρατηρούμε τις ομαδοποιήσεις που προκύπτουν, με δύο διαφορετικά χρώματα, **cluster 0** με το πράσινο χρώμα που αναφέρεται στο ΠΑΣΟΚ και **cluster 1** με το μπλε χρώμα που αναφέρεται στη Νέα Δημοκρατία (ΝΔ). Για να αναδείξουμε αυτές τις ομαδοποιήσεις και να “ανακαλύψουμε” τις προθέσεις των υπόλοιπων ψηφοφόρων / ερωτηθέντων σε σχέση με τη Νέα Δημοκρατία και το ΠΑΣΟΚ, από το μενού “Select Instance”, όπως φαίνεται και στο Σχήμα 4.3.4.1, κάνουμε κλικ στην επιλογή “Polygon” και κατά κάποιο τρόπο ζωγραφίζουμε τις ομαδοποιήσεις που δείχνουν να ξεχωρίζουν, για το κριτήριο “πρόγραμμα”. Έτσι λοιπόν, στο Σχήμα 4.3.4.2, αναδεικνύουμε την απεικόνιση αυτών των συστάδων / ομαδοποιήσεων που προέκυψαν από τον αλγόριθμο k-means. Θα πρέπει να τονίσουμε ότι δεν αναδεικνύουμε τις ομαδοποιήσεις που προκύπτουν σε ΝΔ και ΠΑΣΟΚ, καθώς αποτελούν ήδη τις δύο συστάδες μας (clusters) και άρα δεν προκύπτει κάποια “κρυμμένη” πληροφορία. Από την απεικόνιση, λοιπόν, με τη βοήθεια του “Polygon” εργαλείου προκύπτει ότι:



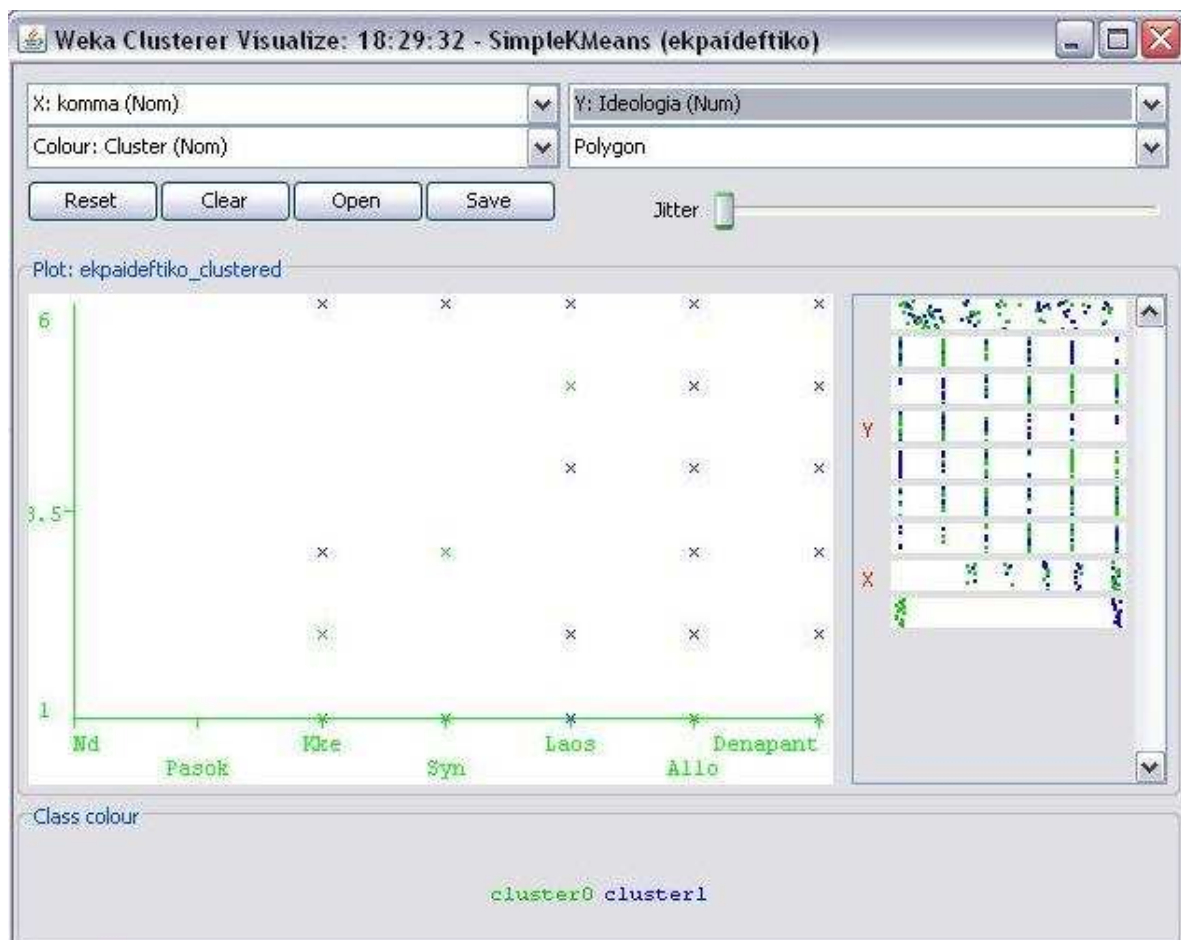
Σχήμα 4.3.4.2 Απεικόνιση ομαδοποιήσεων για το κριτήριο “Πρόγραμμα”

Με την επιλογή “Submit”, δηλώνουμε – καταχωρούμε τις συστάδες που “ζωγραφίσαμε” και με αυτόν τον τρόπο μπορούμε άμεσα, αλλάζοντας το κριτήριο στον Y άξονα, να δούμε τα αποτελέσματα για τα υπόλοιπα κριτήρια και τη συμπεριφορά των ψηφοφόρων / ερωτηθέντων του δείγματός μας σε σχέση με τις δύο μας συστάδες, εν προκειμένω σε σχέση με τη ΝΔ και το ΠΑΣΟΚ. Παρατηρούμε ότι στο κριτήριο “Πρόγραμμα”, όσοι δήλωσαν ότι θα ψηφίσουν ΚΚΕ ταυτίζονται με τη συστάδα 0 (cluster 0) , δηλαδή με το πρόγραμμα του ΠΑΣΟΚ. Οι ερωτηθέντες που μας δήλωσαν ότι θα ψηφίσουν ΣΥ.ΡΙΖ.Α, ΛΑΟΣ και ΑΛΛΟ (άλλο κόμμα, λευκό, άκυρο) βλέπουμε ότι, αν και βρίσκονται και στη συστάδα 0, εν τούτοις ως πρώτη προτίμηση, δείχνουν στο πρόγραμμα της ΝΔ, γι’ αυτό και ομαδοποιούνται στη συστάδα 1 (cluster 1), στην τιμή 1, αλλά και στις τιμές 5 και 6 του άξονα Y (άξονας Y: κλίμακα μέτρησης κριτηρίου). Όσοι από το δείγμα, δεν μας απάντησαν τι σκέφτονται να ψηφίσουν (ΔΕΝ ΑΠΑΝΤΩ), παρατηρούμε ότι στο κριτήριο “Πρόγραμμα” , βρίσκονται εγγύτερα με το ΠΑΣΟΚ παρά με τη ΝΔ, όπως φαίνεται και από το προαναφερθέν Σχήμα 4.3.4.2.



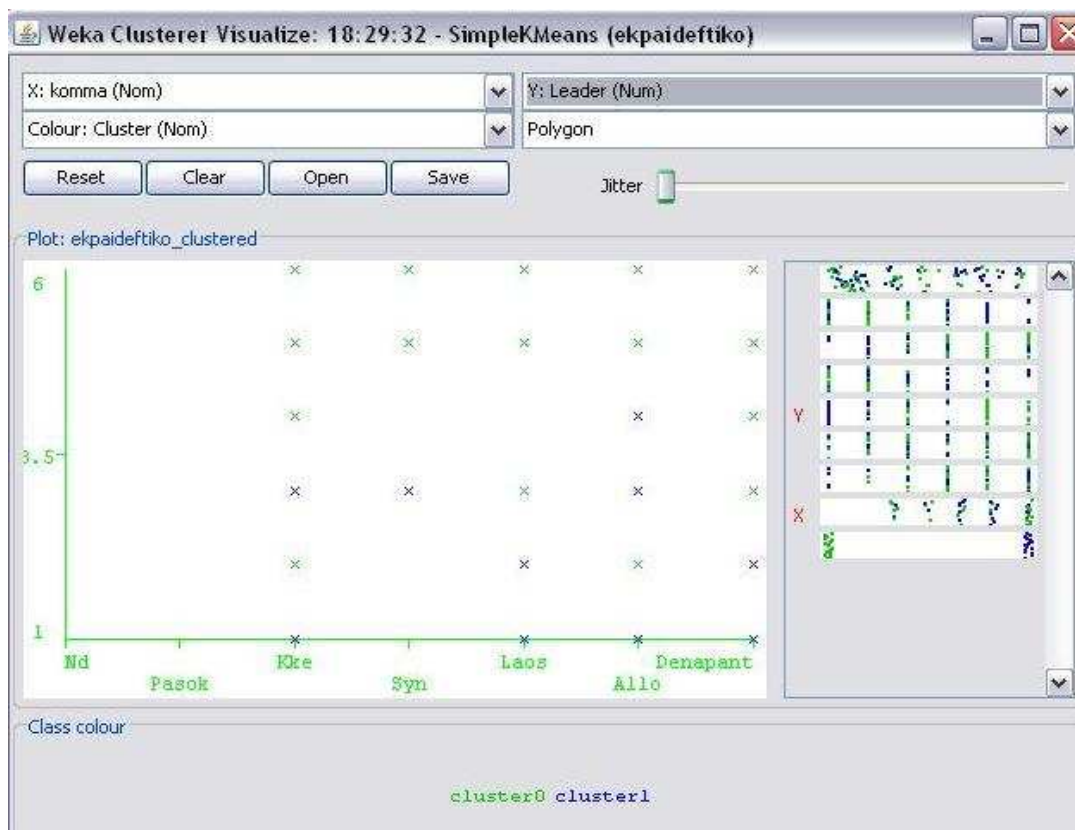
Σχήμα 4.3.4.3 Απεικόνιση ομαδοποιήσεων για το κριτήριο “Βουλευτική ομάδα”

Στο Σχήμα 4.3.4.3 φαίνονται τα αποτελέσματα της συσταδοποίησης για το κριτήριο “Βουλευτές” (“Βουλευτική ομάδα”), όπου και παρατηρούμε ότι οι ερωτηθέντες βρίσκονται εγγύτερα στη Νέα Δημοκρατία (ΝΔ) και δευτερευόντως ομαδοποιούνται στο ΠΑΣΟΚ. Αξίζει να σημειώσουμε ότι οι ψηφοφόροι του ΚΚΕ και του ΣΥ.ΡΙΖ.Α δεν ομαδοποιούνται ως πρώτη επιλογή, σε σχέση με το κριτήριο που αναφέραμε, σε κανένα από τα δύο κόμματα, εν αντιθέσει με όσους μας δήλωσαν ότι θα ψηφίσουν ΛΑΟΣ, ΑΛΛΟ ή δεν μας απάντησαν, όπου έχουν και ως πρώτη επιλογή τη Νέα Δημοκρατία, αναφερόμενοι πάντα στο κριτήριο “Βουλευτική ομάδα”. Συνεχίζοντας την παρουσίαση των αποτελεσμάτων μας, για το κριτήριο “Ιδεολογία”, όπως βλέπουμε και στο Σχήμα 4.3.4.4 που ακολουθεί, οι ψηφοφόροι του ΚΚΕ και του ΣΥ.ΡΙΖ.Α, αλλά και όσοι απάντησαν στην έρευνά μας ΑΛΛΟ και ΔΕΝ ΑΠΑΝΤΩ, προκύπτει ότι στη θέση 1 (τιμή 1 στο συγκεκριμένο κριτήριο) ομαδοποιούνται στο ΠΑΣΟΚ (cluster 0), ενώ όσοι μας απάντησαν ότι θα ψηφίσουν ΛΑΟΣ, ομαδοποιούνται και για τη θέση 1 στη Νέα Δημοκρατία (cluster 1). Εν τούτοις, οι ερωτηθέντες που μας δήλωσαν ΑΛΛΟ και ΔΕΝ ΑΠΑΝΤΩ, στις υπόλοιπες θέσεις-τιμές για το κριτήριο “Ιδεολογία” ομαδοποιούνται στη συστάδα 1, δηλαδή στη ΝΔ.

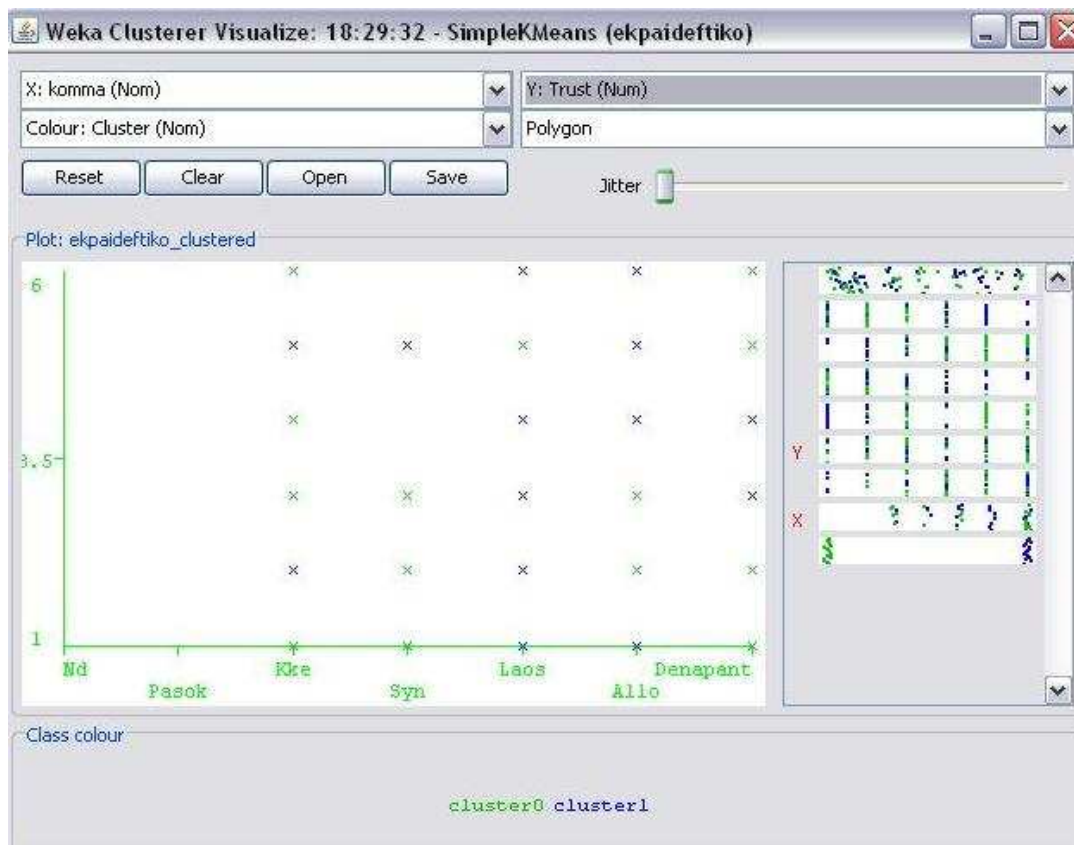


Σχήμα 4.3.4.4 Απεικόνιση ομαδοποιήσεων για το κριτήριο “Ιδεολογία”

Για το κριτήριο “Αρχηγός”, όπως απεικονίζεται και στο Σχήμα 4.3.4.5, ομαδοποιούνται στη ΝΔ, ως πρώτη επιλογή, όσοι από τους ερωτηθέντες μας δήλωσαν ότι θα ψηφίσουν ΚΚΕ, ΛΑΟΣ, ΑΛΛΟ και ΔΕΝ ΑΠΑΝΤΩ. Αξίζει να σημειώσουμε ότι οι ψηφοφόροι του ΚΚΕ και του ΑΛΛΟΥ, ως δεύτερη επιλογή, και οι ψηφοφόροι του ΛΑΟΣ και του ΔΕΝ ΑΠΑΝΤΩ, ως τρίτη επιλογή, ομαδοποιούνται στη συστάδα 0 (cluster 0), δηλαδή στο ΠΑΣΟΚ. Οι ψηφοφόροι του ΣΥ.ΡΙΖ.Α αν και δεν ομαδοποιούνται ούτε ως προς την πρώτη, ούτε ως προς τη δεύτερη επιλογή, εν τούτοις ομαδοποιούνται, ως προς την τρίτη επιλογή - όπως και όσοι μας δήλωσαν ότι θα ψηφίσουν ΚΚΕ και ΑΛΛΟ - στη ΝΔ. Στο Σχήμα 4.3.4.6 απεικονίζονται οι ομαδοποιήσεις σε σχέση με το κριτήριο “Εμπιστοσύνη” (“Εμπιστοσύνη” στο κόμμα, στις θέσεις που παρουσιάζει, στον αρχηγό κ.λπ.). Βλέπουμε ότι οι ψηφοφόροι του ΚΚΕ, του ΣΥ.ΡΙΖ.Α και του ΔΕΝ ΑΠΑΝΤΩ, ομαδοποιούνται, ως προς την πρώτη επιλογή, στο ΠΑΣΟΚ, ενώ οι ψηφοφόροι του ΛΑΟΣ και του ΑΛΛΟ στη ΝΔ. Θα πρέπει να αναδείξουμε ότι όσοι από τους ερωτηθέντες του δείγματός μας, δήλωσαν είτε ότι θα ψηφίσουν ΑΛΛΟ, ως δεύτερη και τρίτη επιλογή, ομαδοποιούνται στο ΠΑΣΟΚ και ως τέταρτη, πέμπτη και έκτη επιλογή του κριτηρίου “Εμπιστοσύνη”, ομαδοποιούνται στη ΝΔ. Αντίστοιχα όσοι από τους ερωτηθέντες του δείγματός μας, δήλωσαν (ότι) ΔΕΝ ΑΠΑΝΤΩ - ως δεύτερη, πέμπτη και έκτη επιλογή - ομαδοποιούνται στο ΠΑΣΟΚ, ενώ ως τρίτη και τέταρτη επιλογή, στη ΝΔ.

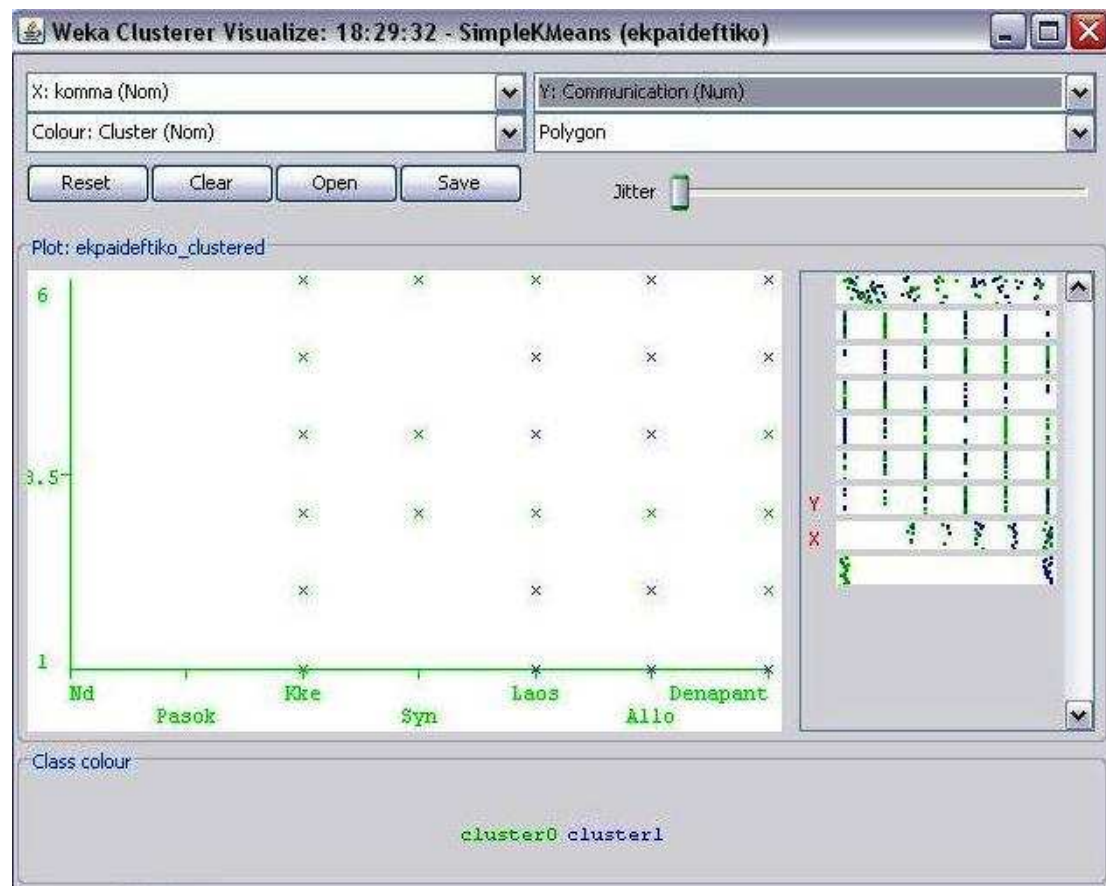


Σχήμα 4.3.4.5 Απεικόνιση ομαδοποιήσεων για το κριτήριο “ Αρχηγός ”



Σχήμα 4.3.4.6 Απεικόνιση ομαδοποιήσεων για το κριτήριο “ Εμπιστοσύνη ”

Τέλος, όσον αφορά το κριτήριο “Επικοινωνία” (επικοινωνιακή πολιτική), παρατηρούμε ότι ως τρίτη επιλογή, οι ψηφοφόροι όλων των κομμάτων ομαδοποιούνται στο ΠΑΣΟΚ. Ως δεύτερη επιλογή, όσοι μας δήλωσαν ΛΑΟΣ ή ΑΛΛΟ ομαδοποιούνται στη ΝΔ, ενώ ΚΚΕ και ΔΕΝ ΑΠΑΝΤΩ ομαδοποιούνται στο ΠΑΣΟΚ. Πάντως ως πρώτη επιλογή του κριτηρίου “Επικοινωνία”, οι ψηφοφόροι του ΛΑΟΣ, ΑΛΛΟ και ΔΕΝ ΑΠΑΝΤΩ ομαδοποιούνται στη ΝΔ, ενώ μόνο οι ψηφοφόροι του ΚΚΕ ομαδοποιούνται στο ΠΑΣΟΚ. Για τους ψηφοφόρους του ΣΥ.ΡΙΖ.Α, ο αλγόριθμος k-means, μας δείχνει ότι ως πρώτη, δεύτερη και πέμπτη επιλογή στο συγκεκριμένο κριτήριο, δεν ανήκουν σε καμία από τις δύο συστάδες που προαναφέραμε (Σχήμα 4.3.4.7).



Σχήμα 4.3.4.7 Απεικόνιση ομαδοποιήσεων για το κριτήριο “ Επικοινωνία ”

Όπως είχαμε αναφέρει και στην παράγραφο 4.3.1, επιλέξαμε τα 300 ερωτηματολόγια που θα μας χρησιμεύσουν στην εξαγωγή γνώσης και συμπερασμάτων ως το εκπαιδευτικό μας δείγμα και τα υπόλοιπα 100 σαν επαληθευτικό δείγμα προκειμένου να ελέγξουμε την όλη διαδικασία Εξόρυξης Γνώσης. Θα παρουσιάσουμε, λοιπόν, παρακάτω τα αποτελέσματα του αλγορίθμου k-means, για δύο (2) συστάδες (clusters), για τα 100 ερωτηματολόγια του επαληθευτικού μας δείγματος και θα συγκρίνουμε τα αποτελέσματά του με αυτά του εκπαιδευτικού μας δείγματος που παρουσιάσαμε στην αρχή της παραγράφου 4.3.4. Το αποτέλεσμα λοιπόν είναι το εξής:

==== Run information ====

Scheme: weka.clusterers.SimpleKMeans -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10

Relation: epalithefsi

Instances: 100

Attributes: 7

Programme

Vouleftes

Ideologia

Leader

Trust

Communication

komma

Test mode: evaluate on training data

==== Model and evaluation on training set ====

kMeans

=====

Number of iterations: 10

Within cluster sum of squared errors: 95.58084848484847

Missing values globally replaced with mean/mode

Cluster centroids:

Attribute	Cluster#		
	Full Data	0	1
	(100)	(55)	(45)
Programme	3.16	2.9091	3.4667
Vouleftes	4.06	3.6182	4.6
Ideologia	2.68	2.8545	2.4667
Leader	2.75	2.1273	3.5111
Trust	3.95	4.4909	3.2889
Communication	4.18	4.1455	4.2222
komma	Nd	Nd	Pasok

Clustered Instances

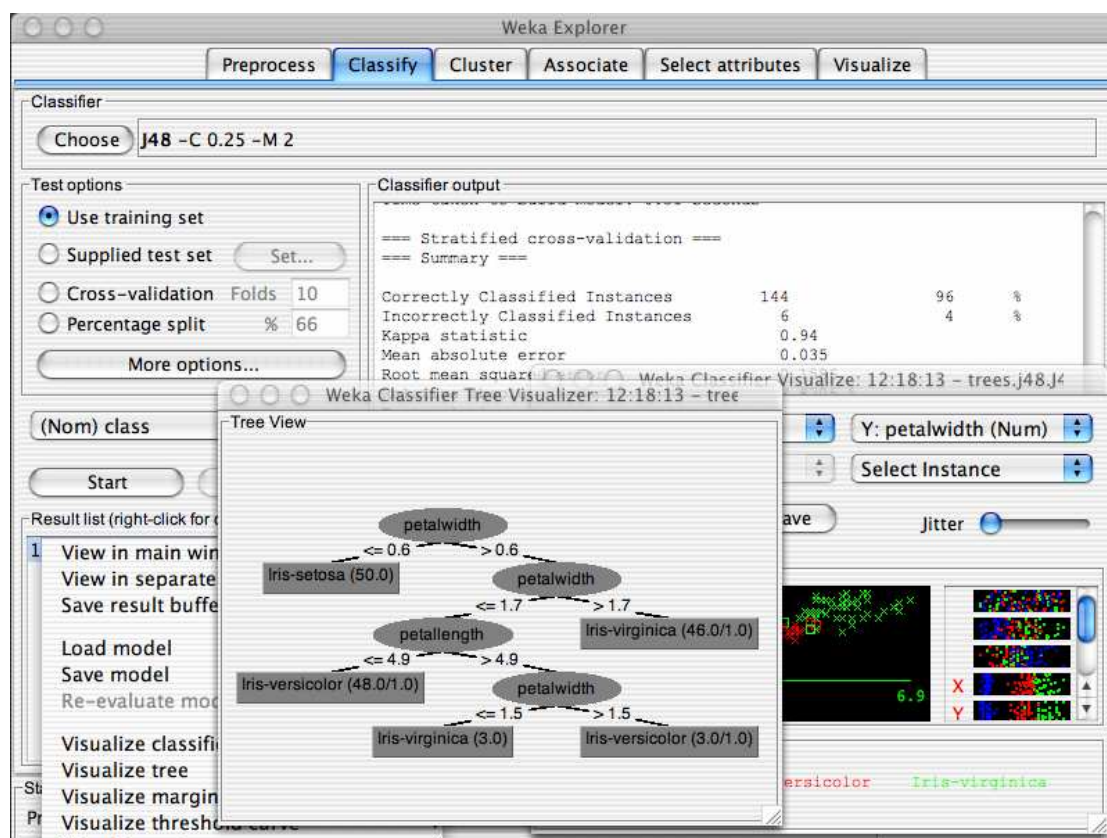
0 55 (55%)

1 45 (45%)

Το αποτέλεσμα λοιπόν που προκύπτει είναι πανομοιότυπο με το αποτέλεσμα του αλγορίθμου στο εκπαιδευτικό μας δείγμα. Παρατηρούμε ότι το 55% των περιπτώσεων ομαδοποιούνται στη ΝΔ και 45% των περιπτώσεων συσταδοποιούνται στο ΠΑΣΟΚ, όπως ακριβώς και στο εκπαιδευτικό μας δείγμα, με μόνη διαφορά την αντιστοίχιση της συστάδας. Στο επαληθευτικό μας δείγμα η συστάδα 0 (cluster 0) αναφέρεται στη ΝΔ, ενώ στο εκπαιδευτικό μας δείγμα στο ΠΑΣΟΚ και αντίστοιχα, η συστάδα 1(cluster 1) στο επαληθευτικό μας δείγμα αναφέρεται στο ΠΑΣΟΚ, ενώ στο εκπαιδευτικό μας δείγμα στη ΝΔ..

4.3.5 Κατηγοριοποίηση Δεδομένων / Ταξινομητές Απόφασης

Η κατηγοριοποίηση δεδομένων (classify) επιτρέπει τη διαμόρφωση και την εφαρμογή των διαφόρων μοντέλων κατηγοριοποίησης στα τρέχοντα δεδομένα. Επίσης, μπορεί να πραγματοποιήσει συγκριτικές μελέτες ή ελέγχους σε ομάδες δεδομένων. Μπορεί κανείς να προβάλλει τα σφάλματα της κάθε κατηγοριοποίησης με ένα αναδυόμενο μενού-εργαλείο, ενώ αν το αποτέλεσμα είναι ένα δένδρο απόφασης μπορεί να προβληθεί γραφικά, όπως φαίνεται και στο παράδειγμα του Σχήματος 4.3.5.1.



Σχήμα 4.3.5.1 Περιβάλλον εργασίας του Weka για την κατηγοριοποίηση δεδομένων (classify)

4.3.5.1 Δέντρα αποφάσεων εποπτευόμενης μάθησης

Τα δέντρα απόφασης αντιπροσωπεύουν μια εποπτευόμενη προσέγγιση στη ταξινόμηση. Ένα δέντρο απόφασης είναι μια απλή δομή, όπου οι μη τερματικοί κόμβοι αντιπροσωπεύουν τις δοκιμές σε ένα ή περισσότερα κριτήρια και οι τελικοί κόμβοι απεικονίζουν τις εκβάσεις (“εξόδους”) απόφασης. Ο J.R. Quinlan⁴, έχει διαδώσει τη προσέγγιση δέντρων απόφασης με την έρευνά του που εκτείνεται περισσότερο από 15 έτη. Η πιο πρόσφατη εφαρμογή του προτύπου του Quinlan είναι ο C4.5 αλγόριθμος. Το πακέτο ταξινομητών του Weka έχει τη δικιά του έκδοση του C4.5, γνωστή ως J48.

Αν και δεν θα παρουσιάσουμε αναλυτικά το συγκεκριμένο αλγόριθμο δέντρου απόφασης εδώ, μπορούμε να συνοψίσουμε τη γενική προσέγγισή του, στα εξής:

1. Επιλέγει ένα κριτήριο που διαφοροποιεί καλύτερα τις εξαγόμενες τιμές των κριτηρίων.

2. Δημιουργεί ξεχωριστά κλαδιά για κάθε τιμή των επιλεγμένων ιδιοτήτων (κριτηρίων).

3. Διαιρεί τις περιπτώσεις σε υποομάδες, ώστε να απεικονιστούν οι τιμές των κριτηρίων του επιλεγμένου κόμβου.

4. Για κάθε υποομάδα, ολοκληρώνει τη διαδικασία επιλογής κριτηρίων εάν:

- α) Όλα τα μέλη μιας υποομάδας έχουν την ίδια τιμή για το παραγόμενο (εξαγόμενο) κριτήριο, ολοκληρώνουν τη διαδικασία επιλογής κριτηρίων για την τρέχουσα πορεία και ονομάζουν τον κλάδο στην τρέχουσα πορεία με τη διευκρινισμένη αξία / τιμή.

- β) Η υποομάδα περιέχει έναν και μοναδικό κόμβο ή εάν κανένα περαιτέρω ιδιαίτερο κριτήριο δεν μπορεί να καθοριστεί. Όπως και στο (α), ονοματίζει τον κλάδο με την εξαγόμενη τιμή που προκύπτει από την πλειοψηφία των περιπτώσεων που απομένουν.

5. Για κάθε υποομάδα που δημιουργείται στο (3) και που δεν έχει χαρακτηριστεί ως τερματική (terminal), επαναλαμβάνει την ανωτέρω διαδικασία.

Ο αλγόριθμος εφαρμόζεται στα “στοιχεία εξάσκησης” (training set). Το δέντρο απόφασης εξετάζεται σε ένα σύνολο στοιχείων δοκιμής (test data), δεδομένου ότι είναι διαθέσιμα. Εάν τα στοιχεία δοκιμής δεν είναι διαθέσιμα, ο J48 αλγόριθμος εκτελεί μια διαγώνια-επικύρωση (cross-validation) χρησιμοποιώντας τα “στοιχεία εξάσκησης”.

4.3.5.2 Επεξήγηση προτύπου Δέντρου Απόφασης του Weka:

Ας δούμε ένα αντιπροσωπευτικό παράδειγμα δομής δέντρων απόφασης:

log_T90 <= 0.40824

| log_T90 <= 0.019947: short (161.0)

⁴ J.R. Quinlan. (1996). “*Improved Use of Continuous Attributes in C4.5*”, Journal of Artificial Intelligence Research, 4:77 - 90

| log_T90 > 0.019947

Κάθε γραμμή αντιπροσωπεύει έναν κόμβο στο δέντρο. Οι δύο επόμενες γραμμές, εκείνες που αρχίζουν με «|», είναι παράγωγοι κόμβοι (child nodes) της πρώτης γραμμής. Στη γενική περίπτωση, ένας κόμβος με έναν ή περισσότερους «|» χαρακτήρες πριν από τον κανόνα της σχέσης, είναι ένας παράγωγος κόμβος (child node) του κόμβου στον οποίο η γραμμή με τους «|» χαρακτήρες ολοκληρώνεται (έχει ένα «|» χαρακτήρα λιγότερο), εάν διαβάσουμε από κάτω προς τα επάνω τη σελίδα των αποτελεσμάτων του δέντρου απόφασης. Το επόμενο μέρος της γραμμής δηλώνει τον κανόνα. Εάν η έκφραση ισχύει για μια δεδομένη περίπτωση τότε, είτε ταξινομείται, εάν ο κανόνας ακολουθείται από μια άνω τελεία και έναν προσδιορισμό κατηγορίας--αυτός ο προσδιορισμός γίνεται η ταξινόμηση του κανόνα--ή, εάν δεν ακολουθείται από μια άνω τελεία, συνεχίζει στον επόμενο κόμβο στο δέντρο. Εάν η έκφραση είναι αντ' αυτού αναληθής (false), συνεχίζουμε στον «αδελφό» κόμβο, του κόμβου που μόλις αξιολογήσαμε, δηλαδή, στον κόμβο που έχει τον ίδιο αριθμό «|» χαρακτήρων πριν από αυτόν, αλλά και που έχει τον ίδιο κόμβο - γονέα. Στους κόμβους που παράγουν μια ταξινόμηση, όπως:

| log_T90 <= 0.019947: short (161.0)

αναγράφεται στο τέλος τους, ένας αριθμός (μερικές φορές δύο) σε παρένθεση. Ο πρώτος αριθμός δηλώνει πόσες περιπτώσεις, στο εκπαιδευτικό μας δείγμα (training set), ταξινομούνται σωστά από αυτόν τον κόμβο, όπου σε αυτήν την περίπτωση είναι 161. Ο δεύτερος αριθμός, εάν υπάρχει (εάν όχι, λαμβάνεται ότι είναι μηδέν [0]), αντιπροσωπεύει τον αριθμό των λανθασμένων ταξινομήσεων στον κόμβο.

Στην παρουσίαση της περιγραφής του δέντρου υπάρχουν / αναφέρονται και τα στοιχεία μέτρησης λάθους στο παραγόμενο δέντρο απόφασης, τα οποία και θα περιγράψουμε στην παρακάτω παράγραφο.

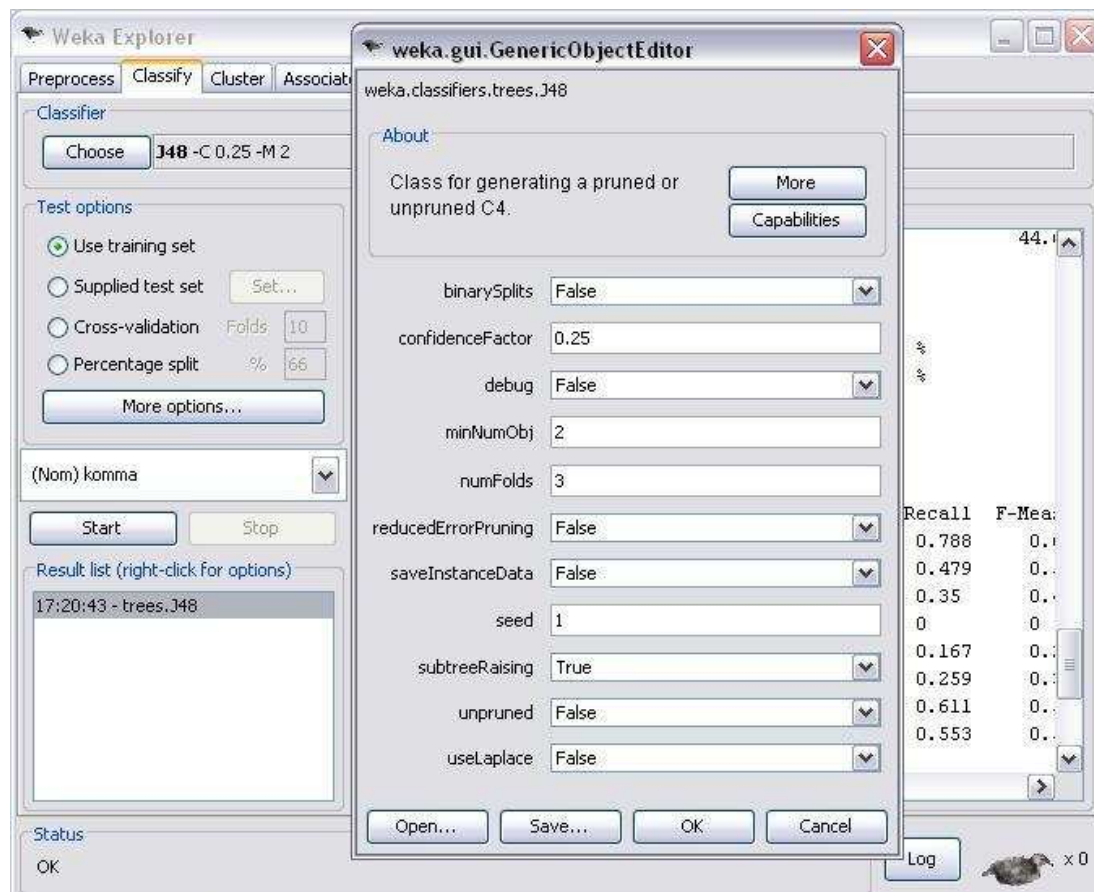
Ιδιότητες :

Πριν “τρέξουμε” τον αλγόριθμο J48, πρέπει να ελέγξουμε / καθορίσουμε τις ακόλουθες πέντε επιλογές:

1. Ποιο κριτήριο είναι το κριτήριο εξόδου (παραγόμενο).
2. Τον παράγοντα εμπιστοσύνης.
3. Τον ελάχιστο αριθμό περιπτώσεων ανά “Φύλλο”.
4. Τον αριθμό πτυχών (folds) για τη διαγώνια-επικύρωση (cross validation).
5. Το σύνολο στοιχείων δοκιμής.

Η πρώτη επιλογή μας επιτρέπει να επιλέξουμε, από έναν κατάλογο κριτηρίων, ποιο κριτήριο θέλουμε να είναι το κριτήριο εξόδου. Είμαστε σε θέση να επιλέξουμε μονάχα από τα κατηγορικά παραγόμενα κριτήρια, δεδομένου ότι ο J48 δεν επιτρέπει σε αριθμητικά κριτήρια, να είναι κριτήρια εξόδου.

Η δεύτερη επιλογή καθορίζει το διάστημα εμπιστοσύνης που χρησιμοποιείται για τη “περικοπή” του δέντρου (αφαιρεί τους κλάδους που κρίνει ότι παρέχουν ελάχιστο ή κανένα κέρδος στη στατιστική ακρίβεια του προτύπου). Συνήθως, επειδή η προκαθορισμένη τιμή, όπως φαίνεται και στο Σχήμα 4.3.5.2.1, του διαστήματος εμπιστοσύνης στο 25% (0,25) παράγει, στις περισσότερες περιπτώσεις, καλά αποτελέσματα, είναι πολύ πιθανόν να μην είναι απαραίτητη η τροποποίηση του εν λόγω στοιχείου. Εντούτοις, εάν το πραγματικό ποσοστό λάθους στα πραγματικά στοιχεία (ή το ποσοστό λάθους στη διαγώνια-επικύρωση) είναι σημαντικά υψηλότερο από το ποσοστό λάθους στο εκπαιδευτικό δείγμα, μπορούμε, αν θελήσουμε, να μειώσουμε τον παράγοντα εμπιστοσύνης για να προκαλέσουμε πιο δραστική περικοπή και ένα πιο γενικό πρότυπο για τα δεδομένα μας. Εάν θέλουμε μια πιο συγκεκριμένη διαμόρφωση, βασισμένη στο εκπαιδευτικό δείγμα, μπορούμε να αυξήσουμε τον παράγοντα εμπιστοσύνης (confidence factor), ο οποίος θα μειώσει το ποσό περικοπής που εμφανίζεται.



Σχήμα 4.3.5.2.1 Ιδιότητες και επιλογές του J48 αλγόριθμου ταξινομητή απόφασης

Η τρίτη επιλογή καθορίζει τον ελάχιστο αριθμό περιπτώσεων που πρέπει να παρουσιαστούν στο εκπαιδευτικό δείγμα, ώστε ένα νέο φύλλο να δημιουργηθεί στο δέντρο απόφασης για να χειριστεί τις συγκεκριμένες περιπτώσεις. Αυτό μπορεί επίσης να δημιουργήσει ένα γενικευμένο ή εξειδικευμένο δέντρο. Ένας υψηλότερος

αριθμός θα δημιουργήσει ένα γενικευμένο δέντρο και ένας χαμηλότερος αριθμός θα δημιουργήσει ένα πιο εξειδικευμένο δέντρο.

Η τέταρτη επιλογή καθορίζει πώς να κατασκευάσει και να εξετάσει έπειτα το πρότυπο, ελλείψει του στοιχείου δοκιμής. Εάν ο αριθμός πτυχών για τη διαγώνια-επικύρωση είναι x , τότε $(x-1) / x$ του εκπαιδευτικού δείγματος χρησιμοποιείται για να κατασκευάσει το πρότυπο και $1/x$ του εκπαιδευτικού δείγματος χρησιμοποιείται για να εξετάσει το πρότυπο. Αυτή η διαδικασία έπειτα επαναλαμβάνεται x φορές, έτσι ώστε όλο εκπαιδευτικό δείγμα να χρησιμοποιηθεί μόνο μια φορά στα στοιχεία δοκιμής. Οι x διαφορετικές εκτιμήσεις λάθους υπολογίζονται κατά μέσο όρο έπειτα για να παραγάγουν μια γενική εκτίμηση λάθους--οι στρωματοποιημένες στατιστικές διαγώνιας-επικύρωσης.

Αν και οι εκτενείς δοκιμές στα πολυάριθμα σύνολα δεδομένων έχουν δείξει ότι δέκα διαγώνιες-επικυρώσεις είναι από τις καλύτερες τιμές που μπορούμε να δώσουμε, προκειμένου να πάρουμε την ακριβέστερη εκτίμηση λάθους, εντούτοις και άλλες τιμές μπορούν να χρησιμοποιηθούν. Η μείωση του αριθμού πτυχών σε σχέση με την προεπιλογή της τιμής 10, θα μειώσει πιθανώς το χρονικό διάστημα που χρειάζεται το δέντρο απόφασης για να παραχθεί, και η αύξηση του αριθμού πτυχών θα αυξήσει πιθανώς το χρονικό διάστημα που χρειάζεται. Φυσικά, η αύξηση του αριθμού πτυχών (folds) θα δημιουργήσει ένα μεγαλύτερο σύνολο δεδομένων για τα δεδομένα μάθησης, τα οποία μπορούν να αυξήσουν την ακρίβεια του δέντρου απόφασης. Ομοίως, η μείωση του αριθμού πτυχών θα δημιουργήσει ένα μικρότερο σύνολο δεδομένων για τα δεδομένα μάθησης, τα οποία μπορούν να μειώσουν την ακρίβεια του δέντρου απόφασης.

Η πέμπτη επιλογή μας επιτρέπει να εξετάσουμε τα δεδομένα σε σχέση με ένα σύνολο δεδομένων δοκιμής, σε αντίθεση με τη δοκιμή του προτύπου μέσω της διαγώνιας-επικύρωσης, όπως την περιγράψαμε στην προηγούμενη παράγραφο.

4.3.5.3 Μέτρηση Σφαλμάτων, παραγόμενα σφάλματα (error output):

Στη συνέχεια, παρουσιάζουμε ένα χαρακτηριστικό παράδειγμα αποτύπωσης στατιστικών λαθών, όπως μας τα δίνει το Weka, των δεδομένων με αριθμητικά παραγόμενα κριτήρια. Τα πρότυπα που έχουν κατηγορικά παραγόμενα κριτήρια, έχουν έναν στατιστικό συντελεστή συσχετισμού αντί του στατιστικού όρου kappa (Kappa statistic) - αυτή είναι και η μόνη διαφορά μεταξύ των αριθμητικών και κατηγορικών παραγόμενων κριτηρίων.

=== Error on training data ===

Correctly Classified Instances	442	92.6625 %
Incorrectly Classified Instances	35	7.3375 %
Kappa statistic	0.8482	
Mean absolute error	0.0617	
Root mean squared error	0.1877	
Relative absolute error	18.8108 %	

Root relative squared error 46.4073 %
 Total Number of Instances 477

==== Confusion Matrix ====

```
a   b   c  <-- classified as
49  14   7 | a = inter
 7  315  0 | b = long
 7   0  78 | c = short
```

Στα στοιχεία μέτρησης των σφαλμάτων υπάρχουν τρία εύκολα εξηγήσιμα στοιχεία: οι σωστά ταξινομημένες περιπτώσεις (Correctly Classified Instances), οι λανθασμένα ταξινομημένες περιπτώσεις (Incorrectly Classified Instances) και ο συνολικός αριθμός των περιπτώσεων (Total Number of Instances). Τα τρία παραπάνω στοιχεία δίνουν, αντίστοιχα, τον αριθμό των περιπτώσεων που το πρότυπο διαμόρφωσε / μοντελοποίησε σωστά, διαμόρφωσε / μοντελοποίησε λανθασμένα, και το σύνολο τους. Για τα υπόλοιπα στοιχεία, λεπτομερέστερα, έχουμε:

- **Συντελεστής συσχετισμού (Correlation coefficient):**

$$\frac{S_{PA}}{S_P S_A} \quad S_{PA} = \frac{\sum_i (p_i - \bar{p})(a_i - \bar{a})}{n-1} \quad S_P = \frac{\sum_i (p_i - \bar{p})^2}{n-1} \quad S_A = \frac{\sum_i (a_i - \bar{a})^2}{n-1}$$

όπου , και

Ο συντελεστής συσχετισμού μετρά το στατιστικό συσχετισμό μεταξύ των προβλεφθεισών και των πραγματικών τιμών. Αυτή η μέθοδος είναι μοναδική δεδομένου ότι δεν αλλάζει με μια κλίμακα τιμών για τις περιπτώσεις δοκιμής. Επίσης, θα πρέπει να σημειώσουμε ότι ένας υψηλότερος αριθμός συντελεστή συσχετισμού σημαίνει και ένα καλύτερο πρότυπο, με τη μονάδα (1) να σημαίνει τον τέλειο στατιστικό συσχετισμό και το μηδέν (0) να σημαίνει ότι δεν υπάρχει κανένας συσχετισμός.

- **Kappa Statistic:**

$$\frac{P(A) - P(E)}{1 - P(E)}$$

, όπου **P(A)** είναι το ποσοστό του πόσες φορές οι τιμές του προτύπου ήταν ίσες με την πραγματική αξία και όπου **P(E)** είναι η αναμενόμενη τυχαία αναλογία. Εκτιμώντας ότι ο συντελεστής συσχετισμού χρησιμοποιείται για τα αριθμητικά στοιχεία, η Κάπα στατιστική (kappa statistic), χρησιμοποιείται ως μέσο ταξινόμησης συμφωνίας στα κατηγορικά δεδομένα. Ένας κάπα συντελεστής (kappa coefficient) ίσος με τη μονάδα (1) σημαίνει μια στατιστικά τέλεια διαμόρφωση, ενώ ίσος με το μηδέν (0) σημαίνει ότι κάθε τιμή του μοντέλου που παρήχθη ήταν διαφορετική από την πραγματική τιμή. Μια Κάπα στατιστική (kappa statistic) που

έχει τιμή **0.7** ή υψηλότερη θεωρείται γενικά ως ένας καλός συσχετισμός στατιστικής, και φυσικά, όσο υψηλότερη τιμή, τόσο καλύτερος ο συσχετισμός.

- **Ρίζα μέσου τετραγωνικού σφάλματος (Root mean-squared error):**

$$\sqrt{\frac{(a_1 - c_1)^2 + (a_2 - c_2)^2 + \dots + (a_n - c_n)^2}{n}}$$

Το μέσο τετραγωνικό σφάλμα είναι ένα από τα συνηθέστερα χρησιμοποιούμενα στοιχεία μέτρησης της αποτελεσματικότητας για την αριθμητική πρόβλεψη. Αυτή η τιμή υπολογίζεται με τη λήψη του μέσου όρου των τετραγωνικών διαφορών μεταξύ κάθε υπολογισμένης τιμής (c_i) και της αντίστοιχης σωστής τιμής (a_i). Η ρίζα του μέσου τετραγωνικού σφάλματος είναι απλά η τετραγωνική ρίζα του μέσου-τετραγωνικού-σφάλματος. Η ρίζα του μέσου τετραγωνικού σφάλματος δίνει στην τιμή σφάλματος την ίδια διαστατικότητα (dimensionality) με τις πραγματικές και προβλεφθείσες τιμές.

- **Μέσο Απόλυτο Σφάλμα (Mean absolute error):**

$$\frac{|a_1 - c_1| + |a_2 - c_2| + \dots + |a_n - c_n|}{n}$$

Το μέσο απόλυτο σφάλμα είναι ο μέσος όρος της διαφοράς μεταξύ της προβλεφθείσας και της πραγματικής τιμής σε όλες τις περιπτώσεις δοκιμής. Είναι το μέσο σφάλμα πρόβλεψης.

- **Ρίζα σχετικού τετραγωνικού σφάλματος (Root relative squared error):**

$$\sqrt{\frac{(a_1 - c_1)^2 + (a_2 - c_2)^2 + \dots + (a_n - c_n)^2}{(a_1 - \bar{a})^2 + (a_2 - \bar{a})^2 + \dots + (a_n - \bar{a})^2}}$$

Το σχετικό τετραγωνικό σφάλμα είναι το συνολικό τετραγωνικό σφάλμα που σχετίζεται με αυτό που το σφάλμα θα ήταν εάν η πρόβλεψη ήταν ο μέσος όρος της απόλυτης τιμής. Όπως γίνεται με τη ρίζα του μέσου τετραγωνικού σφάλματος, η τετραγωνική ρίζα του σχετικού τετραγωνικού σφάλματος λαμβάνεται για να δώσει τις ίδιες διαστάσεις με τις προβλεφθείσες τιμές. Επίσης, ακριβώς όπως με τη ρίζα του μέσου τετραγωνικού σφάλματος, υπερτονίζει τις περιπτώσεις στις οποίες το σφάλμα πρόβλεψης ήταν σημαντικά μεγαλύτερο από το μέσο σφάλμα.

- **Σχετικό Απόλυτο Σφάλμα (Relative absolute error):**

$$\frac{|a_1 - c_1| + |a_2 - c_2| + \dots + |a_n - c_n|}{|a_1 - \bar{a}| + |a_2 - \bar{a}| + \dots + |a_n - \bar{a}|}$$

Το σχετικό απόλυτο σφάλμα είναι το συνολικό απόλυτο σφάλμα, που σχετίζεται με αυτό που το σφάλμα θα ήταν, εάν η πρόβλεψη απλά ήταν ο μέσος όρος των πραγματικών τιμών. Στις τέσσερις προηγούμενες εξισώσεις, c_i είναι η αριθμητική τιμή πρόβλεψης στην i περίπτωση δοκιμής(test case), a_i είναι η πραγματική τιμή στην i περίπτωση δοκιμής, $\bar{a}$ είναι ο μέσος όρος των πραγματικών τιμών όλων των περιπτώσεων δοκιμής, και n είναι ο συνολικός αριθμός των περιπτώσεων δοκιμής. Σε όλες τις παραπάνω μετρήσεις σφάλματος, χαμηλότερη τιμή σημαίνει ακριβέστερο πρότυπο / μοντέλο, κι όπου η τιμή μηδέν (0) θα απεικόνιζε το στατιστικά τέλειο πρότυπο.

▪ Μήτρα “σύγχυσης” (Confusion matrix):

Η μήτρα “σύγχυσης” είναι απλά μια τετραγωνική μήτρα που μας παρουσιάζει τις διάφορες ταξινομήσεις και τις λάθος ταξινομήσεις του προτύπου σε μια συμπαγή περιοχή. Οι στήλες της μήτρας αντιστοιχούν στον αριθμό περιπτώσεων που ταξινομούνται με μία συγκεκριμένη τιμή και οι σειρές αντιστοιχούν στον αριθμό περιπτώσεων της πραγματικής ταξινόμησης. Παραδείγματος χάριν,

=== Confusion Matrix ===

```
a   b   c <-- classified as
49  14   7 | a = inter
 7  315  0 | b = long
 7   0  78 | c = short
```

βλέπουμε στον παραπάνω πίνακα ότι στο εκπαιδευτικό δείγμα οι 49 περιπτώσεις ήταν σωστά ταξινομημένες ως “inter” ενώ 14 “inter” περιπτώσεις ήταν λανθασμένα ταξινομημένες ως “long” και 7 “inter” περιπτώσεις ήταν λανθασμένα ταξινομημένες ως “short”.

4.3.6 Παρουσίαση και περιγραφή αποτελεσμάτων της έρευνάς μας, για την κατηγοριοποίηση - ταξινόμηση των δεδομένων:

Θα παρουσιάσουμε λοιπόν, τα αποτελέσματα που προέκυψαν από την εφαρμογή του J48 αλγόριθμου ταξινόμησης του υπολογιστικού συστήματος Weka, στο εκπαιδευτικό μας δείγμα, δηλαδή στους 300 ερωτηθέντες από το σύνολο των 400 του δείγματός μας, έχοντας ως επαληθευτικό δείγμα τους 100 εναπομείναντες ερωτηθέντες. Από την εφαρμογή λοιπόν του αλγορίθμου προέκυψαν τα εξής:

==== Run information ====

Scheme: weka.classifiers.trees.J48 -C 0.5 -M 2

Relation: ekpaideftiko

Instances: 300

Attributes: 7

Programme

Vouleftes

Ideologia

Leader

Trust

Communication

komma

Test mode: evaluate on training data

==== Classifier model (full training set) ====

J48 pruned tree

Leader <= 2

| Communication <= 5

| | Leader <= 1

| | | Communication <= 4

| | | | Ideologia <= 1: Pasok (7.0/2.0)

| | | | Ideologia > 1

| | | | | Communication <= 2: Laos (3.0)

| | | | | Communication > 2

| | | | | | Programme <= 1: Denapantw (3.0/1.0)

| | | | | | Programme > 1

| | | | | | Communication <= 3: Nd (6.0)

| | | | | | Communication > 3

| | | | | | | Programme <= 2: Nd (13.0/4.0)

| | | | | | | Programme > 2

| | | | | | | | Vouleftes <= 2

| | | | | | | | Ideologia <= 4: Nd (5.0/2.0)

| | | | | | | | | Ideologia > 4: Pasok (2.0/1.0)
 | | | | | | | | | Vouleftes > 2
 | | | | | | | | | Trust <= 5: Nd (17.0/5.0)
 | | | | | | | | | Trust > 5: Laos (8.0/5.0)
 | | | Communication > 4
 | | | | Programme <= 2: Nd (2.0)
 | | | | Programme > 2: Kke (3.0/2.0)
 | | Leader > 1
 | | | Ideologia <= 1
 | | | | Communication <= 4
 | | | | Trust <= 4: Nd (14.0/6.0)
 | | | | Trust > 4: Pasok (19.0/10.0)
 | | | | Communication > 4: Pasok (3.0)
 | | | Ideologia > 1
 | | | | Programme <= 4
 | | | | Communication <= 3: Allo (2.0/1.0)
 | | | | Communication > 3
 | | | | | Communication <= 4
 | | | | | Programme <= 3: Nd (11.0/3.0)
 | | | | | Programme > 3: Allo (3.0)
 | | | | | Communication > 4: Nd (3.0/1.0)
 | | | | Programme > 4: Denapantw (4.0/2.0)
 | Communication > 5
 | | Programme <= 4
 | | | Trust <= 4: Nd (7.0/2.0)
 | | | Trust > 4
 | | | | Programme <= 3: Denapantw (4.0/1.0)
 | | | | Programme > 3: Nd (3.0/2.0)
 | | Programme > 4: Denapantw (5.0/2.0)
 Leader > 2
 | Ideologia <= 3
 | | Ideologia <= 2
 | | | Trust <= 1
 | | | | Vouleftes <= 5
 | | | | Leader <= 3: Pasok (3.0/1.0)
 | | | | Leader > 3: Nd (2.0)
 | | | | Vouleftes > 5: Kke (2.0)
 | | | Trust > 1
 | | | | Communication <= 3
 | | | | Trust <= 3: Kke (4.0/2.0)
 | | | | Trust > 3: Nd (4.0/2.0)
 | | | | Communication > 3

```

| | | | | Leader <= 3
| | | | | | Communication <= 4
| | | | | | | Ideologia <= 1
| | | | | | | | Trust <= 3: Nd (5.0/3.0)
| | | | | | | | Trust > 3: Denapantw (11.0/5.0)
| | | | | | | | Ideologia > 1: Nd (7.0/4.0)
| | | | | | | | Communication > 4: Denapantw (11.0/8.0)
| | | | | | Leader > 3
| | | | | | | Ideologia <= 1
| | | | | | | | Leader <= 4: Kke (3.0/1.0)
| | | | | | | | Leader > 4: Pasok (33.0/21.0)
| | | | | | | | Ideologia > 1
| | | | | | | | Vouleftes <= 5: Denapantw (12.0/5.0)
| | | | | | | | Vouleftes > 5: Pasok (7.0/2.0)
| | | | | | Ideologia > 2
| | | | | | | Vouleftes <= 4
| | | | | | | | Leader <= 4: Denapantw (3.0/1.0)
| | | | | | | | Leader > 4: Pasok (6.0/2.0)
| | | | | | | | Vouleftes > 4
| | | | | | | | Vouleftes <= 5: Allo (4.0/1.0)
| | | | | | | | Vouleftes > 5: Denapantw (3.0/1.0)
| | | | | | Ideologia > 3
| | | | | | | Programme <= 5
| | | | | | | | Leader <= 3
| | | | | | | | | Ideologia <= 5: Nd (4.0/2.0)
| | | | | | | | | Ideologia > 5
| | | | | | | | | Programme <= 3: Kke (2.0/1.0)
| | | | | | | | | Programme > 3: Pasok (2.0/1.0)
| | | | | | | | | Leader > 3
| | | | | | | | | | Ideologia <= 5
| | | | | | | | | | Ideologia <= 4: Nd (8.0/2.0)
| | | | | | | | | | Ideologia > 4: Pasok (6.0/3.0)
| | | | | | | | | | Ideologia > 5
| | | | | | | | | | Leader <= 4: Nd (2.0)
| | | | | | | | | | Leader > 4: Pasok (2.0/1.0)
| | | | | | | | | Programme > 5
| | | | | | | | | Leader <= 3: Nd (3.0)
| | | | | | | | | Leader > 3: Allo (4.0/2.0)

```

Number of Leaves : 49

Size of the tree : 97

Time taken to build model: 0.06 seconds

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	177	59	%
Incorrectly Classified Instances	123	41	%
Kappa statistic	0.4549		
Mean absolute error	0.1504		
Root mean squared error	0.2742		
Relative absolute error	67.7668 %		
Root relative squared error	82.4179 %		
Total Number of Instances	300		

=== Detailed Accuracy by Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.75	0.194	0.672	0.75	0.709	0.865	Nd
	0.648	0.192	0.511	0.648	0.571	0.831	Pasok
	0.4	0.021	0.571	0.4	0.471	0.923	Kke
	0	0	0	0	0	0.954	Syn
	0.333	0.018	0.545	0.333	0.414	0.924	Laos
	0.333	0.015	0.692	0.333	0.45	0.911	Allo
	0.556	0.106	0.536	0.556	0.545	0.873	Denapantw
Weighted Avg.	0.59	0.136	0.584	0.59	0.576	0.872	

=== Confusion Matrix ===

```

a  b  c  d  e  f  g  <-- classified as
78 17  1  0  1  0  7 | a = Nd
16 46  0  0  1  0  8 | b = Pasok
 3  5  8  0  0  1  3 | c = Kke
 0  2  2  0  0  0  2 | d = Syn
 4  3  2  0  6  0  3 | e = Laos
 8  5  1  0  1  9  3 | f = Allo
 7 12  0  0  2  3 30 | g = Denapantw

```

Σύμφωνα με τα όσα αναφέραμε παραπάνω, παρατηρούμε ότι το 59% του εκπαιδευτικού μας δείγματος είναι σωστά ταξινομημένο, ενώ το 41% του δείγματός μας είναι λανθασμένα ταξινομημένο, με τον πίνακα των τιμών των σφαλμάτων να ακολουθεί. Στη συνέχεια, θέλοντας να ερμηνεύσουμε τον πίνακα: “Λεπτομερής Ακρίβεια ανά Κατηγορία” (Detailed Accuracy by Class) πρέπει να αναφέρουμε τα εξής:

- **TP Rate** = True Positive Rate: παραδείγματα στη θετική κατηγορία που προβλέφθηκαν ως θετικά. Η πρώτη τιμή του TP Rate, για παράδειγμα, προκύπτει ως εξής (χρησιμοποιώ τις τιμές από την πρώτη σειρά του Confusion Matrix πίνακα) : $TP Rate = 78 / (78+17+1+0+1+0+7) = 0,75$. Η δεύτερη τιμή προκύπτει (χρησιμοποιώ τις τιμές από τη δεύτερη σειρά του Confusion Matrix πίνακα): $46 / (16+46+0+0+1+0+8) = 0.648$. Ανάλογα προκύπτουν και οι υπόλοιπες τιμές της στήλης του TP Rate.
- **FP Rate** = False Positive Rate: παραδείγματα που προβλέφθηκαν ως θετικά, αλλά στην πραγματικότητα είναι αρνητικά.
- **Precision**: Είναι η πιθανότητα μια ανακτημένη πληροφορία να είναι και σχετική.
- **Recall**: Είναι η πιθανότητα μια σχετική πληροφορία να ανακτάται κατά την αναζήτηση.
- **F-measure**: Ο αρμονικός μέσος της ακρίβειας (precision) και της ανάκλησης (recall): $F = 2 * Recall * Precision / (Recall + Precision) = 2 * 0.75 * 0.672 / (0.75 + 0.672) = 0,709$, που είναι και η πρώτη τιμή στον πίνακα “Detailed Accuracy by Class” για την F-measure.
- **ROC** = receiver operating characteristic: Χαρακτηριστική δέκτη λειτουργίας: χρησιμοποιείται συνήθως για να παρουσιάσει τα αποτελέσματα για τα δυαδικά προβλήματα απόφασης. Οι γραφικές παραστάσεις ROC είναι δυσδιάστατες γραφικές παραστάσεις στις οποίες το ποσοστό του TP Rate σχεδιάζεται στον άξονα Y και το ποσοστό του FP Rate σχεδιάζεται στον άξονα X. Μια γραφική παράσταση ROC απεικονίζει τις σχετικές ανταλλαγές μεταξύ των κερδών (αληθινά θετικά) και των δαπανών (ψεύτικα θετικά).

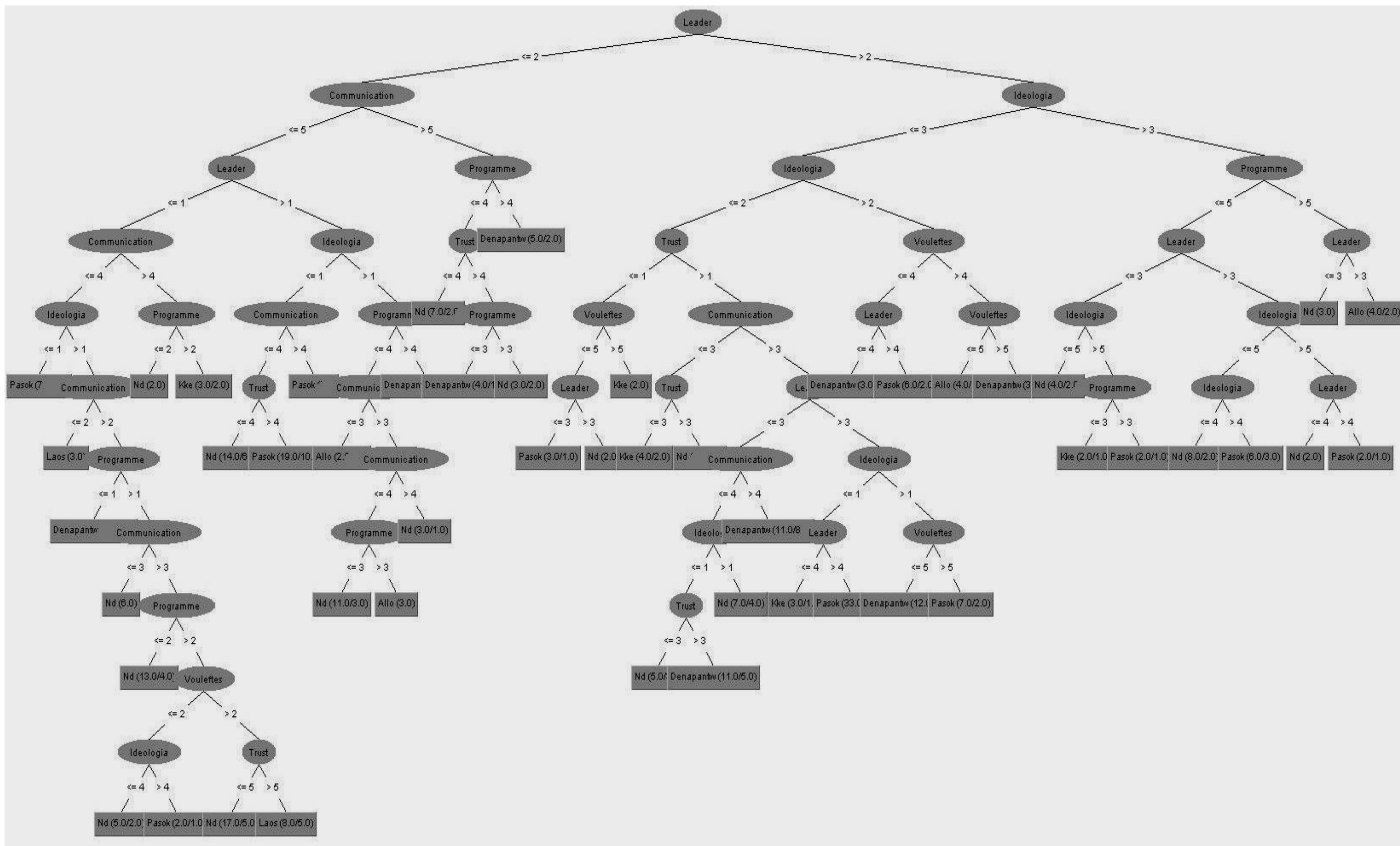
Συνοπτικά, για ένα δυαδικό πρόβλημα απόφασης, ένας ταξινομητής ονομάζει τα παραδείγματα είτε ως θετικά είτε ως αρνητικά, και παράγει τα εξής αποτελέσματα:

Confusion Matrix			Recall	=	$\frac{TP}{TP+FN}$
			Precision	=	$\frac{TP}{TP+FP}$
			True Positive Rate	=	$\frac{TP}{TP+FN}$
			False Positive Rate	=	$\frac{FP}{FP+TN}$
	actual positive	actual negative			
predicted positive	TP	FP			
predicted negative	FN	TN			

Σχήμα 4.3.6.1 Επεξήγηση παραμέτρων, πηγή: “The Relationship Between Precision-Recall and ROC Curves” Jesse Davis, Mark Goadrich (ICML 2006)

Περιγράφοντας και αναλύοντας λίγο περισσότερο τη μήτρα σύγχυσης (Confusion Matrix), θα λέγαμε ότι στην πρώτη σειρά, για παράδειγμα, 78 περιπτώσεις είναι σωστά ταξινομημένες ως ΝΔ, 17 περιπτώσεις είναι λανθασμένα ταξινομημένες ως ΠΑΣΟΚ, 1 περίπτωση είναι λανθασμένα ταξινομημένη ως ΚΚΕ, όπως επίσης και ως ΛΑΟΣ, ενώ 7 περιπτώσεις είναι λανθασμένα ταξινομημένες ως ΔΕΝ ΑΠΑΝΤΩ. Αντίστοιχα θα μπορούσαμε να ερμηνεύσουμε και τις υπόλοιπες σειρές της μήτρας σύγχυσης που μόλις παρουσιάσαμε. Όσον αφορά το δέντρο απόφασης που προκύπτει από τον J48 αλγόριθμο ταξινόμησης, στο Σχήμα 4.3.6.2 απεικονίζεται το αποτέλεσμα της κατηγοριοποίησης των δεδομένων μας απ' όπου τελικά προκύπτει ότι η “Εξόρυξη Γνώσης” μπορεί να αναδειχθεί σε ένα πολύ χρήσιμο εργαλείο ποιοτικής ανάλυσης των δεδομένων που ερευνούμε. Αναλύοντας και περιγράφοντας το δέντρο απόφασης που προέκυψε (η παρουσίασή του αποδείχθηκε δύσκολη λόγω του γραφικού περιβάλλοντος του Weka), παρατηρούμε ότι όσοι, στην κλίμακα 1 έως 6, έχουν απαντήσει >2 στο κριτήριο “Αρχηγός” και >3 στο κριτήριο “Ιδεολογία” και >5 στο κριτήριο “Πρόγραμμα”, εάν έχουν απαντήσει, στην κλίμακα 1 έως 6, για το κριτήριο “Αρχηγός” ≤ 3 τότε ψηφίζουν ΝΔ (3.0 σωστά ταξινομημένες περιπτώσεις), ενώ εάν έχουν απαντήσει >3 τότε ψηφίζουν ΑΛΛΟ (4.0 σωστά ταξινομημένες περιπτώσεις / 2.0 λανθασμένα ταξινομημένες περιπτώσεις). Αν για παράδειγμα οι ψηφοφόροι του δείγματός μας έχουν απαντήσει ≤ 2 για το κριτήριο “Αρχηγός” και >5 για το κριτήριο “Επικοινωνία” και >4 για το κριτήριο “Πρόγραμμα”, τότε παρατηρούμε ότι ψηφίζουν ΔΕΝ ΑΠΑΝΤΩ (5.0 / 2.0), όπου 5 περιπτώσεις είναι σωστά ταξινομημένες και 2 περιπτώσεις είναι λανθασμένα ταξινομημένες. Με παρόμοιο τρόπο μπορούμε να διαβάσουμε και το υπόλοιπο δέντρο απόφασης, απ' όπου μπορούμε να βγάλουμε χρήσιμα συμπεράσματα ή και κρίσιμες πληροφορίες για τη συμπεριφορά των ψηφοφόρων του δείγματός μας και ανάλογα να κατανοήσουμε, να στοχεύσουμε, να προσαρμόσουμε πολιτικές, προεκλογικές στρατηγικές, την τακτική μας.

Το δέντρο απόφασης, λοιπόν, του Σχήματος 4.3.6.2 μας δείχνει τις ταξινομήσεις που προκύπτουν από τις 300 απαντήσεις των ερωτηθέντων του εκπαιδευτικού μας δείγματος. Σε συνέχεια αυτού του δέντρου, θα παρουσιάσουμε τα αποτελέσματα που προκύπτουν από τους 100 υπόλοιπους ερωτηθέντες, θα παρουσιάσουμε δηλαδή το επαληθευτικό μας δείγμα, για να παρατηρήσουμε και να συγκρίνουμε ομοιότητες και διαφορές με το προαναφερθέν εκπαιδευτικό μας δείγμα.



Σχήμα 4.3.6.2 Δέντρο Απόφασης / Δέντρο Ταξινόμησης των δεδομένων του εκπαιδευτικού μας δείγματος

Επαληθευτικό δείγμα:

Θα παρουσιάσουμε, λοιπόν, όπως ήδη προαναφέραμε, τα αποτελέσματα που προέκυψαν από την εφαρμογή του J48 αλγόριθμου ταξινόμησης του υπολογιστικού συστήματος Weka, στο επαληθευτικό δείγμα των 100 ερωτηθέντων. Στη συνέχεια, θα συγκρίνουμε και θα σχολιάσουμε τα αποτελέσματα αυτά σε σχέση με τα αποτελέσματα από το εκπαιδευτικό μας δείγμα. Από την εφαρμογή λοιπόν του αλγορίθμου προέκυψαν τα εξής:

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.5 -M 2

Relation: epalithefsi

Instances: 100

Attributes: 7

Programme

Vouleftes

Ideologia

Leader

Trust

Communication

komma

Test mode: evaluate on training data

=== Classifier model (full training set) ===

J48 pruned tree

Leader <= 3

| Programme <= 2: Nd (23.0/9.0)

| Programme > 2

| | Communication <= 5

| | | Communication <= 3: Pasok (2.0)

| | | Communication > 3

| | | | Programme <= 3: Pasok (7.0/2.0)

| | | | Programme > 3

| | | | | Leader <= 2

| | | | | Programme <= 5

| | | | | | Ideologia <= 4: Nd (13.0/5.0)

| | | | | | Ideologia > 4: Denapantw (3.0/1.0)

| | | | | | Programme > 5

```

| | | | | Trust <= 2: Pasok (6.0/2.0)
| | | | | Trust > 2: Nd (2.0/1.0)
| | | | Leader > 2: Pasok (4.0/2.0)
| | Communication > 5
| | | Ideologia <= 1: Nd (2.0)
| | | Ideologia > 1: Denapantw (4.0/2.0)
Leader > 3
| Trust <= 3
| | Ideologia <= 4
| | | Vouleftes <= 5
| | | Programme <= 2: Syn (4.0/2.0)
| | | Programme > 2: Kke (2.0/1.0)
| | | Vouleftes > 5
| | | | Ideologia <= 1: Pasok (3.0/1.0)
| | | | Ideologia > 1: Denapantw (6.0/3.0)
| | | Ideologia > 4: Nd (2.0)
| Trust > 3
| | Trust <= 5: Pasok (5.0/1.0)
| | Trust > 5
| | | Vouleftes <= 4: Nd (10.0/5.0)
| | | Vouleftes > 4: Pasok (2.0/1.0)

```

Number of Leaves : 18

Size of the tree : 35

Time taken to build model: 0.03 seconds

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	62	62	%
Incorrectly Classified Instances	38	38	%
Kappa statistic	0.4727		
Mean absolute error	0.1474		
Root mean squared error	0.2714		
Relative absolute error	66.9116 %		
Root relative squared error	82.1007 %		
Total Number of Instances	100		

=== Detailed Accuracy by Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.914	0.308	0.615	0.914	0.736	0.856	Nd
0.714	0.125	0.69	0.714	0.702	0.895	Pasok
0.2	0.011	0.5	0.2	0.286	0.939	Kke

	1	0.02	0.5	1	0.667	0.99	Syn
	0	0	0	0	0	0.818	Laos
	0	0	0	0	0	0.794	Allo
	0.636	0.067	0.538	0.636	0.583	0.899	Denapantw
Weighted Avg.	0.62	0.151	0.503	0.62	0.546	0.869	

=== Confusion Matrix ===

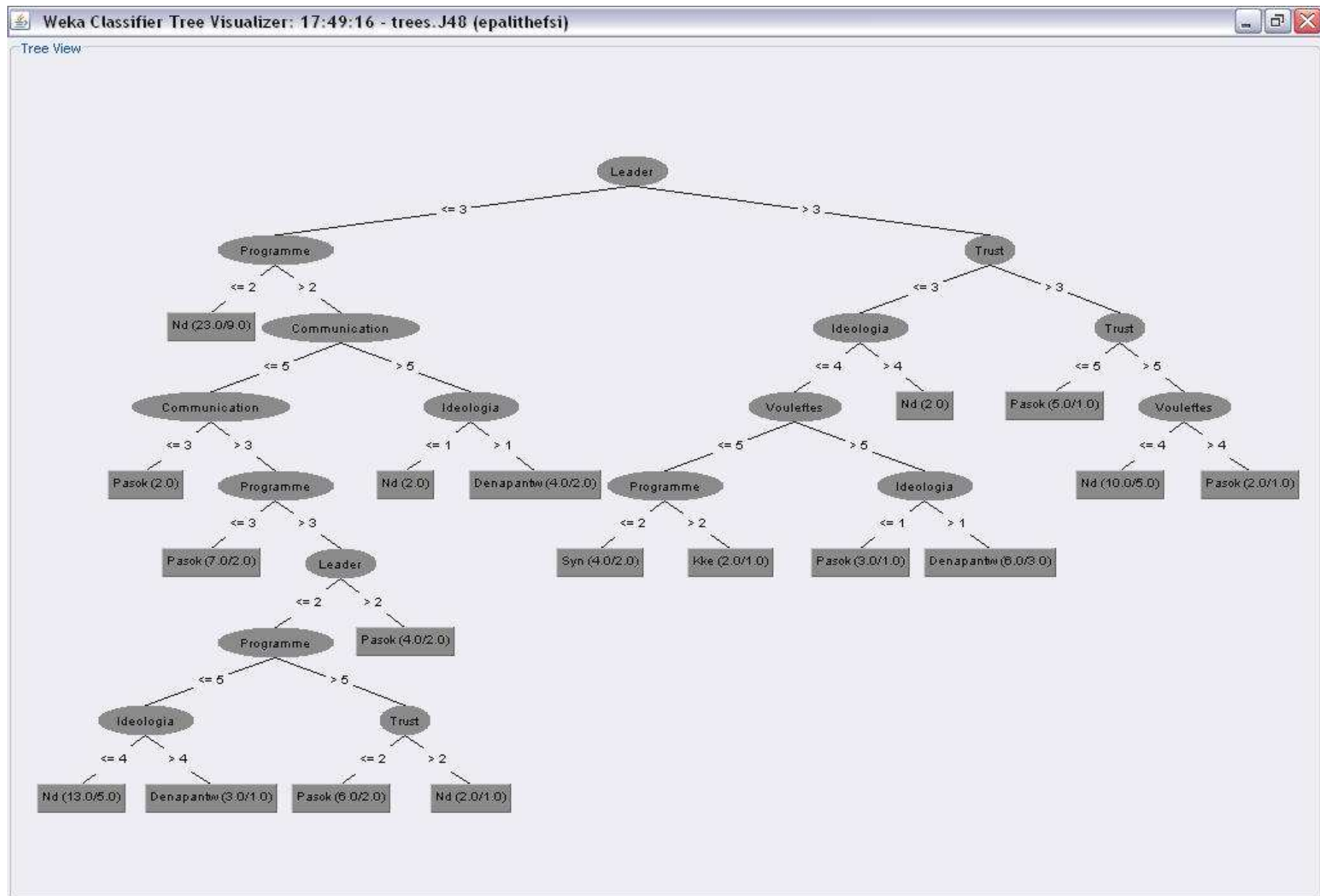
```

a  b  c  d  e  f  g  <-- classified as
32  1  0  0  0  0  2 | a = Nd
 5 20  0  1  0  0  2 | b = Pasok
 0  2  1  1  0  0  1 | c = Kke
 0  0  0  2  0  0  0 | d = Syn
 3  3  0  0  0  0  1 | e = Laos
 8  3  1  0  0  0  0 | f = Allo
 4  0  0  0  0  0  7 | g = Denapantw

```

Παρατηρούμε ότι το 62% του επαληθευτικού μας δείγματος (59% ήταν στο εκπαιδευτικό μας δείγμα) είναι σωστά ταξινομημένο, ενώ το 38% του δείγματός μας είναι λανθασμένα ταξινομημένο. Αυτό σημαίνει ότι η ταξινόμηση που προέκυψε τόσο από το εκπαιδευτικό μας όσο κι από το επαληθευτικό μας δείγμα, σε ποσοστό περίπου 60%, μας δίνει σωστά αποτελέσματα, σωστή κατηγοριοποίηση – ταξινόμηση των δεδομένων μας. Αναλύοντας λίγο περισσότερο τη μήτρα σύγχυσης (Confusion Matrix), προκύπτουν στην πρώτη σειρά, για παράδειγμα, 32 περιπτώσεις να είναι σωστά ταξινομημένες ως ΝΔ, 1 περίπτωση να είναι λανθασμένα ταξινομημένη ως ΠΑΣΟΚ, ενώ 2 περιπτώσεις να είναι λανθασμένα ταξινομημένες ως ΔΕΝ ΑΠΑΝΤΩ. Αντίστοιχα θα μπορούσαμε να ερμηνεύσουμε και τις υπόλοιπες σειρές της μήτρας σύγχυσης του επαληθευτικού μας δείγματος.

Προφανώς, επειδή το δείγμα μας είναι μικρότερο σε σχέση με το εκπαιδευτικό μας δείγμα, οι τιμές του πίνακα: “Λεπτομερής Ακρίβεια ανά Κατηγορία” (Detailed Accuracy by Class) έχουν μία διαφοροποίηση με το εκπαιδευτικό μας δείγμα, αν και παρατηρούμε ότι οι τιμές του ROC Area είναι παραπλήσιες και στα δύο δείγματα που χρησιμοποιήσαμε. Επίσης, λόγω του μικρότερου δείγματος, προκύπτει και μικρότερου μεγέθους (“φύλλα”, “κλαδιά”) δέντρο απόφασης. Θα πρέπει, όμως, να τονίσουμε, ότι και εδώ, στο επαληθευτικό μας δείγμα, το κυρίαρχο κριτήριο είναι αυτό του “Αρχηγού”. Στο Σχήμα 4.6.3.3 απεικονίζεται, ολοκληρωμένα, το δέντρο απόφασης που προέκυψε ως αποτέλεσμα της εφαρμογής του J48 αλγορίθμου στο επαληθευτικό μας δείγμα, απ’ όπου προκύπτουν οι ταξινομήσεις των πολιτών – ψηφοφόρων με βάση τα συγκεκριμένα κριτήρια επιλογής και από το οποίο μπορούμε, με τη χρήση αυτού του “δέντρου”, να καταλήξουμε σε χρήσιμα συμπεράσματα για τη συμπεριφορά των ερωτηθέντων του δείγματός μας.

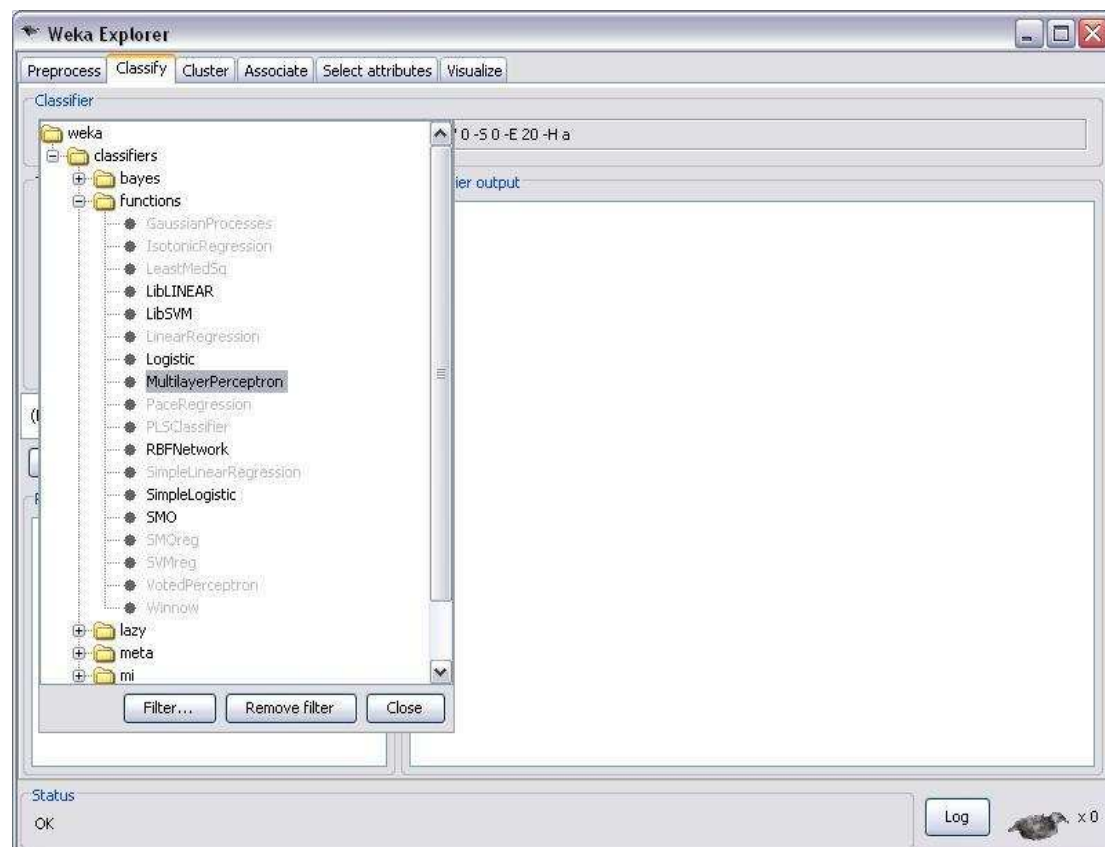


Σχήμα 4.3.6.3 Δέντρο Απόφασης / Δέντρο Ταξινόμησης των δεδομένων του επαληθευτικού μας δείγματος

4.3.7 Μία επιπλέον χρήσιμη ιδιότητα ταξινόμησης του υπολογιστικού συστήματος Weka.

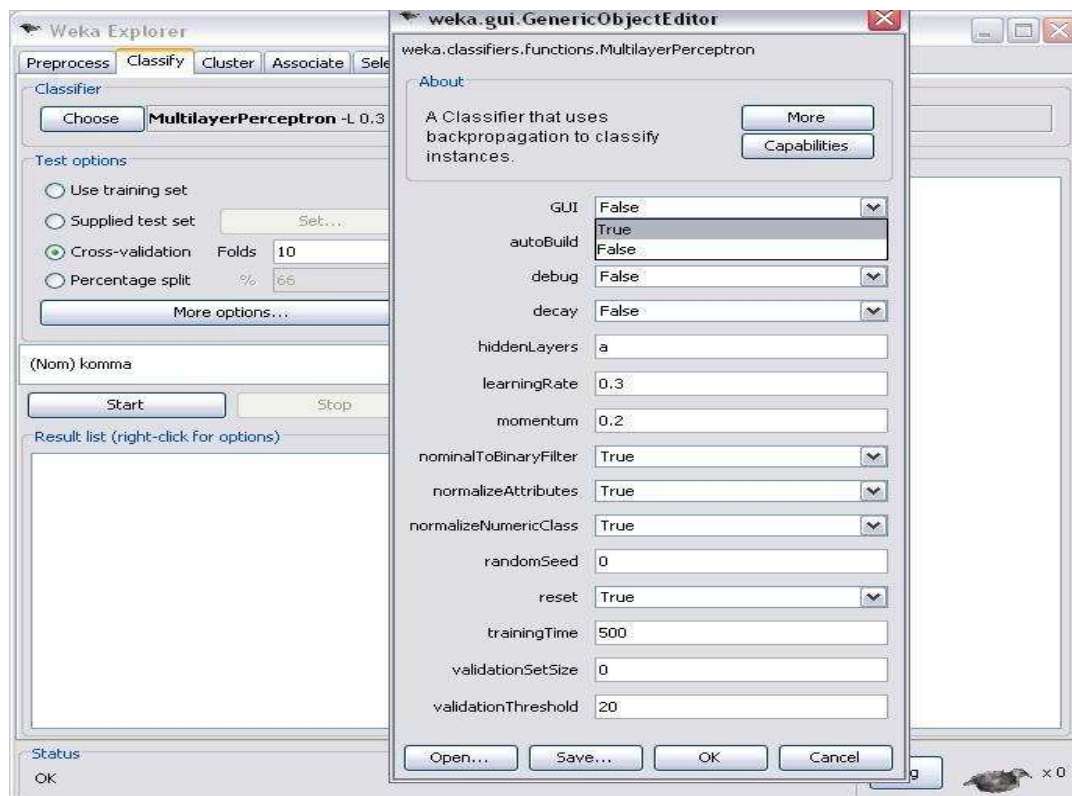
Όπως έχουμε αναφέρει στο κεφάλαιο 3 (παράγραφος 3.7), οι ιδιότητες των εργαλείων του Data Mining έχουν αρκετές εφαρμογές. Μία επιπλέον χρήσιμη ιδιότητα ταξινόμησης του υπολογιστικού συστήματος Weka, είναι αυτή των Νευρωνικών δικτύων. Θα παρουσιάσουμε, λοιπόν, εντελώς συνοπτικά τη δυνατότητα αυτή, όχι τόσο για να τη σχολιάσουμε όσο για να αναδείξουμε τις δυνατότητες που έχει το Weka λογισμικό και η διαδικασία του Data Mining. Ανάλογα, λοιπόν, με το τι θέλει ο χρήστης του συστήματος να αναλύσει και να αναδείξει, μπορεί να εκμεταλλευτεί τα κατάλληλα εργαλεία προκειμένου να εξαχθούν χρήσιμες πληροφορίες, να εξαχθεί “κρυμμένη γνώση”, να γίνει πράξη δηλαδή η “Εξόρυξη Γνώσης”.

Από τον Weka Explorer, λοιπόν, και την καρτέλα “classify”, διαλέγουμε τον “Multilayer Perception” ταξινομητή, που βρίσκεται στο φάκελο functions (Σχήμα 4.3.7.1).



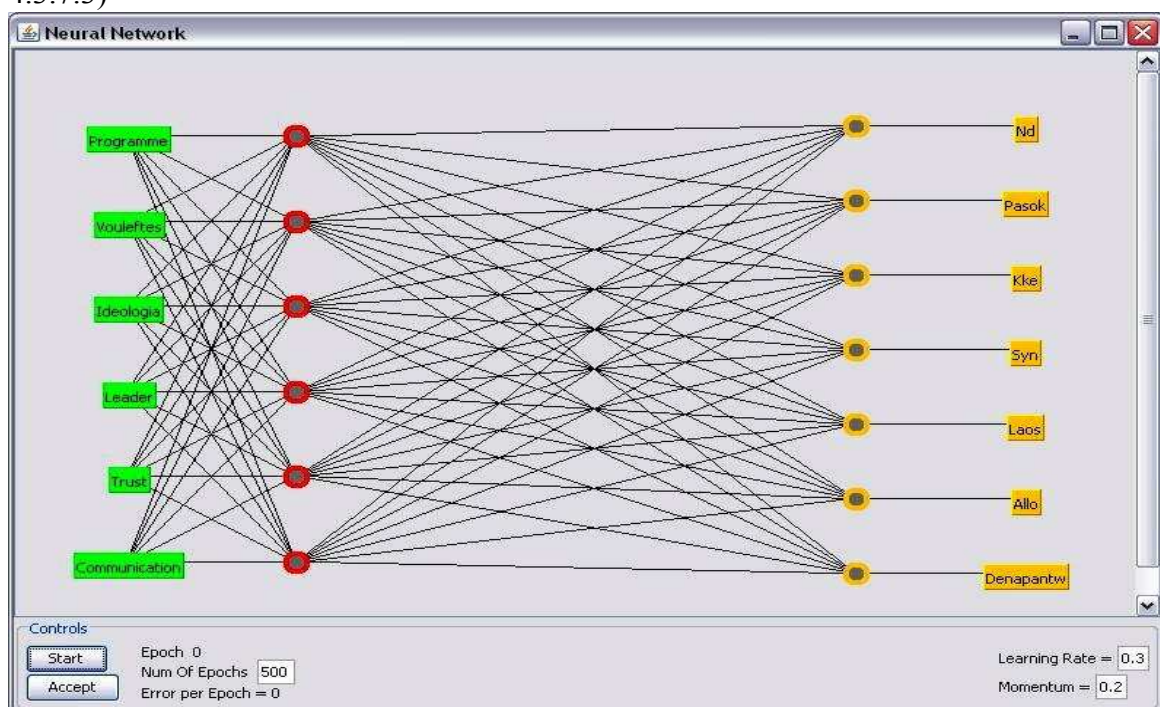
Σχήμα 4.3.7.1 Επιλογή ταξινομητή Νευρωνικού Δικτύου

Στη συνέχεια, για να μπορέσουμε να εμφανίσουμε το Νευρωνικό Δίκτυο, από τις επιλογές του “Multilayer Perception” ταξινομητή, δηλώνουμε την επιλογή **GUI** ως **true** (Σχήμα 4.3.7.2)



Σχήμα 4.3.7.2 Επιλογή εμφάνισης Νευρωνικού Δικτύου

Το αποτέλεσμα της πιο πάνω διαδικασίας είναι η δημιουργία ενός Νευρωνικού δικτύου, όπου στη μία άκρη του υπάρχουν τα αριθμητικά κριτήρια που τέθηκαν προς επιλογή στο δείγμα μας, και από την άλλη μεριά υπάρχει η τελική επιλογή των ερωτηθέντων του δείγματός μας, ως προς το κατηγορικό κριτήριο της ψήφου (Σχήμα 4.3.7.3)



Σχήμα 4.3.7.3 Νευρωνικό Δίκτυο του δείγματος της έρευνάς μας

Αναφέραμε ότι δεν θα σχολιάσουμε τα αποτελέσματα του Νευρωνικού δικτύου διότι, όπως φαίνεται κι από τα ακόλουθα στοιχεία που προέκυψαν από την εφαρμογή του “Multilayer Perception” αλγόριθμου – ταξινομητή, οι λανθασμένες ταξινομήσεις είναι περισσότερες (56.25 %) από τις σωστά ταξινομημένες περιπτώσεις (43.75 %) και η τιμή του kappa – statistic πολύ μικρή (Kappa statistic: 0.2202).

Correctly Classified Instances	175	43.75 %
Incorrectly Classified Instances	225	56.25 %
Kappa statistic	0.2202	
Mean absolute error	0.1949	
Root mean squared error	0.3145	
Relative absolute error	87.9641 %	
Root relative squared error	94.575 %	
Total Number of Instances	400	

Παρόλα αυτά, θεωρήσαμε σημαντικό να αναφερθούμε στη συγκεκριμένη ιδιότητα – δυνατότητα που δίνει στο χρήστη το υπολογιστικό σύστημα Weka, προκειμένου να καταγραφεί ως μία επιπλέον επιλογή για την περαιτέρω ανάλυση των δεδομένων ή της βάσης δεδομένων που θα έχουμε δημιουργήσει και που θα μας βοηθήσει να κάνουμε εφικτή την “Εξόρυξη Γνώσης”.

ΚΕΦΑΛΑΙΟ 5:

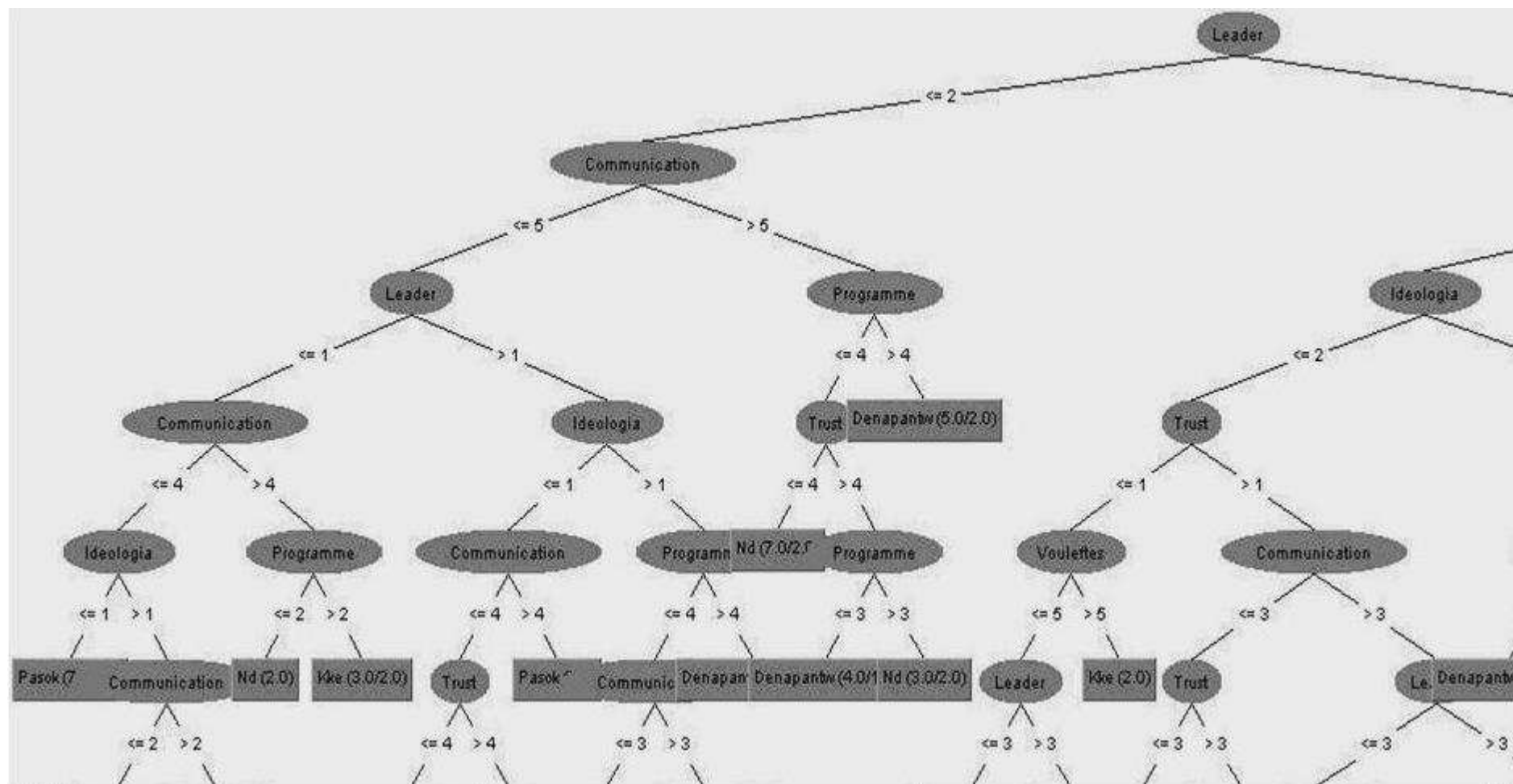
Συμπεράσματα

Υπάρχει, λοιπόν, “κρυμμένη γνώση” σε μια έρευνα κοινής γνώμης; Υπάρχει εκείνη η μέθοδος που μπορεί να πραγματώσει την “εξόρυξη” αυτής της γνώσης και να αναδείξει εκείνες τις πληροφορίες που μπορεί και να φανούν χρήσιμες στον αποφασίζοντα, κατά τη διαδικασία λήψης απόφασης; Με τη παρούσα διπλωματική εργασία πιστεύουμε ότι κάνουμε ένα πολύ σημαντικό πρώτο βήμα προς αυτήν την κατεύθυνση. Αποδείξαμε ότι μέσα σε μια έρευνα υπάρχουν πληροφορίες που μπορούν να δώσουν μια πιο πλήρη και πιο συγκεκριμένη εικόνα των προτιμήσεων των ερωτηθέντων ψηφοφόρων. Και να οδηγηθούμε σε εκείνα τα συμπεράσματα, ως προς τη συμπεριφορά των πολιτών – ψηφοφόρων, που θα κάνουν πιο εύκολη τη λήψη πολιτικών αποφάσεων.

Συνοψίζοντας, όπως αναφέραμε και στην εισαγωγή αυτής της εργασίας, χρειάζεται να κατανοήσουμε και να εξηγήσουμε· δεν αρκεί μόνο να περιγράψουμε και να προβλέψουμε. Αναζητήσαμε, λοιπόν, τα συστήματα-μεθόδους ερευνών αγοράς και ανάλυσης της συμπεριφοράς των πολιτών κατά την εκλογική διαδικασία, πραγματοποιώντας μία βιβλιογραφική αναφορά στις υπάρχουσες μεθόδους ερευνών αγοράς. Στη συνέχεια, αναδείξαμε τα κριτήρια – κλίμακες και την ανάλυση προτιμήσεων – κριτήρια επιλογής (ποιοτικά – ποσοτικά) και καταλήξαμε σε μία προτεινόμενη μεθοδολογία ποιοτικής ανάλυσης συμπεριφοράς των ψηφοφόρων, βασιζόμενοι στη θεωρία της “Εξόρυξης Δεδομένων” (Data Mining) και στη χρήση του Weka λογισμικού.

Η χρήση του υπολογιστικού συστήματος Weka και των εργαλείων που μας παρέχει, μας έδωσε τη δυνατότητα, μέσα από την Προ-επεξεργασία Δεδομένων (Pre-process), τη ταξινόμηση (Classify), τη συσταδοποίηση / ομαδοποίηση (Clustering), να βγάλουμε χρήσιμα συμπεράσματα για τη συμπεριφορά, παραδείγματος χάρη, των ψηφοφόρων που μας δήλωσαν “ΔΕΝ ΑΠΑΝΤΩ”. Έτσι, με τη χρήση του αλγορίθμου συσταδοποίησης k-means, προέκυψε ότι οι συγκεκριμένοι ερωτηθέντες (που μας δήλωσαν “ΔΕΝ ΑΠΑΝΤΩ”), είναι πιο κοντά στη ΝΔ, όταν η ερώτηση αφορούσε τα κριτήρια “Βουλευτές”, “Αρχηγός” και “Επικοινωνία”, ενώ προέκυψε ότι βρίσκονται πιο κοντά στο ΠΑΣΟΚ σε σχέση με τα κριτήρια “Πρόγραμμα”, “Ιδεολογία”, “Εμπιστοσύνη”.

Όμοια, η χρήση του αλγορίθμου ταξινόμησης J48, μας βοήθησε να δημιουργήσουμε κάποιους αρχικούς κανόνες σε σχέση με τη συμπεριφορά των ερωτηθέντων – ψηφοφόρων. Έτσι, για παράδειγμα, όσοι από τους ερωτηθέντες μας απάντησαν/ βαθμολόγησαν, στην κλίμακα μέτρησης από 1 έως 6, το κριτήριο του “Αρχηγού” ≤ 2 , το κριτήριο της “Επικοινωνίας” >5 , το κριτήριο του “Προγράμματος” ≤ 4 και το κριτήριο της “Εμπιστοσύνης” ≤ 4 , τότε ταξινομούνται ως ψηφοφόροι της ΝΔ (7.0 / 2.0), με 7 σωστά ταξινομημένες περιπτώσεις και 2 λανθασμένα ταξινομημένες περιπτώσεις, όπως μας υποδεικνύουν και οι αριθμοί εντός της παρενθέσεως (Σχήμα 5.1).



Σχήμα 5.1 Μέρος του δέντρου απόφασης, που προέκυψε από την εφαρμογή του αλγόριθμου ταξινόμησης J48 στο εκπαιδευτικό μας δείγμα

Παρατηρούμε, λοιπόν, σήμερα, ότι ο αποφασίζων πολίτης, ο ψηφοφόρος, δεν μπορεί να θεωρηθεί δεδομένος, όσον αφορά την τελική επιλογή του. Η ψήφος του, η επιλογή που θα κάνει, δεν μπορεί εκ των προτέρων να προσδιοριστεί με την ίδια ευκολία που γινόταν ίσως τα προηγούμενα χρόνια. Αυτός, πιθανόν, να είναι κι ένας λόγος που οδηγεί τις δημοσκοπήσεις στην αποτυχία πρόβλεψης, καθώς επίσης, όπως αναφέραμε και στο 1^ο κεφάλαιο, και το γεγονός ότι τα δημοσιογραφικά και επικοινωνιακά χαρακτηριστικά επικράτησαν εις βάρος της θεωρίας των ερευνών και της ανάλυσης. Επομένως, σε ένα πιο “ρευστό” πολιτικό περιβάλλον, τα κριτήρια που θα οδηγήσουν στη λήψη απόφασης είναι και πολλαπλά και πιο σύνθετα και διαφορετικής σημασίας / σημαντικότητας. Άρα, η εφαρμογή της πολυκριτήριας ανάλυσης και η προτεινόμενη μας μεθοδολογία της “Εξόρυξης Δεδομένων” έχει, σύμφωνα και με τα όσα αναλύσαμε στην παρούσα διπλωματική, εκείνες τις δυνατότητες ώστε η ποιοτική ανάλυση να συμβάλει όλο και περισσότερο στην εξαγωγή χρήσιμων συμπερασμάτων των επιλογών των πολιτών – ψηφοφόρων, προκειμένου να ληφθούν πιο ορθολογικές πολιτικές αποφάσεις.

Πίσω, λοιπόν, από τα “δημοσκοπικά” χαρακτηριστικά υπάρχει ένας ολόκληρος κόσμος δεδομένων, πληροφοριών και γνώσης που μπορούμε πλέον να αναδείξουμε και να αναλύσουμε. Όπως αναφέραμε χαρακτηριστικά, η εφαρμογή των μεθόδων και των εργαλείων Data Mining οδηγεί στην δημιουργία κανόνων – προτύπων πολιτικής, κομματικής, εκλογικής συμπεριφοράς και επιλογής. Αυτό δεν σημαίνει ότι η έρευνά μας είναι στατική. Αντιθέτως, εφαρμόζοντας τη συγκεκριμένη μεθοδολογία, ανά τακτά χρονικά διαστήματα, μπορούμε να επικαιροποιούμε και να βελτιστοποιούμε αυτούς τους κανόνες. Έτσι, θα μας δοθεί η δυνατότητα να δημιουργήσουμε μια πιο κανονικοποιημένη και μια πιο σταθερή βάση δεδομένων, η οποία, με τη σειρά της, θα μας παρέχει πιο ορθολογικά και πιο “σταθερά” στοιχεία – συμπεράσματα, προκειμένου να λάβουμε τις αποφάσεις μας.

Επίσης, θα πρέπει να τονίσουμε ότι είναι τέτοια η πληθώρα των εργαλείων του Weka λογισμικού, αλλά και άλλων παρόμοιων λογισμικών στατιστικής ανάλυσης και “Εξόρυξης Δεδομένων”, που εξαρτάται από τον μελετητή – ερευνητή το πόσο, το πού και με ποιον τρόπο θα ψάξει να ανακαλύψει, να αναδείξει και να αναλύσει τις “κρυμμένες” πληροφορίες. Να βρει, λοιπόν, όλες εκείνες τις πληροφορίες που θα του φανούν χρήσιμες και για την εξαγωγή συμπερασμάτων και για την όποια υποστήριξή του στη διαδικασία λήψης απόφασης.

Θεωρούμε, λοιπόν, ότι η έρευνα που επιχειρήσαμε και που αναλύσαμε στις προηγούμενες σελίδες, δείχνει ότι η εφαρμογή της Πολυκριτήριας Ανάλυσης και η μεθοδολογία της Εξόρυξης Δεδομένων, μας επιτρέπει να κάνουμε ένα πολύ σημαντικό πρώτο βήμα στο να κατανοήσουμε καλύτερα τη συμπεριφορά των ατόμων, των πολιτών και συνεπώς να πάρουμε, όσο το δυνατόν καλύτερες αποφάσεις.

ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ

- Βερναρδάκης, Χρ. (2005). *Πολιτικά κόμματα και 'μεσαίος χώρος'. Οι ιδεολογικές, πολιτικές και πολιτισμικές συντεταγμένες των σημερινών πολιτικών δυνάμεων*, Σαββάλας, 57-85.
- Βερναρδάκης Χρ. (2007). *Ιδεολογικές Αξίες, Πολιτικές Στάσεις και Ηγεμονία: Οι «ισχυρές μεταβλητές» της πολιτικής και εκλογικής συμπεριφοράς στη συγκυρία των Βουλευτικών εκλογών 2007..*
- Dalton, R. (2006). *Citizen Politics. Public Opinion and Political Parties in Advanced Industrial Democracies*, Washington, CQ Press.
- Ελληνικό Στατιστικό Ινστιτούτο (Ε.Σ.Ι.), (26-29 Απριλίου 2006). *Δημοσκοπήσεις & Στατιστική*, 19^ο Πανελλήνιο Συνέδριο Στατιστικής, Καστοριά.
- Evans, G.A. – Heath, A.F. (1995). *The measurement of Left-Right and Libertarian-Authoritarian Values: A Comparison of Balanced and Un-balanced Scales*, Quality and Quantity 29 (2), 191-206.
- Evans, G.A.-Heath, A.F.-Lalljee, M.G. (1996). *Measuring Left-Right and Libertarian-Authoritarian Values in the British Electorate*, British Journal of Sociology 47 (1), 93-112.
- Νικολακόπουλος Η., (2003). *Οι Πολιτικές Δημοσκοπήσεις στην Ελλάδα: Ιστορία, Προβλήματα, Προοπτικές*, 1^ο ΣΥΝΕΔΡΙΟ ΣΕΔΕΑ.
- Grunberg, G. – Schweisguth, E. (1990). *Liberalisme culturel et liberalisme economique* στο L'électeur français en questions, CEVIPOF, Paris, Presses de FNSP, 45-69.
- Grunberg, G. – Schweisguth, E. (1997). *Recompositions ideologiques* στο Boy, D. – Mayer, N. (dir.), *L'electeur a ses raisons*, Paris, Presses de FNSP, 139-178.
- Lazarsfeld, Paul W., Bernard R. Berelson, and Hazel Gaudet. 1948. *The People's Choice*. New York: Columbia University Press.
- Nonna Mayer, Pascal Perrineau. (1992). *Les comportements politiques*, Armand Colin
- Nonna Mayer, Daniel Boy. (1997). *L'electeur a ses raisons*, Presses de Sciences Po
- Nonna Mayer, Bruno Cautrès. (2004). *Le nouveau désordre électoral*, Presses de Sciences Po
- Michelat, G.(1990). *Les échelles d'attitudes et de comportements* στο *L'electeur français en questions*, CEVIPOF, Paris, Presses de FNSP, 229-236.
- Πάσχου Μ., (2007). *Η στατιστική στην κοινωνική έρευνα*, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών Τμήμα Πολιτικής Επιστήμης και Δημόσιας Διοίκησης.
- Rawls J. (1971). *A Theory of Justice*, The Belknap Press of Harvard University Press
- Tilley, J.R. (2005). *Libertarian-Authoritarian Value Change in Britain, 1974-2001*, Political Studies 53, 442-453.
- Vaus, D. (2002). *Surveys in Social Research*, London, Routledge.

Θεωρία αποφάσεων και Πολυκριτήρια Ανάλυση:

1. Γρηγορούδης, Ε., Μ. Δούμπος, Κ. Ζοπουνίδης και Ν.Φ. Ματσατσίνης (επιμέλεια) (2004). *Πολυκριτήρια ανάλυση αποφάσεων: Μεθοδολογικές προσεγγίσεις και εφαρμογές*, Εκδόσεις Νέων Τεχνολογιών, Αθήνα.
2. Dror Yehezkel.(1964). *Administrative Agencies and Courts: Some Patterns of Interorganizational Relations*, International Review of Administrative Sciences 30: 285-296.
3. Etzioni Amitai (1968). *The Active Society: A Theory of Societal and Political Processes*. New York: Free Press.
4. Δούμπος Μ. (2000). *Πολυκριτήριες Μέθοδοι Ταξινόμησης και Εφαρμογές στην Χρηματοοικονομική Διοίκηση*, Πολυτεχνείο Κρήτης, Τμήμα Μηχανικών Παραγωγής και Διοίκησης, Διδακτορική Διατριβή.
5. Follett M.P. (1947). *The Theory of Social and Economic Organizations*, Talcott Parsons
6. A. Goicoechea, D.R. Hansen and L. Duckstein. (1982). *Multiobjective decision analysis with engineering and business applications*, John Wiley and Sons, New York.
7. C.L. Hwang, A.S.M. Masud, S.R. Paidy and Kl. Yoon. (1979). *Multiple objective decision making – Methods and applications : A state – of – the – Art Survey*, Springer – Verlag, New York.
8. E. Jacquet – Lagreze et J. Siskos, (1982). *Assessing a set of additive utility functions for multicriteria decision – making: The UTA method*, European Journal of Operational Research, Vol. 10, no 2, p. 151-164.
9. E. Jacquet – Lagreze et J. Siskos, (1983). *Multicritère, Hommes et techniques de Methode de decision*, Paris.
10. Keeney R.L. and Raiffa H. (1976). *Multiple objectives in decisions: Preferences and value trade – offs*, John Wiley and Sons, New York.
11. Lindblom, C.E. (1979). “Still Muddling, Not Yet Through”, Public Administration Review 39: 517 – 26.
12. Lindblom C.E. (1979). *Usable Knowledge: Social Science and Social Problem Solving with David K. Cohen*, Yale University Press.
13. Μακροδημήτρης Αντώνης. (1989). *Θεωρία Αποφάσεων*, Σάκκουλας, Ν. Αντώνης.
14. Matsatsinis, N.F., M. Doumpos, C. Zopounidis (1997), *Knowledge acquisition and representation for expert systems in the field of financial analysis*, Expert Systems with Applications, vol. 12, no. 2, pp. 247-262.
15. Ματσατσίνης Νικόλαος. (2004). *Συστήματα Υποστήριξης Αποφάσεων* (Πολυτεχνείο Κρήτης)
16. Roy B. (1985). *Methodologie d’ aide a la decision multicritere*, Economica, Paris.
17. Simon, H.A. (1960). *The new science of management decision*, New York, Harper and Row.

18. Simon, H.A. (1960). *Some further notes on a class of skew distribution functions*, Information and Control, 3, 80-88.
19. Steinbruner J.D.(1974) *The Cybernetic theory of Decision New dimensions of political analysis*, Princeton University Press; 1st Edition edition
20. Weber M. (1954). *A method of multiattribute decision making with incomplete information*, Institute for Operations Research and the Management Sciences, Linthicum, MD, ETATS-UNIS
21. Wildavsky Aaron. (1962). *Dixon-Yates: A Study in Power Politics*, Yale University Press.
22. Zeheny M. (1982). *Multiple criteria decision making*, McGraw – Hill, New York

Data mining:

- Αναστασία Καραναστάση. (2008). *Αποθήκες Δεδομένων-Data Mining-Συστήματα OLAP-Λήψη Αποφάσεων*, Πολυτεχνείο Κρήτης.
- Chan K.P., Stolfo J.S.. (1993). *Toward Parallel and Distributed Learning by meta – Learning*. Working Notes AAAI Work. Knowledge Discovery in Databases, p. 227 – 240.
- Chan K.P., Stolfo J.S. (2001). *On the Accuracy of meta – Learning for Scalable Data Mining*, Journal of Intelligent Information Systems, 8: p. 5 – 28.
- Berry J.A. Michael, Linoff Gordon. (2002). *Data Mining Techniques. Technical report*, Accrue Software.
- David Louis Olson, Dursun Delen. (2008). *Advanced Data Mining Techniques*, p.137- 141 ,Springer.
- Fayyad M. Usama, Piatetsky-Shapiro Gregory, Smyth Padjraic, Uthurusamy Ramasamy. (1996). *Advances in Knowledge Discovery and Data Mining*. MIT Press/AAAI Press.
- Εργαστήριο Νευρωνικών Δικτύων Τμήματος Εφαρμοσμένης Πληροφορικής Πανεπιστημίου Μακεδονίας. (2005). *Εξόρυξη Δεδομένων και Τεχνητά Νευρωνικά Δίκτυα*.
- Fayyad M. Usama, Piatetsky-Shapiro Gregory, Smyth Padjraic, Uthurusamy Ramasamy. (1996). *Advances in Knowledge Discovery and Data Mining*, MIT Press/AAAI Press,.
- Ian H. Witten, Eibe Frank. (2005). *Data mining: practical machine learning tools and techniques*, Morgan Kaufmann.
- Quinlan J.R. (1996). *Improved Use of Continuous Attributes in C4.5*, Journal of Artificial Intelligence Research, 4:77 – 90.
- Matsatsinis, N.F., E. Ioannidou, and E. Grigoroudis (1999), *Customer satisfaction using data mining techniques*, European Symposium on Intelligent Techniques - ESIT '99, Orthodox Academy of Crete, June 1999, Kolymvari, Greece.
- Two Crows Corporation (2005), *Introduction to Data Mining and Knowledge Discovery, Third Edition*, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.).

Πηγές από το διαδίκτυο (internet):

- Το Weka software είναι διαθέσιμο για εγκατάσταση από την ιστοσελίδα:
<http://www.cs.waikato.ac.nz/ml/weka/>
- A comprehensive Weka manual:
<http://www.cs.waikato.ac.nz/~ml/weka/index.html>
- Free Data Mining Tutorial Booklet
<http://www.twocrows.com/booklet.htm>
- The K-Means Algorithm for Unsupervised Learning
http://grb.mnsu.edu/grbts/doc/manual/K_Means.html
- Decision Trees for Supervised Learning
http://grb.mnsu.edu/grbts/doc/manual/J48_Decision_Trees.html
- Performance Measures for Numeric Prediction
http://grb.mnsu.edu/grbts/doc/manual/Error_Measurements.html
- ΕΛΛΗΝΙΚΟ ΣΤΑΤΙΣΤΙΚΟ ΙΝΣΤΙΤΟΥΤΟ
<http://www.esi-stat.gr/>

ΠΑΡΑΡΤΗΜΑ

- **ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ:
“Μεθοδολογία Ποιοτικής Ανάλυσης της Συμπεριφοράς των
Ψηφοφόρων και Λήψη Πολιτικών Αποφάσεων”**

- **Τυπολόγιο στατιστικών δειγματοληπτικών μεθόδων**



ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

- Πείτε μας τη γνώμη σας για το κατά πόσο σας επηρεάζουν οι παρακάτω παράγοντες στην απόφαση που θα πάρετε για το ποιο κόμμα θα ψηφίσετε:

	ΠΟΛΥ	ΑΡΚΕΤΑ	ΜΕΤΡΙΑ	ΛΙΓΟ	ΚΑΘΟΛΟΥ	ΔΕΝ ΓΝΩΡΙΖΩ
Αρχηγός κόμματος						
Ιδεολογία (κοντά στις αρχές μου)						
Πρόγραμμα (προγραμματικές θέσεις)						
Υποψήφιοι βουλευτές						
Εμπιστοσύνη στο κόμμα						
Προεκλογική καμπάνια						
Άλλο.....						

- Κατατάξτε, κατά σειρά σημαντικότητας από το 1 μέχρι το 7, τους παρακάτω παράγοντες που λαμβάνεται υπόψη για να ψηφίσετε (με 1 το σημαντικότερο και με 7 το λιγότερο σημαντικό)

Πρόγραμμα (προγραμματικές θέσεις)	
Υποψήφιοι βουλευτές	
Ιδεολογία (κοντά στις αρχές μου)	
Αρχηγός κόμματος	
Εμπιστοσύνη στο κόμμα	
Προεκλογική καμπάνια – Επικοινωνιακή πολιτική	
Άλλο.....	

- Παρακαλώ απαντήστε στις ακόλουθες ερωτήσεις:

ΚΟΜΜΑΤΑ					ΛΑ. Ο. Σ	ΑΛΛΟ
---------	--	--	--	--	----------	------

Πως κρίνετε το πρόγραμμα του κάθε κόμματος ;	Πολύ καλό						
	Καλό						
	Μέτριο						
	Αδιάφορο						
	Κακό						
	Πολύ κακό						

ΚΟΜΜΑΤΑ					ΛΑ. Ο. Σ	ΑΛΛΟ
----------------	---	---	---	---	-----------------	-------------

Πως κρίνετε τη βουλευτική ομάδα (τους βουλευτές) του κάθε κόμματος ;	Πολύ καλή						
	Καλή						
	Ούτε καλή ούτε κακή						
	Αδιάφορο						
	Κακή						
	Πολύ κακή						

Πως κρίνετε την ιδεολογία του κάθε κόμματος ;	Πολύ καλή						
	Καλή						
	Μέτρια						
	Αδιάφορο						
	Κακή						
	Πολύ κακή						

Πως κρίνετε τον αρχηγό του κάθε κόμματος ;	Πολύ καλός						
	Καλός						
	Ούτε καλός ούτε κακός						
	Αδιάφορο						
	Κακός						
	Πολύ κακός						

Πόση εμπιστοσύνη έχετε σε κάθε κόμμα;	Μεγάλη						
	Αρκετή						
	Μέτρια						
	Αδιάφορο						
	Λίγη						
	Καθόλου						

Πως κρίνετε την επικοινωνιακή πολιτική κάθε κόμματος ;	Πολύ καλή						
	Καλή						
	Μέτρια						
	Αδιάφορο						
	Κακή						
	Πολύ κακή						

- Αξιολογείτε τις επιμέρους πολιτικές που περιλαμβάνονται συνήθως σε ένα πρόγραμμα οποιουδήποτε κόμματος. Πώς κρίνετε την πολιτική του κάθε κόμματος σε θέματα όπως :

ΚΟΜΜΑΤΑ					ΛΑ. Ο. Σ	ΑΛΛΟ
---------	---	---	---	---	----------	------

Παιδεία;	Πολύ καλή						
	Καλή						
	Μέτρια						
	Κακή						
	Πολύ κακή						
	Δεν γνωρίζω						

Εργασία;	Πολύ καλή						
	Καλή						
	Μέτρια						
	Κακή						
	Πολύ κακή						
	Δεν γνωρίζω						

Οικονομία;	Πολύ καλή						
	Καλή						
	Μέτρια						
	Κακή						
	Πολύ κακή						
	Δεν γνωρίζω						

Κοινωνική πολιτική;	Πολύ καλή						
	Καλή						
	Μέτρια						
	Κακή						
	Πολύ κακή						
	Δεν γνωρίζω						

Υγεία;	Πολύ καλή						
	Καλή						
	Μέτρια						
	Κακή						
	Πολύ κακή						
	Δεν γνωρίζω						

ΚΟΜΜΑΤΑ					ΛΑ. Ο. Σ	ΑΛΛΟ
----------------	---	---	---	--	-----------------	-------------

Πολιτισμός;	Πολύ καλή						
	Καλή						
	Μέτρια						
	Κακή						
	Πολύ κακή						
	Δεν γνωρίζω						

Εξωτερική Πολιτική;	Πολύ καλή						
	Καλή						
	Μέτρια						
	Κακή						
	Πολύ κακή						
	Δεν γνωρίζω						

Δημόσια Τάξη;	Πολύ καλή						
	Καλή						
	Μέτρια						
	Κακή						
	Πολύ κακή						
	Δεν γνωρίζω						

Εθνική Άμυνα;	Πολύ καλή						
	Καλή						
	Μέτρια						
	Κακή						
	Πολύ κακή						
	Δεν γνωρίζω						

Αγροτική πολιτική;	Πολύ καλή						
	Καλή						
	Μέτρια						
	Κακή						
	Πολύ κακή						
	Δεν γνωρίζω						

Μεταναστευτική πολιτική;	Πολύ καλή						
	Καλή						
	Μέτρια						
	Κακή						
	Πολύ κακή						
	Δεν γνωρίζω						

- Αξιολογώντας τα πολιτικά κόμματα πάνω σε όλα τα προηγούμενα κριτήρια, κατατάξτε τα κατά σειρά προτίμησης από το 1 μέχρι το 5 (με 1 εκείνο το κόμμα της 1^{ης} προτίμησής σας και 5 της τελευταίας) :

				ΛΑ. Ο. Σ.

- Αξιολογήστε κατά σειρά σημαντικότητας για εσάς, από το 1 μέχρι το 11, τις πολιτικές που περιλαμβάνονται σε ένα πρόγραμμα (με 1 το πιο σημαντικό, μέχρι το 11 για αυτό που θεωρείτε λιγότερο σημαντικό)

ΠΑΙΔΕΙΑ	
ΠΟΛΙΤΙΣΜΟΣ	
ΚΟΙΝΩΝΙΚΗ ΠΟΛΙΤΙΚΗ	
ΥΓΕΙΑ	
ΟΙΚΟΝΟΜΙΑ	
ΘΕΜΑΤΑ ΕΡΓΑΣΙΑΣ	
ΑΓΡΟΤΙΚΗ ΠΟΛΙΤΙΚΗ	
ΕΞΩΤΕΡΙΚΗ ΠΟΛΙΤΙΚΗ	
ΑΜΥΝΑ	
ΜΕΤΑΝΑΣΤΕΥΤΙΚΗ ΠΟΛΙΤΙΚΗ	
ΔΗΜΟΣΙΑ ΤΑΞΗ	
ΑΛΛΟ	

- Παρακαλώ αξιολογήστε τους πολιτικούς αρχηγούς με βάση τα παρακάτω κριτήρια:

<u>ΑΡΧΗΓΟΙ ΚΟΜΜΑΤΩΝ:</u>		ΚΑΡΑΜΑΝΛΗΣ Κ.	ΠΑΠΑΝΔΡΕΟΥ Γ.	ΠΑΠΑΡΗΓΑ Α.	ΑΛΑΒΑΝΟΣ Α.	ΚΑΡΑΤΖΑΦΕΡΗΣ Γ.
------------------------------	--	------------------	------------------	----------------	----------------	--------------------

Πόση ΑΞΙΟΠΙΣΤΙΑ νομίζετε ότι διαθέτει ο / η :	Πολύ υψηλή					
	Υψηλή					
	Μέτρια					
	Χαμηλή					
	Καθόλου					
	ΔΞ/ΔΑ					
Πόσο ΕΙΛΙΚΡΙΝΗΣ πιστεύετε ότι είναι ο / η :	Πολύ					
	Αρκετά					
	Μέτρια					
	Λίγο					
	Καθόλου					
	ΔΞ/ΔΑ					

Πόσο ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΣ πιστεύετε ότι είναι ο / η :	Πολύ					
	Αρκετά					
	Μέτρια					
	Λίγο					
	Καθόλου					
	ΔΞ/ΔΑ					
Πόσο ΕΜΠΕΙΡΟΣ πιστεύετε ότι είναι ο / η :	Πολύ					
	Αρκετά					
	Μέτρια					
	Λίγο					
	Καθόλου					
	ΔΞ/ΔΑ					

ΑΡΧΗΓΟΙ ΚΟΜΜΑΤΩΝ:		ΚΑΡΑΜΑΝΛΗΣ	ΠΑΠΑΝΔΡΕΟΥ	ΠΑΠΑΡΗΓΑ	ΑΛΑΒΑΝΟΣ	ΚΑΡΑΤΖΑΦΕΡΗΣ
------------------------------	--	-------------------	-------------------	-----------------	-----------------	---------------------

Πόσο πιστεύετε ότι ΕΜΠΝΕΕΙ (έχει όραμα) ο / η :	Πολύ					
	Αρκετά					
	Μέτρια					
	Λίγο					
	Καθόλου					
	ΔΞ/ΔΑ					
Πόσο καλή ΑΜΕΣΟΤΗΤΑ ΕΠΙΚΟΙΝΩΝΙΑΣ ΜΕ ΤΟ ΛΑΟ νομίζετε ότι έχει ο / η :	Πολύ καλή					
	Καλή					
	Ουδέτερη					
	Χαμηλή					
	Καθόλου					
	ΔΞ/ΔΑ					

Πόσο ΧΑΡΙΣΜΑΤΙΚΟΣ πιστεύετε ότι είναι ο / η :	Πολύ					
	Αρκετά					
	Μέτρια					
	Λίγο					
	Καθόλου					
	ΔΞ/ΔΑ					
Πόση ΕΛΠΙΔΑ ΚΑΙ ΠΡΟΟΠΤΙΚΗ ΓΙΑ ΤΟ ΜΕΛΛΟΝ νομίζετε ότι προσφέρει ο / η :	Πολύ					
	Αρκετά					
	Μέτρια					
	Λίγο					
	Καθόλου					
	ΔΞ/ΔΑ					

Πόσο ΩΦΕΛΙΜΟΣ για τη χώρα πιστεύετε ότι είναι ο / η :	Πολύ					
	Αρκετά					
	Μέτρια					
	Λίγο					
	Καθόλου					
	ΔΞ/ΔΑ					
Πόσο ΕΝΤΙΜΟΣ πιστεύετε ότι είναι ο / η :	Πολύ					
	Αρκετά					
	Μέτρια					
	Λίγο					
	Καθόλου					
	ΔΞ/ΔΑ					

- **Αξιολογώντας τους αρχηγούς πάνω σε όλα τα προηγούμενα κριτήρια, κατατάξτε τους κατά σειρά προτίμησης από το 1 μέχρι το 5 (με 1 εκείνον της 1^{ης} προτίμησής σας και 5 της τελευταίας) :**

ΚΑΡΑΜΑΝΛΗΣ	ΠΑΠΑΝΔΡΕΟΥ	ΠΑΠΑΡΗΓΑ	ΑΛΑΒΑΝΟΣ	ΚΑΡΑΤΖΑΦΕΡΗΣ

Αν την επόμενη Κυριακή είχαμε βουλευτικές εκλογές, ποιο κόμμα θα ψηφίζατε;

					ΛΑ. Ο. Σ.	ΑΛΛΟ	ΛΕΥΚΟ/ ΑΚΥΡΟ	ΔΕΝ ΑΠΑΝΤΩ
Ψήφος								

ΑΝ ΕΠΙΛΕΞΑΤΕ 'ΑΛΛΟ' ΑΝΑΦΕΡΕΤΕ ΤΟ ΚΟΜΜΑ ΤΗΣ ΠΡΟΤΙΜΗΣΗΣ ΣΑΣ :

.....

- Πόσο βέβαιος/η είστε για την παραπάνω επιλογή σας;

- ☐ ΑΠΟΛΥΤΑ ΒΕΒΑΙΟΣ
- ☐ ΜΑΛΛΟΝ ΒΕΒΑΙΟΣ
- ☐ ΟΧΙ ΚΑΙ ΤΟΣΟ ΒΕΒΑΙΟΣ
- ☐ ΚΑΘΟΛΟΥ ΒΕΒΑΙΟΣ
- ☐ ΔΕΝ ΓΝΩΡΙΖΩ / ΔΕΝ ΑΠΑΝΤΩ

- Ανεξάρτητα από το κόμμα που θα ψηφίζατε αν την Κυριακή είχαμε εκλογές, ποιο κόμμα πιστεύετε ότι θα τις κέρδιζε;

					ΛΑ. Ο. Σ.	ΑΛΛΟ	ΔΕΝ ΓΝΩΡΙΖΩ/ ΔΕΝ ΑΠΑΝΤΩ
Πρόβλεψη							

- Τι ψηφίσατε στις προηγούμενες βουλευτικές εκλογές;

					ΛΑ. Ο. Σ.	ΑΛΛΟ	ΛΕΥΚΟ/ ΑΚΥΡΟ	ΔΕΝ ΓΝΩΡΙΖΩ/ ΔΕΝ ΑΠΑΝΤΩ
Ψήφος								

ΓΕΝΙΚΕΣ ΠΛΗΡΟΦΟΡΙΕΣ

1. ΦΥΛΟ

- ☐ ΑΝΔΡΑΣ
☐ ΓΥΝΑΙΚΑ

2. ΗΛΙΚΙΑ

.....

3. ΜΟΡΦΩΤΙΚΟ ΕΠΙΠΕΔΟ

- ☐ ΑΠΟΦΟΙΤΟΣ ΔΗΜΟΤΙΚΟΥ
☐ ΑΠΟΦΟΙΤΟΣ ΔΕΥΤ/ΘΜΙΑΣ
ΕΚΠΑΙΔΕΥΣΗΣ
☐ ΑΠΟΦΟΙΤΟΣ ΙΕΚ
☐ ΑΠΟΦΟΙΤΟΣ ΑΕΙ - ΤΕΙ
☐ ΑΛΛΟ.....

4. ΕΠΑΓΓΕΛΜΑ

- ☐ ΔΗΜΟΣΙΟΣ ΥΠΑΛΛΗΛΟΣ
☐ ΙΔΙΩΤΙΚΟΣ ΥΠΑΛΛΗΛΟΣ
☐ ΕΛΕΥΘΕΡΟΣ ΕΠΑΓΓ/ΤΙΑΣ
☐ ΣΥΝΤΑΞΙΟΥΧΟΣ
☐ ΟΙΚΙΑΚΑ
☐ ΦΟΙΤΗΤΗΣ
☐ ΑΓΡΟΤΗΣ
☐ ΑΝΕΡΓΟΣ
☐ ΑΛΛΟ.....

5. Συμμετέχετε ενεργά σε κάποιο κομματικό οργανισμό;

- ☐ ΝΑΙ
Αν ναι σε ποιον;
☐ ΟΧΙ

5. Από πού ενημερώνεστε σχετικά με τις πολιτικές εξελίξεις; (κατάταξη)

- ☐ ΤΗΛΕΟΡΑΣΗ
• ΑΘΗΝΑΪΚΑ ΚΑΝΑΛΙΑ ☐
• ΤΟΠΙΚΑ ΚΑΝΑΛΙΑ ☐
☐ ΕΦΗΜΕΡΙΔΕΣ
• ΑΘΗΝΑΪΚΕΣ ☐
• ΤΟΠΙΚΕΣ / ΕΠΑΡΧΙΑΚΕΣ ☐
☐ ΔΙΑΔΙΚΤΥΟ / INTERNET
☐ ΑΠΟ ΕΝΤΥΠΑ / ΜΕΛΗ
ΚΟΜΜΑΤΩΝ
☐ ΣΥΖΗΤΗΣΕΙΣ ΜΕ ΦΙΛΟΥΣ
☐ ΑΛΛΟ.....

ΗΜΕΡΟΜΗΝΙΑ
ΣΥΜΠΛΗΡΩΣΗΣ:

**ΕΥΧΑΡΙΣΤΟΥΜΕ
ΓΙΑ ΤΟΝ ΧΡΟΝΟ ΣΑΣ**

ΤΥΠΟΛΟΓΙΟ

ΑΠΛΗ ΤΥΧΑΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ (Simple random sampling)

Συμβολισμοί – Βασικοί ορισμοί:

Μέγεθος του πληθυσμού: N

Μέγεθος του δείγματος: n

Σύνολο (άθροισμα) του πληθυσμού: $Y = \sum_{i=1}^N Y_i$

Σύνολο (άθροισμα) του δείγματος: $y = \sum_{i=1}^n y_i$

Μέση τιμή του Πληθυσμού: $\bar{Y} = Y/N = \frac{1}{N} \sum_{i=1}^N Y_i$

Μέση τιμή του δείγματος: $\bar{y} = y/n = \frac{1}{n} \sum_{i=1}^n y_i$

Διασπορά του πληθυσμού: $S^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$

Τυπική απόκλιση του πληθυσμού: $S = \sqrt{S^2}$

Διασπορά του δείγματος: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$

Τυπική απόκλιση του δείγματος: $s = \sqrt{s^2}$

Δειγματοληπτικό κλάσμα: $f = n/N$

Εκτιμητές:

$\bar{y} = y/n = \frac{1}{n} \sum_{i=1}^n y_i$ ($E \bar{y} = \bar{Y}$, άρα $\hat{\bar{Y}} = \bar{y}$)

Διακύμανση της μέσης τιμής του δείγματος: $\text{var } \bar{y} = \frac{S^2}{n} \cdot (1-f)$,

Εκτίμηση της διακύμανσης της μέσης τιμής του δείγματος :

$$\text{var } \bar{y} = \frac{s^2}{n} \cdot (1 - f) \quad (\text{αφού } E s^2 = S^2)$$

Εκτίμηση του αθροίσματος του πληθυσμού: $\hat{Y} = N \cdot \bar{y}$ ($E \hat{Y} = Y$)

$$\text{Διασπορά του εκτιμητή } \hat{Y} : \text{var } \hat{Y} = \frac{N^2 S^2}{n} \cdot (1 - f)$$

$$\text{Εκτίμηση της διασποράς του εκτιμητή } \hat{Y} : \text{var } \hat{Y} = \frac{N^2 s^2}{n} \cdot (1 - f) \quad (\text{αφού } E s^2 = S^2)$$

Διαστήματα εμπιστοσύνης:

Για τη μέση τιμή $\bar{Y}$:

$$100 \cdot (1 - \alpha) \% \delta.ε. = \left\{ \bar{y} \pm z_{\alpha/2} \cdot s \cdot \sqrt{\frac{1 - f}{n}} \right\}$$

Για το άθροισμα του πληθυσμού Y:

$$100 \cdot (1 - \alpha) \% \delta.ε. = \left\{ N \cdot \bar{y} \pm z_{\alpha/2} \cdot s \cdot N \cdot \sqrt{\frac{1 - f}{n}} \right\}$$

Μέγεθος δείγματος για την εκτίμηση της μέσης τιμής $\bar{Y}$, με μέγιστο σφάλμα d:

$$n \geq \frac{N}{\left(\frac{d}{z_{\alpha/2} \cdot s} \right)^2 \cdot N + 1} \quad \text{ή} \quad n \geq \frac{z_{\alpha/2}^2 \cdot s^2}{d^2} \quad \text{για μεγάλο } N$$

Μέγεθος δείγματος για την εκτίμηση του αθροίσματος Y του πληθυσμού, με μέγιστο σφάλμα d:

$$n \geq \frac{N}{\left(\frac{d}{z_{\alpha/2} \cdot s} \right)^2 \cdot \frac{1}{N} + 1} \quad \text{ή} \quad n \geq \frac{N^2 \cdot z_{\alpha/2}^2 \cdot s^2}{d^2} \quad \text{για μεγάλο } N$$

Συντελεστές μεταβλητικότητας:

$$CV(\bar{y}) = \frac{SE(\bar{y})}{\bar{y}} = \frac{\sqrt{\text{var } \bar{y}}}{\bar{y}}, \quad CV(\hat{Y}) = \frac{SE(\hat{Y})}{\hat{Y}} = \frac{\sqrt{\text{var } \hat{Y}}}{\hat{Y}}$$

Εκτίμηση ποσοστών:

$$\text{Ποσοστό } P \text{ στον πληθυσμό: } P = \frac{1}{N} \sum_{i=1}^N Y_i$$

$$\text{Ποσοστό } p \text{ στο δείγμα: } p = \frac{1}{n} \sum_{i=1}^n y_i$$

$$E p = P$$

$$\text{Διασπορά στον πληθυσμό: } S^2 = \frac{NP(1-P)}{N-1}$$

$$\text{Διασπορά στο δείγμα: } s^2 = \frac{np(1-p)}{n-1}$$

$$E s^2 = S^2$$

$$\text{Τυπική απόκλιση στον πληθυσμό: } S = \sqrt{S^2}$$

$$\text{Τυπική απόκλιση στο δείγμα: } s = \sqrt{s^2}$$

$$\text{var } p = \frac{1-f}{n} \cdot \frac{NP(1-P)}{N-1}, \quad \text{var } \hat{p} = (1-f) \cdot \frac{p(1-p)}{n-1}$$

$$Y = NP, \quad \hat{Y} = N \cdot p, \quad E \hat{Y} = Y, \quad \text{var } \hat{Y} = \frac{N(N-n)}{n-1} \cdot p \cdot (1-p)$$

$$\underline{100 (1-\alpha)\% \text{ δ.ε. για την } P} = \left(p \pm z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)(1-f)}{n-1}} \right)$$

$$\underline{100 (1-\alpha)\% \text{ δ.ε. για το } NP} = \left(N \cdot p \pm z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)(N-n)N}{n-1}} \right)$$

Μέγεθος δείγματος n για την εκτίμηση του P, με μέγιστο σφάλμα d:

$$n \geq \frac{N \left(1 + \frac{d^2}{z_{a/2}^2 \cdot p \cdot (1-p)} \right)}{1 + \frac{N \cdot d^2}{z_{a/2}^2 \cdot p \cdot (1-p)}}$$

$$n > \frac{N}{1 + \frac{N \cdot d^2}{z_{a/2}^2 \cdot p \cdot (1-p)}}, \text{ όταν το d είναι κοντά στο μηδέν.}$$

$$n > \frac{p(1-p)z_{a/2}^2}{d^2}, \text{ για μεγάλα N.}$$

Μέγεθος δείγματος n για την εκτίμηση του NP, με μέγιστο σφάλμα d:

$$n \geq \frac{p(1-p)N^2 + (d^2 / z_{a/2})}{p(1-p)N + (d^2 / z_{a/2})}$$

Λογοεκτιμητές

$$\text{Η εκτίμηση του λόγου R} = \frac{\sum_{i=1}^N Y_i}{\sum_{i=1}^N X_i} = \frac{Y}{X} = \frac{\bar{Y}}{\bar{X}} \text{ του πληθυσμού είναι } \hat{R} = r = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} = \frac{\bar{y}}{\bar{x}}$$

Για τη διασπορά της εκτίμησης r έχουμε:

$$\text{var } \hat{R} = V(r) = \frac{1-f}{n\bar{X}^2} \cdot \frac{\sum_{i=1}^n (y_i - rx_i)^2}{n-1}, \text{ όπου η πληθυσμιακή μέση τιμή } \bar{X}, \text{ η οποία}$$

όταν είναι άγνωστη εκτιμάται κατά προσέγγιση από τη μέση τιμή $\bar{x}$ του δείγματος.

Το τυπικό σφάλμα είναι: $SE(r) = \sqrt{V(r)}$.

Η εκτίμηση του λόγου του μέσου πληθυσμού $\bar{Y}$ είναι: $\hat{\bar{Y}}_R = r \cdot \bar{X} = \frac{\bar{y}}{\bar{x}} \cdot \bar{X} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \cdot \bar{X}$

Η διακύμανση της εκτίμησης $\hat{\bar{Y}}_R$ είναι: $\text{var}(\hat{\bar{Y}}_R) = \left(\frac{N-n}{n}\right) \cdot \frac{\sum_{i=1}^n (y_i - rx_i)^2}{n(n-1)}$ και το τυπικό σφάλμα είναι: $\text{SE}(\hat{\bar{Y}}_R) = \text{var}(\hat{\bar{Y}}_R)^{1/2}$.

Η εκτίμηση του λόγου του συνολικού πληθυσμού Y είναι:

$$\hat{Y} = r \cdot X = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \cdot X = \frac{\bar{y}}{\bar{x}} \cdot X$$

$$\text{Var}(\hat{Y}) = X^2 \cdot \left(\frac{N-n}{N}\right) \cdot \frac{1}{n\bar{x}^2} \cdot \frac{\sum_{i=1}^n (y_i - rx_i)^2}{n-1}$$

$$\text{SE}(\hat{Y}) = \sqrt{\text{var}(\hat{Y})}$$

ΣΤΡΩΜΑΤΟΠΟΙΗΜΕΝΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ (Stratified random sampling)

Χωρίζουμε τον πληθυσμό N ατόμων σε M υποπληθυσμούς (στρώματα). Το κάθε στρώμα έχει πλήθος ατόμων: N_m , $m=1,2,3,\dots,M$ και $\sum_{m=1}^M N_m = N$. Στο κάθε στρώμα χωριστά κάνουμε Α.Τ.Δ. Θα συμβολίσουμε με Y την τ.μ., που μελετάμε και θα έχουμε:

Y_{mj} : την τιμή της j -μονάδας του m -στρώματος, $j=1,2,\dots,N_m$ και $m=1,2,\dots,M$.

$\bar{Y}_m$: τη μέση τιμή της τ.μ. Y στο στρώμα m . $\bar{Y}_m = \frac{\sum_{i=1}^{N_m} Y_{mi}}{N_m}$, $m=1,2,\dots,M$

Y_m : το άθροισμα των τιμών της τ.μ. Y στο στρώμα m , $m=1,2,\dots,M$.

$$S_m^2: \text{τη διασπορά της τ.μ. } Y \text{ στο στρώμα } m, S_m^2 = \frac{\sum_{i=1}^{N_m} (Y_{mi} - \bar{Y}_m)^2}{N_m - 1}, N_m \geq 2, m=1,2,\dots,M$$

Για το δείγμα θα έχουμε τα σύμβολα: $\bar{y}_m, y_m, s_m^2$ για τη μέση τιμή, το άθροισμα των τιμών και τη διασπορά της τ.μ. Y στο στρώμα $m=1,2,\dots,M$ του δείγματος
 n_m : το μέγεθος του δείγματος στο στρώμα m
 y_{mj} : η τιμή της j -μονάδας στο στρώμα $m=1,2,\dots,M$ του δείγματος, με $j=1,2,\dots, n_m$.

$$\sum_{m=1}^M n_m = n$$

$$\bar{y}_m = \frac{\sum_{i=1}^{n_m} y_{mi}}{n_m}, m=1,2,\dots,M$$

$$s_m^2 = \frac{\sum_{i=1}^{n_m} (y_{mi} - \bar{y}_m)^2}{n_m - 1}, n_m \geq 2, m = 1,2,\dots,M$$

Εκτίμηση μέσου και συνολικού πληθυσμού:

Εκτίμηση μέσου

Η εκτίμηση του μέσου $\bar{Y} = \frac{\sum_{m=1}^M N_m \cdot \bar{Y}_m}{N} = \sum_{m=1}^M W_m \bar{Y}_m$ είναι ο εκτιμητής:

$$\bar{y}_{st} = \frac{\sum_{m=1}^M N_m \cdot \bar{y}_m}{N} = \sum_{m=1}^M W_m \bar{y}_m, \text{ όπου } W_m = \frac{N_m}{N}.$$

Για τη διακύμανση της εκτίμησης του $\bar{y}_{st}$, $V(\bar{y}_{st})$, θα έχουμε ανάλογα με τις περιπτώσεις:

Α) Εάν η κατανομή του δείγματος γίνεται αυθαίρετα στα διάφορα στρώματα (δηλαδή το δείγμα, που επιλέγεται σε ένα στρώμα είναι ανεξάρτητο από οποιοδήποτε δείγμα άλλου στρώματος), θα έχουμε:

$$V(\bar{y}_{st}) = \frac{1}{N^2} \cdot \sum_{m=1}^M N_m^2 \cdot V(\bar{y}_m) = \frac{1}{N^2} \cdot \sum_{m=1}^M N_m^2 \frac{S_m^2}{n_m} \left(\frac{N_m - n_m}{N_m} \right) = \sum_{m=1}^M W_m^2 \cdot \frac{1 - f_m}{n_m} \cdot S_m^2, \text{ όπου}$$

$$f_m = \frac{n_m}{N_m}.$$

$$SE(\overline{y_{st}}) = \sqrt{V(\overline{y_{st}})}$$

Αν δε γνωρίζω το S_m^2 παίρνω την εκτίμησή του, δηλαδή το s_m^2 .

Β) Αναλογική κατανομή του δείγματος: $n_i = n \cdot \frac{N_i}{N}$

$$V(\overline{y_{st,p}}) = \frac{1-f}{n} \cdot \sum_{m=1}^M W_m S_m^2 = \frac{1}{n} \left(\frac{N-n}{N} \right) \sum_{m=1}^M W_m S_m^2, \text{ όπου } f = \frac{n}{N}.$$

Αν $S_m^2 = S^2$, τότε: $V(\overline{y_{st,p}}) = \frac{1-f}{n} \cdot S^2$.

$$SE(\overline{y_{st,p}}) = \sqrt{V(\overline{y_{st,p}})}.$$

Αν δε γνωρίζω το S_m^2 παίρνω την εκτίμησή του, δηλαδή το s_m^2 .

Γ) Κατανομή Neyman:

- Αρίστη κατανομή με σταθερό κόστος ανά μονάδα: $n_i = n \cdot \frac{W_m S_m}{\sum_{m=1}^M W_m S_m}$

$$V(\overline{y_{st}})_{Ne} = \frac{(\sum_{m=1}^M W_m S_m)^2}{n} - \frac{\sum_{m=1}^M W_m S_m^2}{N}$$

$$SE(\overline{y_{st}})_{Ne} = \sqrt{V(\overline{y_{st}})_{Ne}}$$

- Αρίστη κατανομή με διαφορετικό κόστος ανά μονάδα:

$$n_i = n \cdot \frac{W_m S_m / \sqrt{c_m}}{\sum_{m=1}^M (W_m S_m / \sqrt{c_m})}$$

$$V(\overline{y_{st}})_{opt} = \frac{(\sum_{m=1}^M W_m S_m \sqrt{c_m}) \cdot (\sum_{m=1}^M W_m S_m / \sqrt{c_m})}{n} - \frac{\sum_{m=1}^M W_m S_m^2}{N}$$

$$SE(\overline{y_{st}})_{opt} = \sqrt{V(\overline{y_{st}})_{opt}}.$$

Εκτίμηση συνολικού πληθυσμού

A) Αυθαίρετη κατανομή στο δείγμα:

$$V(\hat{Y}) = V(N\overline{y_{st}}) = N^2 \cdot V(\overline{y_{st}}) = N^2 \sum_{m=1}^M W_m^2 \cdot \frac{1-f_m}{n_m} \cdot S_m^2$$

B) Αναλογική κατανομή στο δείγμα:

$$V(\hat{Y}) = V(N\overline{y_{st,p}}) = N^2 \cdot V(\overline{y_{st,p}}) = N^2 \frac{1}{n} \left(\frac{N-n}{N} \right) \sum_{m=1}^M W_m S_m^2$$

Γ) Κατανομή Neyman:

- Αρίστη κατανομή με σταθερό κόστος ανά μονάδα:

$$V(\hat{Y}) = V(N \cdot \overline{y_{st}})_{Ne} = N^2 \left[\frac{(\sum_{m=1}^M W_m S_m)^2}{n} - \frac{\sum_{m=1}^M W_m S_m^2}{N} \right]$$

- Αρίστη κατανομή με διαφορετικό κόστος ανά μονάδα:

$$V(\hat{Y}) = V(N \cdot \overline{y_{st}})_{opt} = N^2 \cdot \left[\frac{(\sum_{m=1}^M W_m S_m \sqrt{c_m}) \cdot (\sum_{m=1}^M W_m S_m / \sqrt{c_m})}{n} - \frac{\sum_{m=1}^M W_m S_m^2}{N} \right]$$

Μέγεθος δείγματος n για την εκτίμηση μέσου πληθυσμού με μέγιστο σφάλμα d:

A) Αυθαίρετη κατανομή στο δείγμα:

$$n = \frac{\sum_{m=1}^M W_m^2 S_m^2 / w_m}{(\frac{d}{z_{\alpha/2}})^2 + \frac{1}{N} \sum_{m=1}^M W_m S_m^2}, \text{ όπου } W_m = \frac{N_m}{N} \text{ και } w_m = \frac{n_m}{n}, m=1,2,\dots,M$$

Β) Αναλογική κατανομή στο δείγμα:

$$n = \frac{N}{1 + \left(\frac{d}{z_{\alpha/2}}\right)^2 \cdot \frac{1}{\sum_{m=1}^M W_m S_m^2} \cdot N}.$$

Γ) Κατανομή Neyman:

- Αρίστη κατανομή με σταθερό κόστος ανά μονάδα:

$$n = \frac{\left(\sum_{m=1}^M W_m S_m\right)^2}{\left(\frac{d}{z_{\alpha/2}}\right)^2 + \frac{1}{N} \cdot \sum_{m=1}^M W_m S_m^2}$$

- Αρίστη κατανομή με διαφορετικό κόστος ανά μονάδα:

$$n = \frac{\left(\sum_{m=1}^M W_m S_m \sqrt{c_m}\right) \cdot \left(\sum_{m=1}^M W_m S_m / \sqrt{c_m}\right)}{\left(\frac{d}{z_{\alpha/2}}\right)^2 + \frac{1}{N} \sum_{m=1}^M W_m S_m^2}$$

Μέγεθος δείγματος n για την εκτίμηση συνολικού πληθυσμού με μέγιστο σφάλμα d:

Α) Αυθαίρετη κατανομή στο δείγμα:

$$n = \frac{\sum_{m=1}^M N_m^2 S_m^2 / w_i}{\left(\frac{d}{z_{\alpha/2}}\right)^2 + \sum_{m=1}^M N_m S_m^2}$$

Β) Αναλογική κατανομή στο δείγμα:

$$n = \frac{N \cdot \sum_{m=1}^M N_m S_m^2}{\left(\frac{d}{z_{\alpha/2}}\right)^2 + \sum_{m=1}^M N_m S_m^2}$$

Γ) Κατανομή Neyman:

- Αρίστη κατανομή με σταθερό κόστος ανά μονάδα:

$$n = \frac{(\sum_{m=1}^M N_m S_m)^2}{(\frac{d}{z_{\alpha/2}})^2 + \sum_{m=1}^M N_m S_m^2}$$

- Αρίστη κατανομή με διαφορετικό κόστος ανά μονάδα:

$$n = \frac{(\sum_{m=1}^M W_m S_m \sqrt{c_m}) \cdot (\sum_{m=1}^M W_m S_m / \sqrt{c_m})}{(\frac{d}{z_{\alpha/2}})^2 + \sum_{m=1}^M W_m S_m^2}$$

Συνάρτηση κόστους : $C = c_0 + \sum_{m=1}^M c_m \cdot n_m$, όπου

c_0 : Γενικά έξοδα

c_m : Κόστος ανά μονάδα στο m-οστό στρώμα $m=1,2,\dots,M$

n_m : Μέγεθος του m-οστού στρώματος στο δείγμα

$$C - c_0 = \sum_{m=1}^M c_m \cdot n_m = C_1$$

$$V(\overline{y_{st}}) = \sum_{m=1}^M W_m^2 \cdot \frac{1-f_m}{n_m} \cdot S_m^2$$

Βέλτιστη επιλογή δείγματος ανά στρώμα (n_m)

$$n_m = n \cdot \frac{\frac{W_m \cdot S_m}{\sqrt{c_m}}}{\sum_{m=1}^M (\frac{W_m \cdot S_m}{\sqrt{c_m}})}, \text{ όπου:}$$

$$i) \ n = \frac{C_1 \cdot \sum_{m=1}^M (W_m \cdot S_m / \sqrt{c_m})}{\sum_{m=1}^M (W_m \cdot S_m \cdot \sqrt{c_m})}, \text{ για σταθερό κόστος } C \text{ και ελάχιστη διασπορά } V(\overline{y_{st}}).$$

$$\text{ii) } n = \frac{\{\sum_{m=1}^M (W_m \cdot S_m / \sqrt{c_m})\} \cdot \{(\sum_{m=1}^M W_m \cdot S_m \cdot \sqrt{c_m})\}}{V(\overline{y_{st}}) + \frac{1}{N} \cdot \sum_{m=1}^M W_m \cdot S_m^2}, \text{ όταν το } V(\overline{y_{st}}) \text{ είναι γνωστό και}$$

σταθερό και θέλουμε ελαχιστοποίηση του κόστους C.

Παρατήρηση: Αν για τουλάχιστον ένα, αλλά λιγότερο από M στρώματα συμβεί να είναι $n_i > N_i$, τότε παίρνουμε για τις αντίστοιχες τιμές του δείκτη $n_i = N_i$ και τα υπόλοιπα μεγέθη στρωμάτων στο δείγμα τα παίρνουμε με βάση την παρακάτω σχέση:

$$n_j = (n - m) \cdot \frac{\frac{W_j \cdot S_j}{\sqrt{c_j}}}{\sum_{i=1}^M \left(\frac{W_i \cdot S_i}{\sqrt{c_i}} \right)}, \text{ όπου } m = \text{άθροισμα των μεγεθών των στρωμάτων, για τα}$$

οποία είχαμε $n_i > N_i$.

Εκτίμηση ποσοστών

Για τον πληθυσμό:

$$\overline{Y} = P, \overline{Y_m} = P_m, S_m^2 = \frac{N_m \cdot P_m \cdot (1 - P_m)}{N_m - 1}, P_{st} = \sum_{m=1}^M \frac{N_m P_m}{N}$$

Για το δείγμα:

$$\overline{y_{st,p}} = p_{st,p}, \overline{y_m} = p_m, S_m^2 = \frac{n_m \cdot p_m \cdot (1 - p_m)}{n_m - 1}, P_{st,p} = \sum_{m=1}^M \frac{N_m p_m}{N}$$

Διασπορά ποσοστού στο δείγμα:

A) Για αυθαίρετη κατανομή του δείγματος στα διάφορα στρώματα η διακύμανση της εκτίμησης του p_{st} θα είναι:

$$V(p_{st}) = \sum_{m=1}^M W_m^2 \cdot \frac{N_m - n_m}{N_m - 1} \cdot \frac{P_m(1 - P_m)}{n_m}$$

$$\text{Όταν } \frac{1}{N_m} \rightarrow 0 \text{ (για το μικρότερο } N_m) \text{ τότε: } V(p_{st}) = \sum_{m=1}^M W_m^2 \cdot \frac{N_m - n_m}{N_m} \cdot \frac{P_m(1 - P_m)}{n_m}$$

Επίσης όταν $f_m \rightarrow 0$, τότε: $V(p_{st}) = \sum_{m=1}^M W_m^2 \frac{P_m(1-P_m)}{n}$

$$SE(p_{st}) = \sqrt{V(p_{st})}$$

Β) Για αναλογική κατανομή του δείγματος στα διάφορα στρώματα η διακύμανση της εκτίμησης του p_{st} θα είναι:

$$V(p_{st})_{pr} = V(p_{st,p}) = \frac{1-f}{n \cdot N} \sum_{m=1}^M N_m^2 \cdot \frac{P_m(1-P_m)}{N_m-1}$$

Όταν $\frac{1}{N_m} \rightarrow 0$, τότε: $V(p_{st,p}) = \frac{1-f}{n \cdot N} \sum_{m=1}^M W_m \cdot P_m \cdot (1-P_m)$

$$SE(p_{st,p}) = \sqrt{V(p_{st,p})}$$

Γ) Στην περίπτωση αρίστης κατανομής έχουμε:

i)Κατανομή Neyman (με σταθερό κόστος ανά μονάδα):

$$V(p_{st,Ne}) = \sum_{m=1}^M W_m^2 \cdot \frac{S_m^2}{n_m} \cdot \left(\frac{N_m - n_m}{N_m} \right),$$

$$\text{όπου: } S_m^2 = \frac{N_m P_m (1-P_m)}{N_m - 1} \text{ και } n_m = n \cdot \frac{N_m \cdot \sqrt{P_m (1-P_m)}}{\sum_{m=1}^M N_m \cdot \sqrt{P_m (1-P_m)}}$$

ii)Διαφορετικό κόστος ανά μονάδα:

$$V(p_{st})_{opt} = \sum_{m=1}^M W_m^2 \cdot \frac{S_m^2}{n_m} \cdot \left(\frac{N_m - n_m}{N_m} \right), \text{ όπου: } n_m = n \cdot \frac{N_m \cdot \sqrt{P_m (1-P_m) / c_m}}{\sum_{m=1}^M N_m \cdot \sqrt{P_m (1-P_m) / c_m}}$$

$$\text{Ιχύει ότι: } \frac{P_m \hat{Q}_m}{n_m} = \frac{p_m q_m}{n_m - 1}, \text{ όπου } Q_m = 1 - P_m$$

Βέλτιστη επιλογή δείγματος ανά στρώμα (n_m) για ποσοστά:

- Για σταθερό κόστος $C = c_0 + \sum_{m=1}^M c_m \cdot n_m$, τα μεγέθη είναι:

$$n_m \cong n \cdot \frac{N_m \sqrt{P_m (1-P_m) / c_m}}{\sum_{m=1}^M N_m \sqrt{P_m (1-P_m) / c_m}}$$

- Για σταθερό μέγεθος του δείγματος n είναι:

$$n_m \cong n \cdot \frac{N_m \sqrt{P_m(1-P_m)}}{\sum_{m=1}^M N_m \sqrt{P_m(1-P_m)}}$$

- Για γνωστή και σταθερή διασπορά $V(p_{st})=V$ και όταν θέλουμε να ελαχιστοποιήσουμε το κόστος C , έχουμε:

$$n_m \cong n \cdot \frac{W_m \cdot \sqrt{\frac{N_m \cdot P_m \cdot (1-P_m)}{(N_m-1) \cdot c_m}}}{\sum_{m=1}^M W_m \cdot \sqrt{\frac{N_m \cdot P_m \cdot (1-P_m)}{(N_m-1) \cdot c_m}}}$$

και

$$n \cong \frac{(\sum_{m=1}^M W_m \cdot \sqrt{\frac{N_m \cdot P_m \cdot (1-P_m) \cdot c_m}{N_m-1}}) \cdot (\sum_{m=1}^M W_m \cdot \sqrt{\frac{N_m \cdot P_m \cdot (1-P_m)}{(N_m-1) \cdot c_m}})}{V + \frac{1}{N} \cdot \sum_{m=1}^M W_m \cdot \frac{N_m \cdot P_m \cdot (1-P_m)}{N_m-1}}$$

και στην περίπτωση, που το $\max \frac{1}{N_m}$ θεωρηθεί αμελητέο, τουλάχιστον ως προς τη μονάδα, έχουμε:

$$n \cong \frac{(\sum_{m=1}^M W_m \cdot \sqrt{P_m \cdot (1-P_m) \cdot c_m}) \cdot (\sum_{m=1}^M W_m \cdot \sqrt{\frac{P_m \cdot (1-P_m)}{c_m}})}{V + \frac{1}{N} \sum_{m=1}^M W_m \cdot P_m \cdot (1-P_m)}$$

και άρα:

$$n_m \cong n \cdot \frac{W_m \cdot \sqrt{\frac{P_m \cdot (1-P_m)}{c_m}}}{\sum_{m=1}^M W_m \cdot \sqrt{\frac{P_m \cdot (1-P_m)}{c_m}}}$$

Στις παραπάνω σχέσεις το $\frac{N_m}{N_m - 1} \cdot P_m \cdot (1 - P_m)$ αντικαθίσταται από τον αμερόληπτο εκτιμητή του, το $\frac{n_m}{n_m - 1} \cdot p_m \cdot (1 - p_m)$.

ΣΥΣΤΗΜΑΤΙΚΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ (systematic sampling theory)

Y: Αθροισμα του πληθυσμού.

$\bar{Y}$: Μέση τιμή του πληθυσμού.

S^2 : Διασπορά του πληθυσμού.

S: Τυπική απόκλιση του πληθυσμού.

$y_{sy,j}$: Αθροισμα του j-οστού δείγματος, $j=1,2,3,\dots,k$, όπου k το πλήθος των πιθανών δειγμάτων.

$\overline{y_{sy,j}}$: Μέση τιμή του j-οστού δείγματος, $j=1,2,3,\dots,k$

$\overline{y_{sy}}$: Δειγματική μέση τιμή.

$s_{sy,j}^2$: Διασπορά του j-οστού δείγματος, $j=1,2,3,\dots,k$

$s_{sy,j}$: Τυπική απόκλιση του j-οστού δείγματος, $j=1,2,3,\dots,k$

$\text{var } \overline{y_{sy}} = V(\overline{y_{sy}})$: Διασπορά της δειγματικής μέσης τιμής.

y_{ij} : Το j-στό στοιχείο του i-οστού δείγματος, $j=1,2,\dots,n$, $i=1,2,\dots,k$.

S_{wsy}^2 : Ο μέσος όρος των k διακυμάνσεων.

$$\bar{Y} = \frac{1}{N} \cdot \sum_{i=1}^N Y_i = \frac{1}{n \cdot k} \cdot \sum_{i=1}^k \sum_{j=1}^n y_{ij}$$

$$S^2 = \frac{1}{N-1} \cdot \sum_{i=1}^N (Y_i - \bar{Y})^2 = \frac{1}{N-1} \cdot \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{Y})^2$$

$$\overline{y_{sy,i}} = \frac{1}{n} \cdot \sum_{j=1}^n y_{ij} = \frac{1}{n} \cdot \sum_{r=1}^n Y_{i+(r-1)k}, \quad i=1,2,\dots,k$$

$$\overline{y_{sy}} = \frac{1}{k} \sum_{i=1}^k \overline{y_{sy,i}}$$

$$S_{\text{wsy}}^2 = \frac{1}{k(n-1)} \cdot \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \overline{y_j})^2 = \frac{1}{k} \cdot \sum_{j=1}^k s_{\text{sy},j}^2$$

$$s_{\text{sy},i}^2 = \frac{1}{n-1} \cdot \sum_{j=1}^n (y_{ij} - \overline{y_{\text{sy},j}})^2, \quad i=1,2,\dots,k$$

$$\text{var } \overline{y_{\text{sy}}} = V(\overline{y_{\text{sy}}}) = \frac{N-1}{N} \cdot S^2 - \frac{k(n-1)}{N} \cdot S_{\text{wsy}}^2$$

Αριθμητική πρόοδος

$$a, b \in \mathfrak{R} \text{ και } Y_{ij} = Y_{(i-1)k+j} = a + ((i-1) \cdot k + j - 1) \cdot b$$

$Y_1 = y_{11} = a$ και $Y_u = y_{(i-1)k+j} = y_{ij} = a + (u-1) \cdot b$, όπου $i=[u/k]+1$, $j=u-(i-1) \cdot k$ και $[x]$ ο μεγαλύτερος ακέραιος, που είναι μικρότερος του $x \in \mathfrak{R}$, $u=1,2,\dots,n$.

$$\overline{y_j} = \frac{1}{n} \cdot \sum_{i=1}^n y_{ij} = \frac{1}{n} \cdot \sum_{i=1}^n y_{(i-1)k+j} : \text{μέση τιμή για το } j\text{-οστό } n\text{-μελές δείγμα}$$

$$s_j^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (y_{ij} - \overline{y_j})^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (y_{(i-1)k+j} - \overline{y_j})^2 : \text{διασπορά στο αντίστοιχο δείγμα.}$$

$$\overline{Y} = a + (N-1) \cdot \frac{b}{2}, \quad S^2 = \frac{b^2}{12} \cdot N \cdot (N+1), \quad S_{\text{wsy}}^2 = \frac{b^2}{12} \cdot k \cdot N \cdot (n+1)$$

$$\overline{y_j} = a + \frac{b}{2} \cdot \{k \cdot (n-1) + 2j - 2\}, \quad s_j^2 = \frac{b^2 \cdot k^2 \cdot n \cdot (n+1)}{12}$$

$$\text{var } \overline{y_{\text{sy}}} = \frac{b^2}{12} \cdot (k^2 - 1)$$

Γεωμετρική πρόοδος

$$g \in \mathfrak{R} - \{0\}, \quad b \in (0,1) \cup (1,\infty) \text{ και } Y_{ij} = Y_{(i-1)k+j} = g \cdot b^{(i-1)k+j-1}$$

$Y_1 = y_{11} = g$ και $Y_u = y_{(i-1)k+j} = y_{ij} = g \cdot b^{u-1}$, όπου $i=[u/k]+1$, $j=u-(i-1) \cdot k$ και $[x]$ ο μεγαλύτερος ακέραιος, που είναι μικρότερος του $x \in \mathfrak{R}$, $u=1,2,\dots,n$.

$$\bar{Y} = \frac{1}{N} \cdot g \cdot \frac{b^N - 1}{b - 1}, S^2 = \frac{g^2}{N \cdot (N - 1)} \cdot \frac{b^N - 1}{b^2 - 1} \cdot \{(b^N + 1) \cdot (N - 1) - 2 \cdot \sum_{i=1}^{N-1} b^i\}$$

$$\overline{y_j} = \frac{g \cdot b^{j-1}}{n} \cdot \frac{b^N - 1}{b^k - 1}$$

$$s_j^2 = \frac{1}{n(n-1)} \cdot \frac{g^2 \cdot (b^2)^{j-1} (b^N - 1)}{(b^k - 1)^2 \cdot (b^k + 1)} \cdot \{(n-1) \cdot (b^{N+k} - 1) - (n+1) \cdot b^k \cdot (b^n - 1)\}$$

$$S_{\text{wsy}}^2 = \frac{1}{N(n-1)} \cdot \frac{g^2 \cdot (b^N - 1)}{(b^k - 1) \cdot (b^2 - 1)} \cdot \{(n-1) \cdot (b^{N+k} - 1) - (n+1) \cdot b^k \cdot (b^n - 1)\}$$

$$\overline{\text{Var} y_{sy}} = \frac{g^2 \cdot (b^N - 1)}{N^2 \cdot (b^k - 1) \cdot (b^2 - 1)} \cdot \{k \cdot (b^N - 1) \cdot (b^k + 1) - (b^N + 1) \cdot (b^k - 1) - 2 \cdot (\sum_{i=1}^{N-1} b^i) \cdot (b^k - 1)\}$$

Για γρήγορη εκτίμηση της μέσης τιμής $\bar{Y}$, αλλά με κάποιους κινδύνους για την ακρίβεια και την αμεροληψία, όταν οι τιμές τις τ.μ. $\bar{Y}$ αποτελούν όρους αριθμητικής ή γεωμετρικής προόδου έχουμε το ακόλουθο τέχνασμα:

A) Αριθμητική πρόοδος:

Παίρνουμε τη διαφορά: $d_j = \overline{y_j} - \bar{Y} = \frac{b}{2} \cdot (2j - k - 1)$ και επιδιώκουμε το μηδενισμό της, αν είναι δυνατόν. Λύνουμε δηλαδή την εξίσωση:

$$d_j = 0 \Leftrightarrow j = \frac{k+1}{2}$$

Αν $k \equiv 1 \pmod{2} \Rightarrow \hat{\bar{Y}} = \overline{y_j}$, ενώ αν $k \equiv 0 \pmod{2} \Rightarrow j = [(k+1)/2]$ και $\hat{\bar{Y}} = \frac{1}{2} \cdot (\overline{y_j} + \overline{y_{j+1}})$.

B) Γεωμετρική πρόοδος:

Παίρνουμε και πάλι τη διαφορά: $d_j = \overline{y_j} - \bar{Y} = \frac{b}{2} \cdot (2j - k - 1)$ και επιδιώκουμε το μηδενισμό της, αν είναι δυνατόν. Λύνουμε δηλαδή την εξίσωση:

$$d_j = 0 \Leftrightarrow j = 1 + \frac{\ln(\frac{b^k - 1}{k(b-1)})}{\ln b}$$

Η παραπάνω λύση δεν είναι γενικά ακέραιος αριθμός. Παίρνουμε σαν τιμή του δείκτη j το ακέραιο μέρος της λύσης, δηλαδή:

$$j = 1 + \left[\frac{\ln\left(\frac{b^k - 1}{k(b-1)}\right)}{\ln b} \right]$$

Ενεργούμε μετά ανάλογα με το αν το αποκοπτόμενο (μη ακέραιο) τμήμα του j είναι ίσο ή όχι με το 0.5. Ειδικά, όταν το μη ακέραιο μέρος είναι $\varepsilon=0.5$, έχουμε:

$$\hat{\bar{Y}} = \bar{y}_{sy} = \frac{\bar{y}_{sy} + \bar{y}_{j+1}}{2} \quad \text{και η τιμή του } j \text{ υπολογίζεται από τη σχέση}$$

$$j = 1 + \left[\frac{\ln\left(\frac{b^k - 1}{k(b-1)}\right)}{\ln b} \right]. \text{ Αν όμως είναι } \varepsilon \in (0,1) - \{0.5\}, \text{ τότε έχουμε:}$$

$$\hat{\bar{Y}} = \bar{y}_{sy} = (1 - \varepsilon) \cdot \bar{y}_j + \varepsilon \cdot \bar{y}_{j+1}$$

ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΚΑΤΑ ΣΥΣΤΑΔΕΣ (cluster sampling) (Ισων μεγεθών)

N : Συμβολισμός συστάδων στον πληθυσμό

m : Συμβολισμός στοιχείων σε κάθε συστάδα

M : Συμβολισμός στοιχείων σε όλο τον πληθυσμό

Y_{ij} : Η τιμή της τ.μ. Y στο στοιχείο j της i -οστής συστάδας του πληθυσμού,
 $i=1,2,\dots,N \quad j=1,2,\dots,m$

Y_i : Άθροισμα της i -οστής συστάδας του πληθυσμού

Y : Άθροισμα όλου του πληθυσμού

$\bar{Y}_i$: Μέση τιμή της i -οστής συστάδας του πληθυσμού

$\bar{\bar{Y}}$: Μέση τιμή του πληθυσμού

$\bar{Y}$: Μέση τιμή των συστάδων

$$M = m \cdot N$$

$$Y_i = \sum_{j=1}^m Y_{ij}, i = 1, 2, \dots, N$$

$$Y = \sum_{i=1}^N Y_i = \sum_{i=1}^N \sum_{j=1}^m Y_{ij}$$

$$\overline{Y}_i = \frac{1}{m} \cdot \sum_{j=1}^m Y_{ij}, i = 1, 2, \dots, N$$

$$\overline{Y} = \frac{1}{N} \cdot \sum_{i=1}^N Y_i = \frac{1}{N} \cdot \sum_{i=1}^N \sum_{j=1}^m Y_{ij}$$

$$\overline{\overline{Y}} = \overline{Y} / m = \frac{1}{M} \cdot \sum_{i=1}^N \sum_{j=1}^m Y_{ij}$$

$$\overline{\overline{Y}} = Y / M = \frac{1}{M} \cdot \sum_{i=1}^N Y_i$$

n: Συμβολισμός συστάδων στο δείγμα

m: Συμβολισμός στοιχείων σε κάθε συστάδα

y_{ij} : Η τιμή της τ.μ. y στο στοιχείο j της i -οστής συστάδας του δείγματος,
 $i=1, 2, \dots, n, j=1, 2, \dots, m$

y_i : Άθροισμα της i -οστής συστάδας του δείγματος

y : Άθροισμα όλου του δείγματος

$\overline{y}_i$: Μέση τιμή της i -οστής συστάδας του δείγματος

$\overline{\overline{y}}$: Μέση τιμή του δείγματος

$\overline{y}$: Μέση τιμή των συστάδων

$$y_i = \sum_{j=1}^m y_{ij}, i = 1, 2, \dots, n$$

$$y = \sum_{i=1}^n y_i = \sum_{i=1}^n \sum_{j=1}^m y_{ij}$$

$$\overline{y}_i = \frac{1}{m} \cdot \sum_{j=1}^m y_{ij}, i = 1, 2, \dots, n$$

$$\overline{y} = \frac{1}{n} \cdot \sum_{i=1}^n y_i = \frac{1}{n} \cdot \sum_{i=1}^n \sum_{j=1}^m y_{ij}$$

$$\overline{\overline{y}} = \overline{y} / m = \frac{1}{m \cdot n} \cdot \sum_{i=1}^n \sum_{j=1}^m y_{ij}$$

$$\overline{\overline{y}} = \frac{y}{m \cdot n} = \frac{1}{m \cdot n} \cdot \sum_{i=1}^n y_i$$

$$E \overline{\overline{y}} = \overline{\overline{Y}}$$

$$\text{var } \overline{\overline{y}} = \frac{1-f}{n} \cdot S_1^2, \text{ όπου: } S_1^2 = \frac{1}{N-1} \cdot \sum_{i=1}^N (\overline{Y}_i - \overline{\overline{Y}})^2$$

$$Es_1^2 = S_1^2, \text{ οπότε: } \text{var } \overline{\overline{y}} = \frac{1-f}{n} \cdot s_1^2$$

$$100(1-\alpha)\% \text{ δ.ε.της } \bar{\bar{Y}} = \{\bar{y} \pm t_{n-1; \alpha/2} \cdot \sqrt{\hat{\text{var}} \bar{y}}\}$$

$$100(1-\alpha)\% \text{ δ.ε.της } Y = M \cdot \{\bar{y} \pm t_{n-1; \alpha/2} \cdot \sqrt{\hat{\text{var}} \bar{y}}\}$$

$$n \geq \frac{N}{1 + \left(\frac{d}{t_{n-1; \alpha/2} \cdot s_1}\right)^2 \cdot N}, \text{ για την εκτίμηση του } \bar{\bar{Y}} \text{ και } n \geq \frac{N}{1 + \left(\frac{h}{M \cdot t_{n-1; \alpha/2} \cdot s_1}\right)^2 \cdot N},$$

για την εκτίμηση του Y, όπου d και h μέγιστα σφάλματα.

Εκτίμηση ποσοστών:

Για πληθυσμό:

$$P_i = \bar{Y}_i = \frac{W_i}{m}$$

$$W_i = m \cdot P_i = Y_i$$

$$W = m \cdot \sum_{i=1}^N P_i = Y$$

$$P = \frac{W}{M} = \bar{\bar{Y}}$$

$$\frac{W}{N} = \frac{m}{N} \cdot \sum_{i=1}^N P_i = \bar{Y}$$

$$S_1^2 = \frac{1}{N-1} \cdot \sum_{i=1}^N (P_i - P)^2$$

Για δείγμα:

$$p_i = y_i$$

$$w_i = m \cdot p_i = y_i$$

$$w = m \cdot \sum_{i=1}^n p_i = y$$

$$p = \bar{y} = \frac{1}{n} \sum_{i=1}^n p_i$$

$$\frac{w}{n} = \frac{m}{n} \cdot \sum_{i=1}^n p_i = \bar{y}$$

$$s_1^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (p_i - p)^2$$

$$Ep = P$$

$$\text{var } p = \frac{1-f}{n} \cdot S_1^2$$

$$Es_1^2 = S_1^2, \text{ οπότε: } \hat{\text{var}} p = \frac{1-f}{n} \cdot s_1^2$$

$$100(1-\alpha)\% \text{ δ.ε.της } P = \{p \pm t_{n-1;\alpha/2} \cdot \sqrt{\widehat{\text{var } p}}\}$$

$$100(1-\alpha)\% \text{ δ.ε.της } W = M \cdot \{p \pm t_{n-1;\alpha/2} \cdot \sqrt{\widehat{\text{var } p}}\}$$

$$n \geq \frac{N}{1 + \left(\frac{d}{t_{n-1;\alpha/2} \cdot s_1}\right)^2 \cdot N}, \text{ για την εκτίμηση του } P \text{ και } n \geq \frac{N}{1 + \left(\frac{h}{M \cdot t_{n-1;\alpha/2} \cdot s_1}\right)^2 \cdot N},$$

για την εκτίμηση του W, όπου d και h μέγιστα σφάλματα.