ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

ΣΤΑΤΙΣΤΙΚΕΣ ΤΕΧΝΙΚΕΣ ΓΙΑ ΑΝΑΛΥΣΗ ΨΗΦΟΦΟΡΙΩΝ

Θεοδώρα-Παρασκευή Γ. Σταμπέλου

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική

Πειραιάς
Σεπτέμβριος 2014

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

ΣΤΑΤΙΣΤΙΚΕΣ ΤΕΧΝΙΚΕΣ ΓΙΑ ΑΝΑΛΥΣΗ ΨΗΦΟΦΟΡΙΩΝ

Θεοδώρα-Παρασκευή Γ. Σταμπέλου

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική

Πειραιάς
Σεπτέμβριος 2014

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίσθηκε από τη ΓΣΕΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. …….. συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική

Τα μέλη της Επιτροπής ήταν:

- Καθηγητής, Μάρκος Κούτρας (Επιβλέπων)
- Καθηγητής, Κλέωνας Τσίμπος
- Επίκουρος Καθηγητής, Γεώργιος Τζαβελάς

Η έγκριση της Διπλωματική Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα.

UNIVERSITY OF PIRAEUS

DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE

POSTGRADUATE PROGRAM IN
APPLIED STATISTICS

STATISTICAL TECHNIQUES FOR VOTING ANALYSIS

By

Theodora-Paraskevi G. Stampelou

MSc Dissertation

submitted to the Department of Statistics and Insurance Science of the University of Piraeus in partial fulfilment of the requirements for the degree of Master of Science in Applied Statistics

Piraeus, Greece
September 2014

Στους γονείς μου

Γιώργο και Μαρία

Ευχαριστίες

Θα ήθελα αρχικά να ευχαριστήσω θερμά τον καθηγητή μου κ. Μάρκο Κούτρα για την επίβλεψη της παρούσας εργασίας, για την συνεχή καθοδήγηση και την εμπιστοσύνη που μου επέδειξε.

Ιδιαίτερα ευχαριστώ στον καθηγητή κ. Κλέωνα Τσίμπο και στον επίκουρο καθηγητή κ. Γεώργιο Τζαβελά για τη συμμετοχή τους στην τριμελή επιτροπή μου αλλά και για τις πολύτιμες γνώσεις που μου παρείχαν καθ’ όλη τη διάρκεια των προπτυχιακών και μεταπτυχιακών σπουδών μου.

Ευχαριστώ τη Μαρία και τη Ματίνα για το ότι ήταν “πάντα εκεί” και έκαναν τα χρόνια αυτά μία αξέχαστη εμπειρία.

Τέλος, το μεγαλύτερο ευχαριστώ το οφείλω σε δύο ανθρώπους που όλα αυτά τα χρόνια είναι δίπλα μου, με συμβουλεύουν και με στηρίζουν. Σε κάθε μου όνειρο, σε κάθε μου βήμα, σε κάθε μου προσπάθεια, σε κάθε μου επιτυχία ή και αποτυχία… Στους γονείς μου…

Σταμπέλου Θεοδώρα Παρασκευή

Πειραιάς, Σεπτέμβριος 2014

Περίληψη

Στις μέρες μας, στην εποχή της οικονομικής κρίσης, των ραγδαίων πολιτικών εξελίξεων και της γενικευμένης πολιτικής ρευστότητας που επικρατεί παγκοσμίως, το τελικό αποτέλεσμα κάθε εκλογικής αναμέτρησης δε μπορεί σε καμία περίπτωση να θεωρηθεί εκ των προτέρων γνωστό. Το γεγονός αυτό, αποτελεί τη μεγαλύτερη πρόκληση για τις εταιρείες δημοσκοπήσεων οι οποίες πριν από κάθε εκλογική αναμέτρηση επενδύουν τεράστια χρηματικά ποσά για την διεξαγωγή του *exit poll* έχοντας ως απώτερο στόχο την αξιόπιστη και όσο το δυνατόν ακριβέστερη πρόβλεψη του εκλογικού αποτελέσματος. Σημαντικό εργαλείο γι' αυτό, αποτελούν κάποια στατιστικά μοντέλα τα οποία χρησιμοποιούνται από τους αναλυτές για την πρόβλεψη του αποτελέσματος.

Κύριο αντικείμενο της παρούσας διπλωματικής εργασίας είναι να περιγράψει τη διαδικασία κατασκευής τέτοιων μοντέλων πρόβλεψης, καθώς και τις στατιστικές τεχνικές ανάλυσης των αποτελεσμάτων του exit poll που, πέραν της πρόβλεψης του εκλογικού αποτελέσματος, παράγουν επιπλέον πληροφορίες σχετικά με τα κοινωνικά χαρακτηριστικά του εκλογικού σώματος.

Επιπλέον, θα επιδιώξουμε να προβλέψουμε τα τελικά εκλογικά αποτελέσματα με μία τεχνική διαφορετική από εκείνη του exit poll και λιγότερο διαδεδομένη στον κύκλο των εταιρειών δημοσκοπήσεων, την τεχνική της *προσομοίωσης*.

Abstract

In these days of economic crisis, rapid political changes and global political fluidity, the final outcome of every voting, cannot be considered known beforehand. This fact constitutes the biggest challenge for Public Opinion Research companies, which, before every general election invest huge amounts of money on carrying out exit polls, aiming at a considerably precise and reliable prediction of the result. An important tool to achieve that, is a number of statistical models used by analysts in order to predict the outcome.

The main objective of this thesis is to describe the process of building such prediction models, as well as the statistical techniques of analyzing the exit poll's results which, except for predicting the outcome, provide more information about the social characteristics of the electoral body.

Moreover, we will illustrate a prediction method different to that of the exit polls and less widespread in the public opinion research companies, the method of simulation.

Περιεχόμενα

Κατάλογος Πινάκων

Κατάλογος Σχημάτων

Κατάλογος Συντομογραφιών

Η.Π.Α	Ηνωμένες Πολιτείες Αμερικής
ΝΔ	Νέα Δημοκρατία
ΣΥ.ΡΙΖ.Α	Συνασπισμός Ριζοσπαστικής Αριστεράς
ΠΑΣΟΚ	Πανελλήνιο Σοσιαλιστικό Κίνημα
Χ.Α	Χρυσή Αυγή
Κ.Κ.Ε	Κομμουνιστικό Κόμμα Ελλάδας
OR	Odds Ratio
df	degrees of freedom
ε.μ.π	εκτιμητές μέγιστης πιθανοφάνειας
ε.σ	επίπεδο σημαντικότητας
τ.μ	τυχαία μεταβλητή
δ.ε	διάστημα εμπιστοσύνης
MSE	Mean Squared Error
AR	Auto Regressive
ARIMA	Auto Regressive Integrated Moving Average
SST	Total Sum of Squares
SSR	Regression Sum of Squares
SSE	Error Sum of Squares
MST	Mean Square of Treatments
MSR	Regression Mean Square
MSE	Error Mean Square
AIC	Akaike’s Information Criterion
BIC	Bayesian Information Criterion

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

1.1 Ιστορική αναδρομή

Η έννοια της ψήφου συναντάται για πρώτη φορά στην αρχαία Αθήνα τον 6ο αιώνα π.Χ. όταν ο Κλεισθένης καθιέρωσε το *θεσμό του οστρακισμού* για την προστασία του δημοκρατικού πολιτεύματος. Μία φορά το χρόνο οι Αθηναίοι πολίτες, ύστερα από απόφαση της εκκλησίας του Δήμου, συγκεντρώνονταν στην Αγορά και ψήφιζαν εναντίον του πολίτη εκείνου που λόγω υπερβολικά μεγάλης δύναμης θεωρούνταν επικίνδυνος για την κατάλυση του δημοκρατικού πολιτεύματος και έπρεπε να εξοριστεί από την πόλη για δέκα χρόνια. Ωστόσο, για να εξοριστεί ένας πολίτης έπρεπε να έχει συγκεντρώσει τουλάχιστον 6000 ψήφους. Η ιστορία των εκλογικών συστημάτων ξεκινά την εποχή της Γαλλικής Επανάστασης και συγκεκριμένα το 1770, όταν ο *Jean-Charles de Borda* πρότεινε μια μέθοδο για την εκλογή των μελών της Γαλλικής Ακαδημίας Επιστημών. Στην ιδέα του Borda αντιτάχθηκε ο Marquis de Condorcet ο οποίος πρότεινε μία εναλλακτική μέθοδο, εκείνη των *αντικρουόμενων ζευγών*, θέτοντας έτσι τις βάσεις για τη δημιουργία των δύο πρώτων πλειοψηφικών εκλογικών συστημάτων. Τα αναλογικά συστήματα αρχίζουν να αναπτύσσονται στην Αμερική τον 18ο αιώνα προκειμένου να βρεθεί ο τρόπος με τον οποίο θα κατανέμονταν οι θέσεις της Βουλής των Αντιπροσώπων στις διάφορες πολιτείες ανάλογα με τον πληθυσμό τους, όπως όριζε το σύνταγμα. Στην Ευρώπη, η μελέτη των αναλογικών συστημάτων ξεκινά τον 19ο αιώνα, μετά την εμφάνιση των πρώτων πολιτικών παρατάξεων. Η καθυστέρηση αυτή της Ευρώπης έναντι της Αμερικής έχει ως αποτέλεσμα συστήματα που είναι ίδια να διαθέτουν διπλή ονομασία. Χαρακτηριστικά παραδείγματα αποτελούν τα συστήματα των *d'Hondt, Sainte- Laguë* και *Hare* που αναπτύχθηκαν στην Ευρώπη και είναι ίδια με αυτά των *Jefferson, Webster* και *Hamilton* που αναπτύχθηκαν στην Αμερική. Τα μικτά εκλογικά συστήματα, δημιουργήθηκαν από ένα συνδυασμό αναλογικών και πλειοψηφικών συστημάτων καθώς διαπιστώθηκε ότι κανένα αμιγώς αναλογικό ή πλειοψηφικό σύστημα δε συμβάλει στην ειρηνική διακυβέρνηση από σταθερές κυβερνήσεις.

1.2 Είδη εκλογικών συστημάτων

Με τον όρο εκλογικό σύστημα εννοούμε όλη εκείνη τη διαδικασία, μαθηματικής, τεχνικής και νομικής μορφής, με βάση την οποία καθορίζεται η ανάδειξη αιρετών αντιπροσώπων μέσα από τις διαδικασίες της ψηφοφορίας. Ουσιαστικά, το εκλογικό σύστημα περιγράφει τον τρόπο και τους όρους μετατροπής των ψήφων σε εκλεγμένους αντιπροσώπους. Στην πλειονότητα των χωρών του κόσμου, το εκλογικό σύστημα καθιερώνεται μέσα από μία σειρά κανόνων δικαίου, από τους οποίους άλλοι περιέχονται στο Σύνταγμα και άλλοι στην κοινή -εκλογική- νομοθεσία, παραμένοντας έτσι ενιαίο για όλες τις πολιτικές αναμετρήσεις. Αντίθετα, στην Ελλάδα πολύ σπάνια εφαρμόστηκε το ίδιο εκλογικό σύστημα έστω και σε δύο συνεχόμενες εκλογές, εκτός από τις επαναληπτικές, καθώς δεν υπάρχει συνταγματικά κατοχυρωμένη μόνιμη ρύθμιση που να αφορά την ανάδειξη των βουλευτών και έτσι το εκάστοτε εκλογικό σύστημα καθορίζεται πάντοτε από τον αντίστοιχο εκλογικό νόμο που ψηφίζει η κυβέρνηση που κηρύσσει τις εκλογές. Από το 1844, χρονιά που καθιερώθηκε συνταγματικά η ψηφοφορία στην Ελλάδα, μέχρι και το 1923, οι βουλευτικές εκλογές γίνονταν με το πλειοψηφικό σύστημα και μάλιστα με σφαιρίδια και όχι με ψηφοδέλτια. Από το 1926 έως και το 1956 υπήρξε εναλλαγή πλειοψηφικού και αναλογικού συστήματος, ενώ μετά το 1956, βασικό εκλογικό σύστημα είναι το αναλογικό, σε πολλές παραλλαγές του. Αξίζει να σημειωθεί 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 την πλειοψηφία των δύο τρίτων του όλου αριθμού των βουλευτών. Σήμερα, σχεδόν σε όλες τις δημοκρατικές χώρες εφαρμόζονται αναλογικά ή μεικτά εκλογικά συστήματα με εξαίρεση ορισμένες από τις πρώην βρετανικές αποικίες όπου εφαρμόζονται πλειοψηφικά. Στη συνέχεια, ακολουθεί μία ενδελεχής επισκόπηση στα τρία βασικά είδη εκλογικών συστημάτων και κατ' επέκταση στις επιμέρους μεθόδους που τα απαρτίζουν.

1.2.1 Τα Πλειοψηφικά εκλογικά συστήματα

Γενικός κανόνας των πλειοψηφικών συστημάτων είναι ότι σε κάθε εκλογική περιφέρεια εκλέγεται ο συνδυασμός ή ο υποψήφιος που λαμβάνει την πλειοψηφία των ψήφων (*αρχή της πλειοψηφίας*). Το πλειοψηφικό σύστημα είθισται να εφαρμόζεται σε μονοεδρικές εκλογικές περιφέρειες, καθώς εκείνος που συγκεντρώνει τις περισσότερες ψήφους λαμβάνει και τη μοναδική έδρα. Παρόλα αυτά, είναι δυνατόν να εφαρμοστεί και σε πολυεδρικές εκλογικές περιφέρειες. Στην περίπτωση αυτή, νικητές ανακηρύσσονται εκείνοι που συγκεντρώνουν τις περισσότερες ψήφους και είναι τόσοι όσες και οι έδρες που χρειάζονται να καλυφθούν. Το πλειοψηφικό εκλογικό σύστημα αποτελείται από επιμέρους συστήματα τα οποία και θα αναλύσουμε στη συνέχεια.

α) Σύστημα Σχετικής Πλειοψηφίας

Αποτελεί την απλούστερη μέθοδο πλειοψηφικού συστήματος και είναι παγκοσμίως γνωστό με τη φράση *"the first past the post"* (ο πρώτος πάνω από το όριο). Η συγκεκριμένη φράση προέρχεται από την ορολογία των ιπποδρομιών. Σ' αυτούς τους αγώνες νικητής αναδεικνύεται αυτός που θα διέλθει πρώτος από ένα ορισμένο σημείο της διαδρομής, το οποίο καθιστά όλους τους άλλους αγωνιζόμενους που ακολουθούν ηττημένους. Στο συγκεκριμένο εκλογικό σύστημα οι ψηφοφόροι επιλέγουν έναν μόνο υποψήφιο και νικητής εκλέγεται από τον πρώτο γύρο εκείνος ο οποίος λαμβάνει την πλειοψηφία των ψήφων χωρίς αυτή να χρειάζεται να είναι απόλυτη, δηλαδή μεγαλύτερη από το 50% των ψήφων της περιφέρειας συνολικά.

β) Σύστημα Απόλυτης Πλειοψηφίας

Οι ψηφοφόροι επιλέγουν έναν μόνο υποψήφιο και νικητής εκλέγεται εκείνος ο οποίος συγκεντρώνει την απόλυτη πλειοψηφία των ψήφων. Εάν κανένας από τους υποψηφίους δε λάβει τόσες ψήφους, τότε, όλοι οι υποψήφιοι εκτός από τους δύο που έλαβαν τις περισσότερες ψήφους εξαιρούνται από τη διαδικασία και ακολουθεί δεύτερος γύρος εκλογών. Εκεί, είναι βέβαιο, από τη στιγμή που έχουμε δύο μόνο υποψηφίους $\left(X_1, X_2\right)$, ότι ο ένας από τους δύο θα συγκεντρώσει την απόλυτη πλειοψηφία ($X_1 > 50\%$ ή $X_2 > 50\%$) και θα βγει νικητής. Σε κάποιες παραλλαγές του συστήματος, κάθε υποψήφιος απαιτείται να έχει συγκεντρώσει ένα συγκεκριμένο ποσοστό ψήφων προκειμένου να μπορέσει να συμμετάσχει στο δεύτερο γύρο εκλογών. Στην περίπτωση αυτή, είναι δυνατόν να συναφθούν συμμαχίες μεταξύ δύο ή περισσότερων υποψηφίων με κοινή ιδεολογία με σκοπό να συγκεντρώσουν το ποσοστό που έχει τεθεί ως όριο.

γ) Εξαντλητική Ψήφος

Οι ψηφοφόροι επιλέγουν έναν μόνο υποψήφιο και νικητής εκλέγεται εκείνος ο οποίος λαμβάνει την απόλυτη πλειοψηφία των ψήφων. Σε αντίθετη περίπτωση, ακολουθεί δεύτερος γύρος εκλογών στον οποίο εξαιρείται από τη διαδικασία ο υποψήφιος που έλαβε τις λιγότερες ψήφους στον πρώτο γύρο εκλογών. Η εκλογική διαδικασία επαναλαμβάνεται με επόμενους γύρους εκλογών αποκλείοντας κάθε φορά τον υποψήφιο με τις λιγότερες ψήφους μέχρις ότου κάποιος καταλάβει την απόλυτη πλειοψηφία.

δ) Η Μέθοδος Condorcet

Το 1785 ο Γάλλος μαθηματικός Μαρκήσιος του Condorcet δημοσίευσε μία εργασία με τίτλο "*Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix".* Σκοπός της εργασίας του ήταν να υπολογίσει την πιθανότητα ορθών αποφάσεων που λαμβάνονται με πλειοψηφία, λαμβάνοντας ταυτόχρονα υπόψη ότι οι ψηφοφόροι μπορεί και να σφάλουν. Σε αυτήν την εργασία ο Μαρκήσιος του Condorcet εισήγαγε τη μέθοδο της ταξινομικής ψήφου. Σύμφωνα με τη μέθοδο αυτή, οι ψηφοφόροι εκφράζουν μία σειρά προτίμησης για όλους τους υποψηφίους, δηλαδή τους ταξινομούν. Ανάλογα με τη χώρα στην οποία εφαρμόζεται και τη μέθοδο που χρησιμοποιείται ορίζεται πόσους υποψηφίους έχει το δικαίωμα να ψηφίσει ο κάθε ψηφοφόρος καθώς και εάν επιτρέπεται να υπάρχουν περισσότεροι από ένας στην ίδια σειρά προτίμησης. Για κάθε ψηφοδέλτιο γίνεται σύγκριση καθενός από τους υποψηφίους με τους αντιπάλους του και νικητής ανακηρύσσεται τελικά εκείνος που ικανοποιεί το κριτήριο του Condorcet, εκείνος δηλαδή που όταν αναμετρηθεί με όλους τους άλλους υποψηφίους ανά ζεύγη υπερτερεί σε ψήφους ("pairwise champion"). Η μέθοδος του Condorcet συνίσταται στην προσομοίωση όλων των δυνατών μονομαχιών μεταξύ των υποψηφίων οι οποίες ποικίλουν ανάλογα με τον αριθμό των εκάστοτε υποψηφίων. Όταν προσπαθήσει κανείς να εξάγει μία ολική κατάταξη βρίσκεται αντιμέτωπος με το λεγόμενο *παράδοξο του Condorcet*, διότι οι πλειοψηφικές προτιμήσεις δεν έχουν τη μεταβατική ιδιότητα και υπάρχει το ενδεχόμενο κανένας υποψήφιος να μην νικήσει όλους του τους αντιπάλους του ή ακόμα και το ενδεχόμενο της ισοπαλίας μεταξύ δύο υποψηφίων με αποτέλεσμα να μην μπορεί να υπάρξει ξεκάθαρος νικητής. Για να κατανοήσουμε τη μη-μεταβατικότητα της σχέσης προτίμησης, θα θεωρήσουμε ότι έχουμε τρεις υποψηφίους *Α*,*Β* και *Γ* και θα δεχτούμε ως πιθανό αποτέλεσμα το παρακάτω:

- 20% προτίμησαν τη σειρά *ΑΒΓ*
- 19% προτίμησαν τη σειρά *ΒΓΑ*
- 19% προτίμησαν τη σειρά *ΓΑΒ*

- 14% προτίμησαν τη σειρά *ΑΓΒ*
- 14% προτίμησαν τη σειρά *ΒΑΓ*
- 14% προτίμησαν τη σειρά *ΓΒΑ*

Το συμπέρασμα που προκύπτει από τις ανά δύο συγκρίσεις ζευγών είναι ότι η πλειοψηφία προτιμά τον *Α* έναντι του *Β*, τον *Β* έναντι του *Γ* και τον *Γ* έναντι του *Α*. Δηλαδή, οδηγούμαστε σε έναν κύκλο χωρίς απόφαση. Ο κύκλος αυτός , μπορεί να λυθεί, θεωρώντας τη λιγότερο ισχυρή πλειοψηφική απόφαση ως "εσφαλμένη" πιθανότητα. Αν τροποποιήσουμε λίγο το αποτέλεσμα για τους τρεις υποψηφίους του προηγούμενου παραδείγματος μπορεί εύκολα να φανεί και η επιλεκτική λειτουργία μίας ταξινομικής μεθόδου ψηφοφορίας. Έστω λοιπόν ότι:

- 30% προτίμησαν τη σειρά *ΑΒΓ*
- 26% προτίμησαν τη σειρά *ΒΓΑ*
- 25% προτίμησαν τη σειρά *ΓΒΑ*
- 19% προτίμησαν τη σειρά *ΑΓΒ*

Στην περίπτωση αυτή ο *Α* συγκεντρώνει το 49% των πρώτων προτιμήσεων, ο *Β* το 26% και ο *Γ* το 25%. Το συμπέρασμα που προκύπτει από τις ανά δύο συγκρίσεις ζευγών είναι ότι η πλειοψηφία προτιμά τον *Β* έναντι του *Α* και του *Γ* και τον *Γ* έναντι του *Α*. Επομένως, εάν έπρεπε να εκλεγούν δύο από τους τρεις υποψηφίους η μέθοδος Condorcet θα απέκλειε τον υποψήφιο *Α* παρά το γεγονός ότι συγκέντρωσε σχεδόν τόσες πρώτες προτιμήσεις όσες οι υποψήφιοι *Β* και *Γ* μαζί. Είναι φανερό, από αυτό το παράδειγμα ότι η μέθοδος Condorcet έχει ως σκοπό την εκλογή του υποψηφίου εκείνου που είναι ευρέως αποδεκτός από την πλειοψηφία των πολιτών ακόμα και αν δεν αποτέλεσε την πρώτη τους επιλογή και όχι κάποιου που, ναι μεν, ψηφίστηκε από την πλειονότητα των πολιτών αλλά αποδοκιμάζεται από τους υπόλοιπους. Πιο συγκεκριμένα, η ταξινομική ψήφος δεν επιδιώκει να αναδείξει τη βέλτιστη λύση, με την έννοια του μέγιστου οφέλους, αλλά επιδιώκει να εξασφαλίσει ότι θα αποκλειστούν οι προτάσεις που μία ετερόκλητη πλειοψηφία θα μπορούσε να θεωρήσει επικίνδυνες περιορίζοντας τα υποτιθέμενα λάθη των ψηφοφόρων.
Η μέθοδος του Condorcet αποτελείται από μία σειρά επιμέρους μεθόδων τις οποίες και θα παρουσιάσουμε και θα αναλύσουμε στη συνέχεια.

i) Η μέθοδος Copeland

Αποτελεί την απλούστερη μέθοδο Condorcet σύμφωνα με την οποία νικητής ανακηρύσσεται εκείνος που έχει νικήσει στις περισσότερες συγκρίσεις ζευγών. Συγκεκριμένα, ο νικητής κάθε σύγκρισης παίρνει έναν βαθμό ενώ σε περίπτωση ισοπαλίας παίρνουν και οι δύο από μισό βαθμό. Νικητής τελικά εκλέγεται εκείνος που έχει συγκεντρώσει τους περισσότερους βαθμούς. Παρόλα αυτά, υπάρχει πάντα ο κίνδυνος της ισοπαλίας μεταξύ των υποψηφίων και γι’ αυτό το λόγο δεν προτιμάται.

ii) Η μέθοδος MiniMax

Νικητής ανακηρύσσεται εκείνος που έχει νικήσει όλους τους αντιπάλους του στη σύγκριση ζευγών. Σε ενδεχόμενη ισοπαλία μεταξύ δύο υποψηφίων, νικητής ανακηρύσσεται εκείνος του οποίου η χειρότερη ήττα στη σύγκριση ζευγών, δηλαδή η ήττα με τη μεγαλύτερη διαφορά ψήφων, είναι καλύτερη από τις αντίστοιχες των υπόλοιπων υποψηφίων. Σε αυτήν τη μέθοδο, οι πιθανότητες ισοπαλίας μεταξύ των υποψηφίων είναι ελάχιστες διότι είναι σχεδόν απίθανο δύο υποψήφιοι να έχουν ακριβώς την ίδια διαφορά ψήφων στις χειρότερες ήττες τους.

iii) Η μέθοδος Kemeny-Young

Οι ψηφοφόροι τοποθετούν με σειρά προτίμησης όλους τους υποψήφιους έχοντας τη δυνατότητα να τοποθετήσουν περισσότερους από έναν στην ίδια σειρά. Για κάθε ψηφοδέλτιο γίνεται σύγκριση καθενός από τους υποψηφίους με τους αντιπάλους του και ο υποψήφιος εκείνος που έχει νικήσει στις περισσότερες συγκρίσεις ζευγών θεωρείται πρώτη προτίμηση των ψηφοφόρων, αυτός που τον ακολουθεί δεύτερη προτίμηση κ.ο.κ. Στη συνέχεια, δημιουργείται ένας δεύτερος πίνακας ο οποίος παρουσιάζει τα ζεύγη των υποψηφίων και χωρίζεται σε τρεις στήλες, μία για τον έναν υποψήφιο κάθε ζεύγους, μία για τον άλλον και μία και για τους δύο σε περίπτωση που έχουν τοποθετηθεί στην ίδια σειρά προτίμησης. Για κάθε ζευγάρι υπάρχει και ένας νικητής στον οποίο προστίθεται μία ψήφος στην αντίστοιχη στήλη του πίνακα. Τελικά, νικητής ανακηρύσσεται εκείνος που έχει συγκεντρώσει το μεγαλύτερο άθροισμα ψήφων σε σύγκριση με τους αντιπάλους του. Θα πρέπει να σημειωθεί ότι σε περίπτωση που δύο υποψήφιοι έχουν καταταχθεί στην ίδια σειρά προτίμησης, η ψήφος που προστίθεται στη στήλη που αντιστοιχεί και στους δύο δε λαμβάνεται υπόψη στον υπολογισμό του τελικού αποτελέσματος.

iv) Η μέθοδος Tideman (Ranked Pairs)

Οι ψηφοφόροι τοποθετούν με σειρά προτίμησης όλους τους υποψήφιους και για κάθε ψηφοδέλτιο γίνεται σύγκριση καθενός από τους υποψηφίους με τους αντιπάλους του. Στη συνέχεια γίνεται κατάταξη των ζευγαριών από το μεγαλύτερο στο μικρότερο, με μεγαλύτερο το ζευγάρι όπου ο νικητής έχει επικρατήσει έναντι του αντιπάλου του με τη μεγαλύτερη διαφορά ψήφων. Τέλος, ένα-ένα τα ζευγάρια "κλειδώνουν" ζωγραφίζοντας κάθε φορά ένα βέλος που στοχεύει τον υποψήφιο που έχει χάσει στη σύγκριση ζευγών με εξαίρεση τα ζευγάρια εκείνα που δημιουργούν κύκλους στο διάγραμμα και αποκλείονται από τη διαδικασία. Νικητής ανακηρύσσεται εκείνος που δεν έχει καμία ήττα στις συγκρίσεις ζευγών ή με άλλα λόγια δεν υπάρχει κανένα βέλος που να στοχεύει πάνω σε αυτόν.

v) Η μέθοδος Schulze

Η μέθοδος Schulze είναι στις μέρες μας η πιο διαδεδομένη μέθοδος Condorcet. Οι ψηφοφόροι τοποθετούν με σειρά προτίμησης όλους τους υποψήφιους και όπως και στις προηγούμενες μεθόδους, για κάθε ψηφοδέλτιο γίνεται σύγκριση καθενός από τους υποψηφίους με τους αντιπάλους του. Για την ανάδειξη του τελικού νικητή, ακολουθείται μία περίπλοκη διαδικασία την οποία και θα παραθέσουμε στη συνέχεια.

Έστω, $d[X,Y]$ ο αριθμός των ψηφοφόρων που προτιμούν τον υποψήφιο X από τον υποψήφιο Y. Ένα *μονοπάτι (path)* από τον υποψήφιο X στον υποψήφιο Y με *δύναμη* (*strength*) p, είναι μια σειρά από υποψηφίους $C(1),...,C(n)$ με τις εξής ιδιότητες:

- $C(1) = X$ και $C(n) = Y$
- Για κάθε $i = 1,...,(n-1) : d[C(i),C(i+1)] > d[C(i+1),C(i)]$
- Η δύναμη p ενός μονοπατιού είναι η ελάχιστη τιμή όλων των $d[C(i),C(i+1)]$ για $i = 1,...,(n-1)$ ("the strength of a chain is as much as the strength of its weakest link")

Η *δύναμη του πιο ισχυρού μονοπατιού* $p[X,Y]$ από τον υποψήφιο X στον υποψήφιο Y, είναι η μεγαλύτερη τιμή ώστε να υπάρχει ένα μονοπάτι από τον υποψήφιο X στον υποψήφιο Y με εκείνη τη δύναμη. Αν δεν υπάρχει κανένα μονοπάτι από τον υποψήφιο X στον υποψήφιο Y, τότε $p[X,Y] := 0$.

Ο υποψήφιος X είναι *καλύτερος* από τον υποψήφιο Y αν και μόνο αν $p[X,Y] > p[Y,X]$ ενώ ο υποψήφιος X είναι τελικά *πιθανός νικητής* αν και μόνο αν $p[X,Y] \geq p[Y,X]$ για κάθε άλλον υποψήφιο Y.

Η μέθοδος της ταξινομικής ψήφου δεν είναι ευρέως διαδεδομένη όσον αφορά τις περιφερειακές εκλογές για την ανάδειξη μεμονωμένων υποψηφίων ή συνδυασμών κομμάτων. Στην Ελλάδα όμως, βρίσκει πρακτική εφαρμογή καθώς χρησιμοποιείται για την εκλογή των εσωτερικών μελών του Συμβουλίου κάθε Πανεπιστημίου. Σύμφωνα με το νέο Νόμο Πλαίσιο "Τα εσωτερικά μέλη εκλέγονται από το σύνολο των καθηγητών του οικείου ιδρύματος με ενιαίο ψηφοδέλτιο και με σημείωση από τους εκλογείς δίπλα από το όνομα των υποψηφίων της σειράς προτίμησης με διαδοχικούς ακέραιους αριθμούς". (άρθρο 8, παρ.4, εδ.β) Στην αιτιολογική έκθεση για το Νομοσχέδιο αναφέρεται ότι "Για την εκλογική διαδικασία ανάδειξης των καθηγητών–μελών του Συμβουλίου εισάγεται για πρώτη φορά η εφαρμογή του συστήματος της ταξινομικής ψήφου, έτσι ώστε να διασφαλιστεί η ισόρροπη εκπροσώπηση της κοινότητας του ιδρύματος στο Συμβούλιο, χωρίς αποκλεισμούς ή κυριαρχία επιμέρους αντιλήψεων". Η πρόταση αυτή, έρχεται αντιμέτωπη με την επιστήμη των μαθηματικών και της στατιστικής καθώς όπως αποδείξαμε στην αρχή της παραγράφου, ένας υποψήφιος που συγκεντρώνει το 49% των πρώτων προτιμήσεων είναι δυνατόν να αποκλειστεί έπειτα από μία "καλή συνεννόηση" μεταξύ των άλλων υποψηφίων. Επομένως, η εφαρμογή της συγκεκριμένης μεθόδου όχι μόνο δεν εμποδίζει τους "αποκλεισμούς" όπως ισχυρίζονται οι υποστηρικτές του, αλλά στην πραγματικότητα τους διευκολύνει.

ε) Το σύστημα Borda

Οι ψηφοφόροι τοποθετούν με σειρά προτίμησης όλους τους υποψηφίους και νικητής ανακηρύσσεται εκείνος που έχει συγκεντρώσει τους περισσότερους βαθμούς. Ο γενικός κανόνας διανομής των βαθμών μεταξύ των υποψηφίων είναι ο ακόλουθος: "Αν K ο αριθμός των υποψηφίων, αυτός που βρίσκεται πρώτος στο ψηφοδέλτιο θα λάβει $K-1$ βαθμούς, αυτός που βρίσκεται δεύτερος $K-2$ βαθμούς κ.ο.κ ενώ αυτός που βρίσκεται τελευταίος δε θα λάβει κανένα βαθμό".

Συγκρίνοντας τα συστήματα που προτάθηκαν από τον Condorcet και το Borda, θα μπορούσε να πει κανείς ότι οι δύο αυτές μέθοδοι έχουν σχεδιαστεί για πετύχουν δύο διαφορετικούς σκοπούς. Η μέθοδος του Condorcet έχει ως απώτερο σκοπό την εκλογή του υποψηφίου που είναι ευρέως αποδεκτός από την απόλυτη πλειοψηφία των πολιτών ακόμα και αν δεν αποτέλεσε την πρώτη τους επιλογή ενώ το σύστημα Borda έχει ως απώτερο σκοπό την εκλογή του υποψηφίου που είναι ευρέως αποδεκτός από όλους τους πολίτες και γι' αυτό το λόγο λαμβάνει υπόψη και τους υποψηφίους που βρίσκονται πιο χαμηλά στην προτίμηση των ψηφοφόρων.

στ) Το σύστημα Bucklin

Οι ψηφοφόροι τοποθετούν με σειρά προτίμησης όλους τους υποψηφίους. Στην πρώτη καταμέτρηση των ψήφων λαμβάνονται υπόψη μόνο όσοι υποψήφιοι ήρθαν πρώτοι στις προτιμήσεις των ψηφοφόρων τους και νικητής ανακηρύσσεται εκείνος που έχει συγκεντρώσει την απόλυτη πλειοψηφία των ψήφων. Σε διαφορετική περίπτωση, ακολουθεί δεύτερος γύρος στον οποίο προσμετρούνται και οι ψήφοι όσων ήρθαν δεύτεροι στις προτιμήσεις των ψηφοφόρων τους και οι ψήφοι τους προστίθενται σε εκείνους που κατατάχθηκαν πρώτοι σε κάθε ψηφοδέλτιο. Η διαδικασία αυτή συνεχίζεται με επόμενους γύρους μέχρις ότου κάποιος καταλάβει την απόλυτη πλειοψηφία και ανακηρυχθεί νικητής. Το συγκεκριμένο εκλογικό σύστημα χρησιμοποιήθηκε στις Η.Π.Α στις αρχές του 20^{ου} αιώνα, εγκαταλείφθηκε όμως γρήγορα καθώς χαρακτηρίστηκε αντισυνταγματικό.

ζ) Το σύστημα Coomb

Οι ψηφοφόροι τοποθετούν με σειρά προτίμησης όλους τους υποψηφίους και νικητής ανακηρύσσεται εκείνος που έχει καταταχθεί πρώτος στην απόλυτη πλειοψηφία των ψηφοδελτίων οπότε και η διαδικασία ολοκληρώνεται στον ένα γύρο. Σε διαφορετική περίπτωση, ακολουθεί δεύτερος γύρος στον οποίο ο υποψήφιος που έχει καταλάβει την τελευταία θέση στα περισσότερα ψηφοδέλτια εξαιρείται από τη διαδικασία και οι ψήφοι του μεταφέρονται στο δεύτερο κατά σειρά προτίμησης υποψήφιο όπως προέκυψε από τον πρώτο γύρο. Η εκλογική διαδικασία επαναλαμβάνεται με επόμενους γύρους εκλογών με τις ψήφους των αποκλεισθέντων να μεταφέρονται κάθε φορά στους τρίτους, στους τέταρτους κ.ο.κ κατά σειρά προτίμησης υποψηφίους των προηγούμενων γύρων μέχρις ότου κάποιος καταλάβει την απόλυτη πλειοψηφία.

η) Εναλλακτική Ψήφος

Η εναλλακτική ψήφος, δανείζεται στοιχεία από τα συστήματα της απόλυτης πλειοψηφίας και της εξαντλητικής ψήφου ενώ παρουσιάζει πολλές ομοιότητες και με το σύστημα Coomb. Οι ψηφοφόροι τοποθετούν με σειρά προτίμησης όλους τους υποψηφίους και νικητής ανακηρύσσεται εκείνος που έχει καταταχθεί πρώτος στην απόλυτη πλειοψηφία των ψηφοδελτίων. Σε διαφορετική περίπτωση, ακολουθεί δεύτερος γύρος εκλογών στον οποίο ο υποψήφιος που έχει καταλάβει την πρώτη θέση στα λιγότερα ψηφοδέλτια (η μόνη διαφορά με το σύστημα Coomb) εξαιρείται από τη διαδικασία και οι ψήφοι του μεταφέρονται στο δεύτερο κατά σειρά προτίμησης υποψήφιο όπως προέκυψε από τον πρώτο γύρο. Η εκλογική διαδικασία επαναλαμβάνεται με επόμενους γύρους εκλογών με τις ψήφους των αποκλεισθέντων να μεταφέρονται κάθε φορά στους τρίτους, στους τέταρτους κ.ο.κ κατά σειρά προτίμησης υποψηφίους των προηγούμενων γύρων μέχρις ότου κάποιος καταλάβει την απόλυτη πλειοψηφία. Στο συγκεκριμένο εκλογικό σύστημα, οι ψηφοφόροι δε χρειάζεται να ψηφίζουν σε κάθε γύρο εκλογών καθώς η μέτρηση κάθε φορά γίνεται από τα ήδη υπάρχοντα ψηφοδέλτια του πρώτου γύρου. Αν και εξαιρετικά σπάνιο, έχει συμβεί στο παρελθόν δύο υποψήφιοι να ισοψηφήσουν στις τελευταίες θέσεις των προτιμήσεων των ψηφοφόρων τους (αναπληρωματικές εκλογές για πλήρωση κενής θέσης στη Βουλή –by-elections- στις επαρχίες του Haltemprice και του Howden το 2008). Σύμφωνα με τη νομοθεσία, στην περίπτωση αυτή, χρησιμοποιείται η μέθοδος του κλήρου προκειμένου να επιλεγεί ποιος από τους δύο υποψήφιους θα εξαιρεθεί πρώτος. Συγκεκριμένα, ο υποψήφιος που κερδίζει τον κλήρο, κερδίζει μία επιπλέον ψήφο που του επιτρέπει να παραμείνει στην εκλογική διαδικασία ενώ ο άλλος εξαιρείται από αυτήν.

Αξίζει να σημειωθεί ότι στις 5 Μαίου 2011, οι πολίτες τις Μεγάλης Βρετανίας απέρριψαν σε δημοψήφισμα την πρόταση να αντικαταστήσουν το εκλογικό τους σύστημα, με ένα σύστημα εναλλακτικής ψήφου στις βουλευτικές εκλογές. Δηλαδή, να μην κυριαρχεί άνευ όρων το πλειοψηφικό σύστημα, αλλά οι ψηφοφόροι να μπορούν να σημειώσουν δεύτερη ή τρίτη επιλογή, σε περίπτωση δεύτερης κατανομής εδρών. Η συμμετοχή στο δημοψήφισμα ήταν της τάξης του 42,2 %, με το 68 % να καταψηφίζει την πρόταση για αλλαγή του εκλογικού νόμου.

1.2.2 Τα Αναλογικά εκλογικά συστήματα

Οι έδρες κάθε εκλογικής περιφέρειας κατανέμονται στους συνδυασμούς και στους μεμονωμένους υποψηφίους που μετέχουν στις εκλογές ανάλογα με την εκλογική τους δύναμη σε κάθε εκλογική περιφέρεια. Δηλαδή, ανάλογα με τον αριθμό των ψήφων που έλαβαν. Εφαρμόζεται αποκλειστικά σε πολυεδρικές εκλογικές περιφέρειες και η κατανομή των εδρών γίνεται με βάση το λεγόμενο *εκλογικό μέτρο* το οποίο εκφράζει τον ελάχιστο αριθμό ψήφων που απαιτείται να συγκεντρώσει ένας συνδυασμός ή ένας μεμονωμένος υποψήφιος προκειμένου να καταλάβει μία έδρα. Για τον καθορισμό του εκλογικού μέτρου, ο τρόπος υπολογισμού του οποίου αποτελεί και τη σημαντικότερη διαφορά μεταξύ των αναλογικών εκλογικών συστημάτων, έχουν αναπτυχθεί διάφορα συστήματα τα οποία και θα αναλύσουμε στη συνέχεια.

α) Το σύστημα d'Hondt

Στο συγκεκριμένο εκλογικό σύστημα, κάθε συνδυασμός (ή μεμονωμένος υποψήφιος) καταλαμβάνει έδρες ανάλογα με τον αριθμό των ψήφων του και τον αριθμό των εδρών που έχει ήδη στην κατοχή του. Πιο αναλυτικά, κάθε έδρα παραχωρείται στο συνδυασμό με το μεγαλύτερο πηλίκο

$$\frac{V_i}{S_i(t)+1}$$

όπου V_i είναι ο συνολικός αριθμός ψήφων του *i*-συνδυασμού και $S_i(t)$ είναι ο συνολικός αριθμός εδρών που έχει λάβει μέχρι τη στιγμή t ο *i*-συνδυασμός. Είναι σκόπιμο να αναφερθεί ότι όλοι οι υποψήφιοι συνδυασμοί ξεκινούν από μηδέν έδρες δηλαδή, $S_i(0)=0$ και αυτές αυξάνονται κάθε φορά που καταλαμβάνουν μία. Η διαδικασία επαναλαμβάνεται με επόμενα στάδια μέχρι να συμπληρωθούν όλες οι διαθέσιμες έδρες, με το σύνολο των ψήφων V_i του κάθε συνδυασμού να διαιρείται σταδιακά με μεγαλύτερους αριθμούς κάθε φορά που αυτός καταλαμβάνει μία έδρα. Σε ορισμένες παραλλαγές του συστήματος υπάρχει εκλογικό όριο το οποίο δεν επιτρέπει σε ένα συνδυασμό(ή μεμονωμένο υποψήφιο) να καταλάβει έδρες αν δεν υπερβαίνει κάποιο όριο. Ενδεικτικά αναφέρουμε ότι το εκλογικό όριο είναι για την Ισπανία 3%, για τη Σλοβενία 4%, ενώ για την Τουρκία 10%. Σε μία άλλη παραλλαγή του συστήματος, ένας συγκεκριμένος αριθμός εδρών διανέμεται αυτόματα στο συνδυασμό που συγκέντρωσε τις περισσότερες ψήφους (“the majority bonus”) με τις υπόλοιπες να διανέμονται αναλογικά στους υπόλοιπους συνδυασμούς συμπεριλαμβανομένου και του πρώτου.

β) Το σύστημα Sainte- Laguë

Το σύστημα Sainte- Laguë είναι παρόμοιο με το σύστημα d’Hondt με τη μόνη διαφορά ότι ως διαιρέτες χρησιμοποιούνται μόνο περιττοί αριθμοί. Κάθε έδρα παραχωρείται στο συνδυασμό με το μεγαλύτερο πηλίκο

$$\frac{V_i}{2S_i(t)+1}$$

και η διαδικασία επαναλαμβάνεται με επόμενα στάδια μέχρι να συμπληρωθούν όλες οι διαθέσιμες έδρες.

γ) Το σύστημα Hare

Το εκλογικό μέτρο υπολογίζεται διαιρώντας το σύνολο των ψήφων V προς το σύνολο των διαθέσιμων εδρών S.

$$\text{Εκλογικό Μέτρο} = \frac{V}{S}$$

Κάθε συνδυασμός (ή μεμονωμένος υποψήφιος) καταλαμβάνει τόσες έδρες όσες φορές η εκλογική του δύναμη καλύπτει το εκλογικό μέτρο, δηλαδή όσο είναι το ακέραιο πηλίκο της διαίρεσης των ψήφων που συγκέντρωσε με το εκλογικό μέτρο. Όσες έδρες δεν διατίθενται με τον τρόπο αυτό, μεταφέρονται κατά σειρά στους συνδυασμούς που παρουσιάζουν τα μεγαλύτερα υπόλοιπα ψήφων και η διαδικασία επαναλαμβάνεται μέχρι να συμπληρωθούν όλες οι διαθέσιμες έδρες. Σε περίπτωση που κάποιος υποψήφιος δεν καταφέρει παρόλα αυτά να συμπληρώσει το εκλογικό μέτρο, τότε ο υποψήφιος με τις λιγότερες ψήφους αποκλείεται από τη διαδικασία και οι ψήφοι του μεταφέρονται σε εκείνον με τις περισσότερες ψήφους προκειμένου να συμπληρώσει το εκλογικό μέτρο και να εκλεγεί.

δ) Το σύστημα Droop

Το εκλογικό μέτρο υπολογίζεται διαιρώντας το σύνολο των ψήφων V προς το σύνολο των εδρών αυξημένο κατά μία $(S+1)$ και προσθέτοντας στο αποτέλεσμα της διαίρεσης μία μονάδα.

$$\text{Εκλογικό Μέτρο} = \frac{V}{S+1}+1$$

Το σύστημα Droop είναι παρόμοιο με εκείνο του Hare. Η μόνη διαφορά, που το καθιστά και αποτελεσματικότερο, έγκειται στο μικρότερο εκλογικό μέτρο το οποίο βοηθά στην άμεση συμπλήρωση των εδρών χωρίς να χρειάζονται πολλές επόμενες κατανομές.

ε) Το σύστημα Hagenbach-Biscoff

Το εκλογικό μέτρο υπολογίζεται διαιρώντας το σύνολο των ψήφων V προς το σύνολο των εδρών S αυξημένο κατά μία μονάδα.

$$\text{Εκλογικό Μέτρο} = \frac{V}{S+1}$$

Το σύστημα Hagenbach-Biscoff πολλές φορές ταυτίζεται με εκείνο του Droop εξαιτίας των ομοιοτήτων που παρουσιάζουν. Παρόλα αυτά, είναι υποδεέστερό του καθώς λόγω του μικρού εκλογικού μέτρου που παράγει, υπάρχει το ενδεχόμενο να εκλεγούν περισσότεροι υποψήφιοι από τις διαθέσιμες έδρες πράγμα το οποίο δε συμβαίνει στο σύστημα Droop.

στ) Το σύστημα Imperiali

Το εκλογικό μέτρο υπολογίζεται διαιρώντας το σύνολο των ψήφων V προς το σύνολο των εδρών S συν δύο.

$$\text{Εκλογικό Μέτρο} = \frac{V}{S+2}$$

Η διαδικασία που ακολουθείται για την κατανομή των εδρών είναι παρόμοια με εκείνη των συστημάτων Hare και Droop. Το βασικό μειονέκτημα του Imperiali είναι ότι λόγω του πολύ μικρού εκλογικού μέτρου, υπάρχει το ενδεχόμενο να εκλεγούν περισσότεροι υποψήφιοι από τις διαθέσιμες έδρες πράγμα το οποίο δε συμβαίνει στα δύο προηγούμενα συστήματα.

ζ) Το σύστημα Niemeyer

Για τον υπολογισμό των εδρών που δικαιούται να καταλάβει κάθε συνδυασμός (ή μεμονωμένος υποψήφιος), οι συνολικές ψήφοι κάθε συνδυασμού V_i πολλαπλασιάζονται με το σύνολο των εδρών S και το γινόμενο αυτό διαιρείται με το σύνολο των ψήφων V.

$$\text{Έδρες}_{\text{i}} = \frac{V_i * S}{V}$$

Με την ολοκλήρωση της διαδικασίας, τυχόν αδιάθετες έδρες κατανέμονται στους συνδυασμούς με τα μεγαλύτερα υπόλοιπα όπως εκείνα προέκυψαν από την παραπάνω διαίρεση.

1.2.3 Το μικτό εκλογικό σύστημα

Δεν αποτελεί ξεχωριστό είδος εκλογικού συστήματος αλλά, όπως δηλώνει και το όνομά του, αποτελεί συνδυασμό τεχνικών των δύο βασικών εκλογικών συστημάτων για την ανάδειξη των νικητών και την κατανομή των εδρών. Πιο αναλυτικά, κάθε μικτό εκλογικό σύστημα έχει ως βάση κάποιο αναλογικό ή κάποιο πλειοψηφικό και δανείζεται επιπρόσθετα σημαντικά στοιχεία από το άλλο σύστημα για την τελειοποίησή του. Στις μέρες μας, αρκετές χώρες εφαρμόζουν μεικτά συστήματα τα οποία έχουν ως απώτερο σκοπό να περιορίσουν τα μειονεκτήματα των δύο αμιγών εκλογικών συστημάτων και να εκμετελευτούν τα πλεονεκτήματά τους. Χαρακτηριστικό παράδειγμα αποτελεί η Ελλάδα. Το εκλογικό σύστημα που εφαρμόστηκε στις εκλογές της $6^{ης}$ Μαΐου του 2012 (σύμφωνα με το νόμο Ν.3231/2004 όπως τροποποιήθηκε από το νόμο 3636/2008) είναι η ενισχυμένη αναλογική, ένα μεικτό σύστημα που συνδυάζει στοιχεία απλής αναλογικής και πλειοψηφικού. Συγκεκριμένα, από τις τριακόσιες βουλευτικές έδρες που πρέπει να κατανεμηθούν,

- οι διακόσιες πενήντα κατανέμονται απολύτως αναλογικά σε όσους συνδυασμούς ξεπεράσουν το 3% (στοιχείο απλής αναλογικής με όριο)
- οι άλλες πενήντα χαρίζονται στον πρώτο συνδυασμό (στοιχείο πλειοψηφικού-“bonus majority”)

Με αυτόν τον τρόπο αρκεί στον πρώτο συνδυασμό το 40.5% των ψήφων για να έχει αυτοδυναμία 151 εδρών στη Βουλή. Όσες δε περισσότερες είναι οι ψήφοι των συνδυασμών που δεν ξεπερνούν το 3%, τόσο χαμηλώνει το ποσοστό που απαιτείται για αυτοδύναμη πλειοψηφία.

1.3 Πρόβλεψη του εκλογικού αποτελέσματος

Η πρόβλεψη του εκλογικού αποτελέσματος αποτελεί ένα -διαρκώς αυξανόμενου ενδιαφέροντος- αντικείμενο μελέτης της σύγχρονης πολιτικής επιστήμης. Οι εταιρίες δημοσκοπήσεων επενδύουν πριν από κάθε εκλογική αναμέτρηση μεγάλα χρηματικά ποσά για την διεξαγωγή των *exit polls (“δημοσκοπήσεις εξόδου”)* έχοντας ως απώτερο στόχο την αξιόπιστη και όσο το δυνατόν ακριβέστερη πρόβλεψη του τελικού εκλογικού αποτελέσματος. Εμπνευστής τους θεωρείται η Warren Mitofsky, μία πολιτική δημοσκόπος, η οποία το 1967 διεξήγαγε το πρώτο exit poll για ένα δίκτυο μέσων μαζικής ενημέρωσης κατά την εκλογή του κυβερνήτη της πολιτείας του Κεντάκι. Στην Ελλάδα, η πρώτη απόπειρα να διεξαχθεί exit poll έγινε στις βουλευτικές εκλογές του 1993 κάτι το οποίο τελικά δεν πραγματοποιήθηκε λόγω σχετικής απαγόρευσης που είχε θέσει η κυβέρνηση του κ. Κ. Μητσοτάκη. Στις ευρωεκλογές του 1994, στις οποίες δεν ίσχυαν οι κυβερνητικές απαγορεύσεις, η εταιρεία δημοσκοπήσεων Opinion του κ. Ηλία Νικολακόπουλου διεξήγαγε τελικά σε συνεργασία με τη γαλλική PVA το πρώτο επίσημο exit poll ως εργαλείο μέτρησης της ψήφου. Από την εφαρμογή της πρώτης δημοσκόπησης εξόδου το 1967 μέχρι και σήμερα, οι εταιρείες δημοσκοπήσεων σε παγκόσμιο επίπεδο έχουν δεχθεί εύσημα για την ακρίβειά τους αλλά και πυρά για τις αποκλίσεις των προβλέψεών τους. Οι άνθρωποι των εταιρειών υπερασπίζονται την επιστημονική ακρίβεια των exit polls λέγοντας ότι η αξιοπιστία τους είναι δεδομένη και διαχωρίζοντάς τα από τις δημοσκοπήσεις. Ωστόσο η μελέτη των ευρημάτων των τελευταίων ετών αποκαλύπτει ότι πλέον οι ερωτώμενοι δεν είναι τόσο ειλικρινείς όσο στο παρελθόν, γεγονός που δυσχεραίνει ακόμη περισσότερο το έργο των δημοσκόπων.

1.3.1 Η σημασία του exit poll

Με τον όρο exit poll αναφερόμαστε σε μία ειδικού τύπου δημοσκόπηση, σε μία εικονική επανάληψη της ψηφοφορίας κατά την οποία κάθε ερωτώμενος ψηφοφόρος καλείται να επαναλάβει μία πράξη την οποία μόλις έκανε, δηλαδή να απαντήσει στην ερώτηση *“τι ψηφίσατε μόλις τώρα”*. Αυτό αποτελεί και την κύρια διαφορά μεταξύ exit poll και δημοσκοπήσεων καθώς οι δεύτερες εκφράζουν τη στάση των ερωτηθέντων για κάτι το οποίο θα συμβεί στο μέλλον. Απαντούν δηλαδή στην ερώτηση *“τι θα ψηφίσετε στις επερχόμενες* *εκλογές"*. Για τις εταιρίες δημοσκοπήσεων αυτό που παίζει κυρίαρχο ρόλο για το αποτέλεσμα είναι ο αριθμός των ψηφοφόρων που θα κληθούν να απαντήσουν στα ερωτηματολόγια των ερευνητών, δηλαδή ο καθορισμός του μεγέθους του δείγματος n, καθώς και το βήμα με το οποίο θα επιλέγεται καθένας από τους n ψηφοφόρους (είναι διαφορετικό σε κάθε εκλογικό κέντρο και επηρεάζεται από τον αριθμό τον εγγεγραμμένων ψηφοφόρων σε αυτό) έτσι ώστε το δείγμα να θεωρείται τυχαίο. Η - κατόπιν εκλογών - χρησιμότητα του exit poll κρίνεται ιδιαίτερα σημαντική, καθώς είναι το μοναδικό μέσο παροχής πληροφοριών σχετικά με τις μετακινήσεις των ψηφοφόρων αλλά και διαφόρων άλλων κοινωνικών χαρακτηριστικών του εκλογικού σώματος όπως, το πώς ψήφισαν οι πολίτες ανάλογα με την ηλικία, το επάγγελμα και το επίπεδο μόρφωσης, τα κριτήρια που καθόρισαν την ψήφο τους κτλ.

1.3.2 Συλλογή των δεδομένων

Την ημέρα των εκλογών οι εταιρείες δημοσκοπήσεων στέλνουν έναν αριθμό *ερευνητών* και *καταμετρητών* σε συγκεκριμένα σχολεία – εκλογικά κέντρα της περιφέρειας προκειμένου να διεξάγουν τα exit polls. Οι καταμετρητές είναι υπεύθυνοι να μετρούν όλους όσους εξέρχονται από την κεντρική πόρτα του σχολείου και ανάλογα με το βήμα του κάθε εκλογικού κέντρου να υποδεικνύουν στους ερευνητές το άτομο από το οποίο πρέπει να πάρουν προσωπική συνέντευξη. Στόχος των ερευνητών είναι να πλησιάζουν τον ψηφοφόρο που τους υποδεικνύει ο καταμετρητής τους προκειμένου να απαντήσει στο ανώνυμο ερωτηματολόγιο. Τα άτομα – ψηφοφόροι που δέχονται να συμμετέχουν στη διαδικασία (συνέντευξη), συμπληρώνουν τη φόρμα του ερωτηματολογίου και το ρίχνουν μέσα σε μία κάλπη. Στην περίπτωση που κάποιος αρνηθεί να απαντήσει, τότε ο ερευνητής συμπληρώνει το φύλο και την εκτιμώμενη ηλικία του ψηφοφόρου με την ένδειξη "άρνηση" και το ρίχνει και πάλι μέσα στην κάλπη καθώς, στατιστικά, μπορεί και αυτό να παράγει κάποια πληροφορία.

Ένα τυπικό ερωτηματολόγιο exit poll που καλείται να συμπληρώσει ένας ψηφοφόρος περιλαμβάνει ερωτήσεις που αφορούν:

- το φύλο,
- την ηλικία,
- το κόμμα που ψήφισε σήμερα,
- το κόμμα που είχε ψηφίσει στις προηγούμενες εκλογές,

- το πότε αποφάσισε ποιο κόμμα θα ψηφίσει,
- το επίπεδο μόρφωσης,
- το επάγγελμα.

1.3.3 Ακρίβεια και αξιοπιστία

Σύμφωνα με τους εκλογικούς αναλυτές, αν θεωρήσουμε ότι η επιλογή του δείγματός είναι επιστημονική, τότε ο μόνος κίνδυνος που υπάρχει να "πέσουν έξω" τα exit polls είναι να μην είναι ειλικρινείς οι απαντήσεις των ψηφοφόρων. Σε αυτόν ακριβώς τον αστάθμητο παράγοντα της ανθρώπινης συμπεριφοράς, αλλά και στη γενικευμένη πολιτική ρευστότητα αποδίδουν οι εταιρείες δημοσκοπήσεων την "αποτυχία" των exit polls τα τελευταία χρόνια σε Ιταλία, Ηνωμένο Βασίλειο, Γερμανία, Ουκρανία και φυσικά και στην Ελλάδα με πιο πρόσφατη εκείνη του πρώτου γύρου των δημοτικών και περιφερειακών εκλογών του Μαΐου του 2014. Τότε, το κοινό exit poll των εταιριών δημοσκοπήσεων Opinion, Alco, GPO, Marc, Metron Analysis και MRB έδειχνε τον κ. Γαβριήλ Σακελλαρίδη να έχει άνετο προβάδισμα στο δήμο της Αθήνας με ποσοστό που θα κυμαίνονταν μεταξύ του [20% , 24%]. Στα επίσημα εκλογικά αποτελέσματα, ο κ. Σακελλαρίδης βρίσκονταν στη δεύτερη θέση με ποσοστό κοντά στο 20% ενώ προηγείτο ο κ. Καμίνης με περίπου 2 ποσοστιαίες μονάδες διαφορά. Μεγάλο προβάδισμα και άνετη επικράτηση έδιναν τα exit polls και στην περίπτωση της κ. Δούρου την οποία εμφάνιζαν να προηγείται με ποσοστό που θα μπορούσε να αγγίξει ως και το 31%, με τον κ. Σγουρό να έρχεται δεύτερος με ποσοστά που κινούνταν μεταξύ [20.5% , 24.5%]. Παρόλα αυτά, στα επίσημα εκλογικά αποτελέσματα η διαφορά της κ. Δούρου και του κ. Σγουρού δεν ξεπέρασε τελικά την 1.5 ποσοστιαία μονάδα.

Παρόμοιο ήταν το σκηνικό και στις επαναληπτικές βουλευτικές εκλογές του Ιουνίου του 2012 όπου το κοινό exit poll έδινε τη ΝΔ να κινείται μεταξύ [27.5% , 30.5%], με τον ΣΥ.ΡΙΖ.Α να ακολουθεί σε απόσταση αναπνοής με ποσοστά που κινούνταν μεταξύ [27% , 30%]. Αντίστοιχη πρόβλεψη της Public Issue εμφάνιζε τον ΣΥ.ΡΙΖ.Α να προηγείται με ποσοστό που θα μπορούσε να φθάσει ως και το 31%, με τη ΝΔ να έρχεται δεύτερη συγκεντρώνοντας το πολύ το 30% των ψήφων. Στα επίσημα εκλογικά αποτελέσματα, η ΝΔ (29.66%) επικράτησε άνετα του δεύτερου ΣΥ.ΡΙΖ.Α (26.89%) με τη μεταξύ τους διαφορά να αγγίζει σχεδόν τις 3 ποσοστιαίες μονάδες.

Οι δύο τελευταίες εκλογικές αναμετρήσεις αποτέλεσαν πράγματι δοκιμασία για την προβλεπτική ικανότητα και την αξιοπιστία των ελληνικών exit polls.

Τα σημαντικότερα προβλήματα που ανέκυψαν και δυσχέραιναν το έργο των εκλογικών αναλυτών ήταν τα εξής:

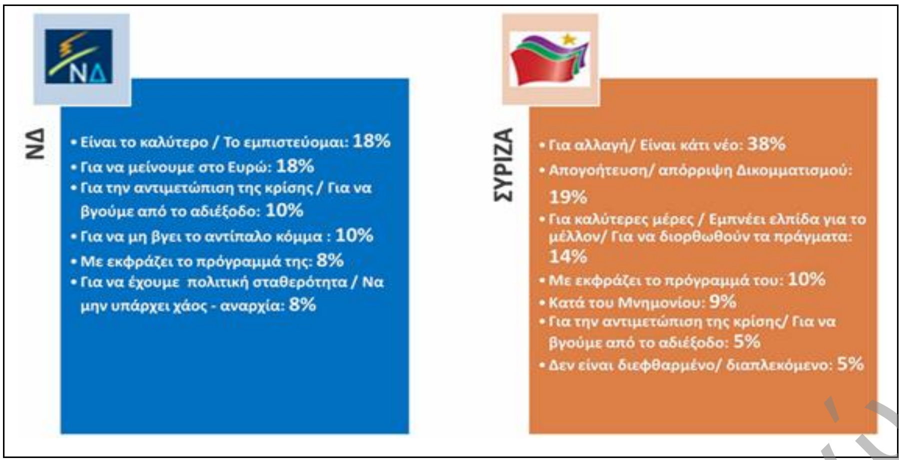
- Το μεγάλο ποσοστό αποχής που άγγιξε στις εκλογές του Ιουνίου το 37.5% και στις εκλογές του Μαΐου το 40%.
- Το φαινόμενο της μη-σωστής αναπαραγωγής η και απόκρυψης της ψήφου από τους ερωτώμενους. Για παράδειγμα, οι ψηφοφόροι του ΣΥ.ΡΙΖ.Α στις εκλογές του 2014 είχαν την τάση να θέλουν να δηλώσουν ανοιχτά τι ψήφισαν ενώ οι εκλογείς των υποψηφίων που υποστηρίζονταν από τη ΝΔ ή την ΕΛΙΑ έδειχναν περισσότερο μυστικοπαθείς και συντηρητικοί. Το ίδιο φάνηκε να συνέβη και στις εκλογές του Ιουνίου όπου κάποιοι μπορεί να ψήφισαν τα δύο μεγάλα κόμματα αλλά να είχαν αναστολές να το δηλώσουν ενώ κάποιοι άλλοι μπορεί να ψήφισαν μικρότερα κόμματα και να δήλωσαν ότι ψήφισαν ένα από τα δύο μεγάλα κόμματα. Παρατηρώντας τον Πίνακα 1.1, είναι αποκαλυπτικό της πρωτοφανούς εκλογικής αβεβαιότητας ότι σε σύγκριση με το 17% των εκλογέων στις εκλογές του 2009 που αποφάσισαν για την ψήφο τους το τελευταίο δεκαπενθήμερο ή και την ημέρα των εκλογών, τον Μάιο του 2012 το αντίστοιχο ποσοστό εμφανίσθηκε υπερδιπλάσιο αγγίζοντας το 41%**.**

Πίνακας 1.1: Χρόνος απόφασης ψήφου 2009-2012

	6/2012**	5/2012*	10/2009*
Σήμερα	**14**	**19,5**	10,9
Τις τελευταίες 15 ημέρες*/ Την τελευταία εβδομάδα**	12	20,8	5,8
Εδώ και καιρό*/ Την προτελευταία εβδομάδα/Με την προκήρυξη των εκλογών**	37	31,8	34
Ψηφίζω πάντα το ίδιο	37	24,8	46,2
ΔΑ		3,1	3,1

- Η *ψήφος τακτικής (tactical voting).* Από τον Πίνακα 1.2 παρατηρούμε ότι η νίκη της ΝΔ στις περασμένες βουλευτικές εκλογές οφείλεται σε μεγάλο βαθμό στο φόβο που κυριάρχησε στο εκλογικό σώμα και πήρε τη μορφή της ψήφου τακτικής. Είναι χαρακτηριστικό, ότι 2 στους 10 ψηφοφόρους το 2012 ψήφισαν τη ΝΔ «για να παραμείνουμε στο ευρώ», το 8% των ψηφοφόρων της «για να μην υπάρξει χάος και αναρχία» ενώ το 10% των ψηφοφόρων της δήλωσε ως αιτιολογία για την επιλογή ψήφου ότι θα την ψήφιζε «για να μην κερδίσει ο ΣΥ.ΡΙΖ.Α».

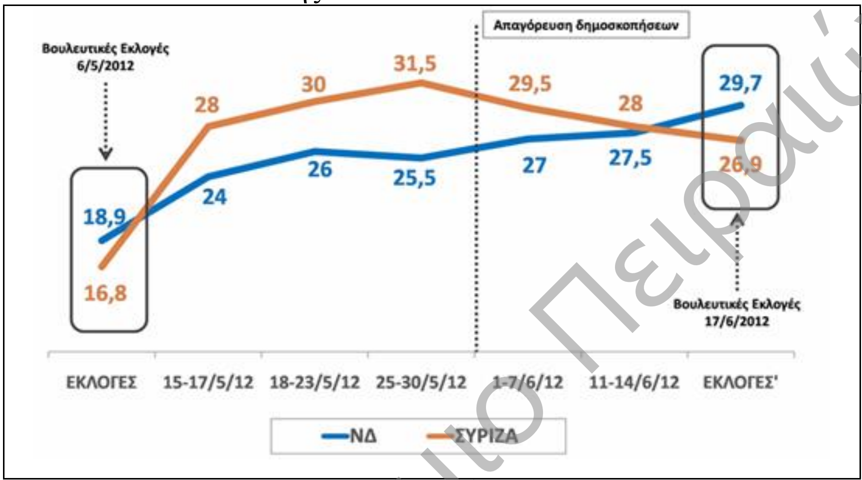
Πίνακας 1.2: Αιτιολόγηση της ψήφου για ΝΔ και ΣΥ.ΡΙΖ.Α



- Η τεράστια και πρωτοφανής άρνηση των ψηφοφόρων να συμμετέχουν στις δημοσκοπήσεις εξόδου των αυτοδιοικητικών εκλογών του 2014, εκφράζοντας με αυτόν τον τρόπο μία αγανακτισμένη συλλογικότητα απέναντι στα κανάλια και στο ίδιο το σύστημα. Η συμπεριφορά αυτή μπορεί να ερμηνευθεί καλύτερα από την *ψυχολογία της μάζας* η οποία εξηγεί το πως απλοί καθημερινοί άνθρωποι αποκτούν ιδιαίτερη δύναμη όταν λειτουργούν συλλογικά. Σύμφωνα με τον Gustave Le Bon, η ιδιαιτερότητα αυτού του είδους της συλλογικής ψυχολογίας είναι ότι δεν απαιτείται πάντα η συγκέντρωση ενός μεγάλου αριθμού ατόμων στο ίδιο σημείο έτσι ώστε να γεννηθεί μία κοινή τάση απέναντι σε κάποιο σημαντικό γεγονός (στη συγκεκριμένη περίπτωση απέναντι στο θεσμό των exit polls). Κάτω από την επήρεια αυτού του γεγονότος χιλιάδες άτομα που δεν έχουν έρθει ποτέ σε επαφή και ζουν σε διαφορετικές περιοχές μπορούν να αποκτήσουν τα χαρακτηριστικά του πλήθους και να επιφέρουν σημαντικές αλλαγές απέναντι στο εκάστοτε σύστημα.
- Τέλος, η απαγόρευση δημοσίευσης κάθε είδους δημοσκόπησης κατά το τελευταίο δεκαπενθήμερο της προεκλογικής περιόδου φάνηκε να έπαιξε με τη σειρά του καθοριστικό ρόλο στην "αποτυχία" του exit poll του 2012 και αυτός ήταν και ο κύριος λόγος που στις αυτοδιοικητικές εκλογές του 2014 η απαγόρευση αυτή δεν εφαρμόστηκε. Όπως υποστήριξαν οι αναλυτές, η επίδραση της προεκλογικής περιόδου στο εκλογικό αποτέλεσμα δεν είναι κάθε φορά η ίδια. Υπάρχουν εκλογικές αναμετρήσεις, όπως του 2004 ή του 2007, στις οποίες αποδείχθηκε ότι σε μεγάλο βαθμό το εκλογικό αποτέλεσμα είχε από καιρό ξεκαθαρίσει και κατά συνέπεια, η επίδραση της προεκλογικής εκστρατείας των πολιτικών κομμάτων δεν διαδραμάτισε ουσιαστικό ρόλο. Στην περίπτωση των

εκλογών της 17ης Ιουνίου συνέβη το ακριβώς αντίθετο. Η δυναμική της πρόσφατης προεκλογικής περιόδου αποδείχθηκε η μεγαλύτερη που έχει παρατηρηθεί ιστορικά μέχρι σήμερα, σαφώς μεγαλύτερη από την αντίστοιχη που παρατηρήθηκε και στις πρώτες εκλογές του Μαΐου. Οι εξελίξεις των δύο τελευταίων εβδομάδων υπήρξαν ραγδαίες, συνεχίστηκαν μέχρι και την ημέρα των εκλογών και επηρέασαν καθοριστικά το αποτέλεσμα όπως φαίνεται και στο Διάγραμμα 1.1.

Διάγραμμα 1.1: Προεκλογική εκτίμηση της εκλογικής επιρροής της ΝΔ και του ΣΥ.ΡΙΖ.Α



1.4 Δομή διπλωματικής εργασίας

Η παρούσα διπλωματική εργασία πραγματεύεται τις στατιστικές μεθόδους που χρησιμοποιούνται για την πρόβλεψη του τελικού εκλογικού αποτελέσματος και τη μετέπειτα ανάλυση των ψηφοφοριών.

Συγκεκριμένα, στο παρόν κεφάλαιο γίνεται μία εισαγωγή στα είδη και στην ιστορική διαδρομή των κυριότερων εκλογικών συστημάτων. Επιπλέον, προσδιορίζεται τι ακριβώς είναι το exit poll, ποια είναι η κύρια διαφορά του από τις δημοσκοπήσεις καθώς και η σημασία του για την πρόβλεψη του εκλογικού αποτελέσματος. Τέλος, γίνεται αναφορά στην αξιοπιστία του exit poll ως εργαλείο μέτρησης της ψήφου.

Στο δεύτερο κεφάλαιο, παρουσιάζονται ορισμένες στατιστικές τεχνικές, πέραν του exit poll, που χρησιμοποιούνται για την πρόβλεψη του τελικού εκλογικού αποτελέσματος καθώς και για την πρόβλεψη της αποχής από την εκλογική διαδικασία. Συγκεκριμένα, γίνεται αναφορά στη μέθοδο της λογιστικής παλινδρόμησης, στη μέθοδο του απλού κινητού μέσου, στη μέθοδο της απλής εκθετικής εξομάλυνσης, στην πρόβλεψη της εκλογικής επιρροής, στην πρόβλεψη με βάση τα πραγματικά εκλογικά αποτελέσματα, καθώς και στη γραμμική παλινδρόμηση.

Στο τρίτο κεφάλαιο, γίνεται αναφορά στην τεχνική της προσομοίωσης και στον τρόπο με τον οποίο μπορεί να χρησιμοποιηθεί από έναν αναλυτή προκειμένου να προβλέψει το τελικό εκλογικό αποτέλεσμα μίας ψηφοφορίας στην οποία εφαρμόστηκε η ταξινομική μέθοδος. Επιπλέον, εκτιμάται ποιο θα πρέπει να είναι το απαιτούμενο μέγεθος του δείγματος που θα πρέπει να συγκεντρωθεί, έτσι ώστε το τελικό εκλογικό αποτέλεσμα να μπορεί να προβλεφθεί με όσο το δυνατόν μεγαλύτερη ακρίβεια.

Στο τελευταίο κεφάλαιο, θα επιδιώξουμε αρχικά να εξετάσουμε κατά πόσο ορισμένες μεταβλητές (π.χ φύλο, ηλικία, επάγγελμα) είναι ικανές να επηρεάσουν στατιστικά σημαντικά την ψήφο του i-ψηφοφόρου ενώ, στη συνέχεια θα κατασκευάσουμε ένα στατιστικό μοντέλο λογιστικής παλινδρόμησης που θα προβλέπει, με βάση τα δεδομένα του exit poll, τα τελικά εκλογικά αποτελέσματα.

Η ανάλυση των δεδομένων έγινε με τη χρήση του στατιστικού πακέτου IBM SPSS Statistics 22.0.

ΚΕΦΑΛΑΙΟ 2

Στατιστικές τεχνικές πρόβλεψης εκλογικών αποτελεσμάτων

2.1 Εισαγωγή

Η μέθοδος του exit poll αποτελεί στον κύκλο των εταιρειών δημοσκοπήσεων τη δημοφιλέστερη μέθοδο πρόβλεψης του τελικού εκλογικού αποτελέσματος. Δεν παύουν όμως να υπάρχουν και άλλες, εξίσου σημαντικές, τεχνικές πρόβλεψης γνωστές για την εγκυρότητα και την αξιοπιστία τους. Παραδείγματα τέτοιων μεθόδων αποτελούν η *λογιστική παλινδρόμηση, η μέθοδος του απλού κινητού μέσου, η μέθοδος της απλής εκθετικής εξομάλυνσης, η πρόβλεψη της εκλογικής επιρροής* και *η πρόβλεψη με βάση τα πραγματικά εκλογικά αποτελέσματα*.

2.2 Λογιστική παλινδρόμηση

Το πιο διαδεδομένο γραμμικό μοντέλο για δίτιμα ή διωνυμικά δεδομένα είναι το μοντέλο της λογιστικής παλινδρόμησης. Η λογιστική συνάρτηση στην οποία βασίζεται παίρνει πάντοτε τιμές στο διάστημα $[0,1]$ και επομένως το μοντέλο χρησιμοποιείται κυρίως για να περιγράψει τη σχέση της πιθανότητας ενός χαρακτηριστικού με διάφορους παράγοντες που πιθανόν το επηρεάζουν. Η διαφορά του με το αντίστοιχο μοντέλο της απλής παλινδρόμησης είναι ότι η μεταβλητή απόκρισης είναι δίτιμη, δηλαδή περιγράφει «επιτυχία-αποτυχία». Συγκεκριμένα, στην περίπτωση των εκλογών περιγράφει την κατάσταση ο i-ψηφοφόρος να «ψήφισε» (επιτυχία) ή να «μην ψήφισε» (αποτυχία) το κάθε κόμμα και από τη στιγμή που θέλουμε να προβλέψουμε ποσοστά θεωρείται το καταλληλότερο μοντέλο παλινδρόμησης. Προτού προχωρήσουμε στην ανάλυση και στην ερμηνεία του λογιστικού μοντέλου, κρίνεται απαραίτητο να αναφερθούμε σε κάποιες βασικές έννοιες της λογιστικής παλινδρόμησης. Εκείνες της *σχετικής πιθανότητας*, του *λόγου σχετικών πιθανοτήτων* και του *σχετικού κινδύνου* τις οποίες θα αναλύσουμε εκτενέστερα στις επόμενες παραγράφους.

2.2.1 Σχετική πιθανότητα

Η έννοια της *σχετικής πιθανότητας (odds)* είναι ιδιαίτερα σημαντική για την ερμηνεία των παραμέτρων σε ένα μοντέλο λογιστικής παλινδρόμησης. Η σχετική πιθανότητα ενός ενδεχομένου A, έστω το ενδεχόμενο "ο i-ψηφοφόρος να ψηφίσει στις εκλογές ΠΑΣΟΚ", ορίζεται ως ο λόγος

$$\text{odds} = \frac{P(A)}{1-P(A)}$$

όπου, $P(A)$ είναι η πιθανότητα να συμβεί το ενδεχόμενο A.

- αν η τιμή της σχετικής πιθανότητας είναι μεγαλύτερη του 1 (*odds > 1*), τότε το ενδεχόμενο *A* που εμφανίζεται στον αριθμητή είναι πιο πιθανό να συμβεί από το να μη συμβεί.
- αν η τιμή της σχετικής πιθανότητας είναι μικρότερη του 1 (*odds < 1*), τότε το ενδεχόμενο *A* είναι λιγότερο πιθανό να συμβεί από το να μη συμβεί.
- αν η τιμή της σχετικής πιθανότητας είναι ίση με 1 (*odds = 1*), τότε το ενδεχόμενο *A* είναι εξίσου πιθανό να συμβεί ή να μη συμβεί, δηλαδή έχει πιθανότητα εμφάνισης (και μη εμφάνισης) 50%.

2.2.2 Λόγος σχετικής πιθανότητας

Έστω ότι έχουμε δύο ενδεχόμενα A και B, όπου A το ενδεχόμενο "ο i-ψηφοφόρος να ψήφισε ΠΑΣΟΚ και να είναι γυναίκα" και B το ενδεχόμενο "ο i-ψηφοφόρος να ψήφισε ΠΑΣΟΚ και να είναι άντρας". Τότε ο *λόγος σχετικών πιθανοτήτων (odds ratio)* του A ως προς το B είναι:

$$\text{OR} = \frac{odds(A)}{odds(B)} = \theta_{\text{AB}} = \frac{\dfrac{P(A)}{1-P(A)}}{\dfrac{P(B)}{1-P(B)}} = \frac{P(A)(1-P(B))}{P(B)(1-P(A))}$$

και εκφράζει πόσες φορές η σχετική πιθανότητα του ενδεχομένου A είναι μεγαλύτερη από τη σχετική πιθανότητα του ενδεχομένου B δηλαδή, πόσες φορές μεγαλύτερη είναι η σχετική πιθανότητα να ψηφίσει ΠΑΣΟΚ κάποια γυναίκα από τη σχετική πιθανότητα να ψηφίσει ΠΑΣΟΚ κάποιος άντρας.

2.2.3 Σχετικός κίνδυνος

Μία έννοια που συνδέεται με το λόγο σχετικής πιθανότητας είναι ο *σχετικός κίνδυνος (relative risk)*. Ο όρος προέρχεται από τη Βιοστατιστική όπου εκεί η μεταβλητή απόκρισης είναι συχνά η εμφάνιση ή όχι κάποιας ασθένειας. Ως σχετικός κίνδυνος ορίζεται η ποσότητα

$$\frac{P(A)}{P(B)}$$

και στην περίπτωσή μας θα μπορούσε να εκφράζει πόσες φορές πιθανότερο είναι μία γυναίκα να ψηφίσει ΠΑΣΟΚ σε σχέση με έναν άντρα. Αν ο σχετικός κίνδυνος ισούται με τη μονάδα, τότε οι δύο πιθανότητες $P(A)$ και $P(B)$ θα είναι ίσες.

Τέλος, να αναφέρουμε ότι ο λόγος σχετικής πιθανότητας και ο σχετικός κίνδυνος συνδέονται μεταξύ τους μέσω της σχέσης

$$\text{OR} = \frac{\dfrac{P(A)}{1-P(A)}}{\dfrac{P(B)}{1-P(B)}} = \text{relative risk}$$

2.2.4 Το μοντέλο

Έστω ένα διάνυσμα $\mathbf{X} = (X_1, X_2, ..., X_k)$ με k-ανεξάρτητες μεταβλητές, οι οποίες εκφράζουν τα χαρακτηριστικά του κάθε ψηφοφόρου που θεωρούμε ότι επηρεάζουν την τελική απόφασή του. Το μοντέλο της λογιστικής παλινδρόμησης, στην περίπτωση που έχουμε μόνο μία ανεξάρτητη μεταβλητή X_1 (απλή λογιστική παλινδρόμηση), περιγράφεται από τη συνάρτηση

$$\text{logit}[P(\mathbf{X})] := \log\left[\frac{P(X_1)}{1-P(X_1)}\right] = a + \beta_1 X_1$$

με πιθανότητα επιτυχίας,

$$P(\mathbf{X}) = \frac{e^{\beta_0+\beta_1 X_1}}{1+e^{\beta_0+\beta_1 X_1}}$$

ενώ στην περίπτωση που έχουμε k-ανεξάρτητες μεταβλητές (πολλαπλή λογιστική παλινδρόμηση) το μοντέλο περιγράφεται από τη συνάρτηση

$$\text{logit}[P(\mathbf{X})] := \log\left[\frac{P(X_1, X_2, \ldots X_k)}{1 - P(X_1, X_2, \ldots, X_k)}\right] = a + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

με πιθανότητα επιτυχίας,

$$P(\mathbf{X}) = \frac{e^{a+\beta_k X_k}}{1 + e^{a+\beta_k X_k}}.$$

Και στις δύο περιπτώσεις το αριστερό μέλος της εξίσωσης περιέχει τις τιμές της εξαρτημένης μεταβλητής με τη μορφή του λογαρίθμου της σχετικής πιθανότητας (*odds*), ενώ το δεξί μέλος της εξίσωσης δημιουργείται από έναν γραμμικό συνδυασμό

$$a + \sum_{i=1}^{k} \beta_i X_i$$

των ανεξάρτητων μεταβλητών $X_1, X_2, \ldots, X_k$ που συμμετέχουν στο μοντέλο της παλινδρόμησης. Το πρόσημο της παραμέτρου a μας δείχνει τη μονοτονία της συνάρτησης, ενώ το μέγεθος της παραμέτρου a καθορίζει το ρυθμό μεταβολής της πιθανότητας επιτυχίας P. Τέλος, η μεταβλητή απόκρισης θα παίρνει την τιμή 1 ("επιτυχία") εάν ο ψηφοφόρος ψήφισε το κόμμα και την τιμή 0 ("αποτυχία") εάν δεν το ψήφισε ενώ οι ανεξάρτητες μεταβλητές $X_1, X_2, \ldots, X_k$ μπορεί να είναι είτε δίτιμες είτε συνεχείς.

Ένας από τους κύριους λόγους που το μοντέλο λογιστικής παλινδρόμησης προτιμάται από τα υπόλοιπα ΓΓΜ είναι η ευκολότερη διαισθητική ερμηνεία των αποτελεσμάτων με βάση τη σχετική πιθανότητα.

Υπενθυμίζουμε ότι:

$$\text{logit}[P(\mathbf{X})] := \log\left[\frac{P(X_1, X_2, \ldots, X_k)}{1 - P(X_1, X_2, \ldots, X_k)}\right] = a + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k.$$

Η συνάρτηση logit αναφέρεται στο λογάριθμο της σχετικής πιθανότητας του ενδεχομένου που μας ενδιαφέρει και μπορεί να γραφτεί ισοδύναμα ως εξής:

$$odds = \frac{P(X_1, X_2, ..., X_k)}{1 - P(X_1, X_2, ..., X_k)} = \exp(a + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k).$$

Αν θεωρήσουμε ένα μοντέλο με δύο μεταβλητές

$$\log\left[\frac{P(X_1, X_2)}{1 - P(X_1, X_2)}\right] = a + \beta_1 X_1 + \beta_2 \mathrm{X}_2$$

τότε,

- αύξηση της τιμής του X_1 κατά μία μονάδα μεταβάλλει την τιμή του λογαρίθμου της σχετικής πιθανότητας κατά β_1,
- αύξηση της τιμής του X_1 κατά μία μονάδα προκαλεί πολλαπλασιαστική αύξηση της σχετικής πιθανότητας κατά e^{β_1} όταν η μεταβλητή X_2 παραμένει σταθερή.

2.2.5 Εκτίμηση των παραμέτρων του μοντέλου

Έστω το logit μοντέλο με k ερμηνευτικές μεταβλητές. Σύμφωνα με τη μέθοδο της μέγιστης πιθανοφάνειας, η εκτίμηση των παραμέτρων του μοντέλου, δηλαδή το διάνυσμα των παραμέτρων $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_k)$ προκύπτει είτε μεγιστοποιώντας τη συνάρτηση πιθανοφάνειας

$$L\ (p; y) = \exp\left[\sum_{i=1}^{n} y_i \log\frac{p_i}{1 - p_i} + \sum_{i=1}^{n} \log(1 - p_i)\right] + c_n(y)$$

όπου $p = (p_1, ..., p_n)'$, $y = (y_1, ..., y_n)'$ και η συνάρτηση c_n δεν εξαρτάται από τις παραμέτρους p_i ή ισοδύναμα, μεγιστοποιώντας το λογάριθμο της συνάρτησης πιθανοφάνειας

$$l(p; y) = \sum_{i=1}^{n} y_i \log\frac{p_i}{1 - p_i} + \sum_{i=1}^{n} \log(1 - p_i).$$

2.2.6 Έλεγχοι υποθέσεων για τις παραμέτρους του μοντέλου

Για να ελέγξουμε τη σημαντικότητα μιας μεταβλητής σε ένα μοντέλο λογιστικής παλινδρόμησης

$$\text{logit}[P(\mathbf{X})] := \log\left[\frac{P(X_1, X_2, ..., X_k)}{1 - P(X_1, X_2, ..., X_k)}\right] = a + \sum \beta_i X_i$$

χρησιμοποιούμε δύο ελέγχους, τον *έλεγχο λόγου πιθανοφανειών* και τον *έλεγχο Wald*.

α) Έλεγχος λόγου πιθανοφανειών

Θεωρούμε ένα μοντέλο M_1 (*προσαρμοσμένο μοντέλο*) του οποίου οι μεταβλητές είναι υποσύνολο ενός άλλου μοντέλου M_2 (*κορεσμένο μοντέλο*). Επιπλέον, έστω $L(M_1)$ και $L(M_2)$ οι αντίστοιχες συναρτήσεις πιθανοφάνειας των δύο μοντέλων. Ο έλεγχος του λόγου πιθανοφανειών γίνεται με τη βοήθεια της στατιστικής συνάρτησης

$$D_1 - D_2 = -2(\log L(M_1) - \log(M_2))$$

η οποία ακολουθεί (ασυμπτωτικά) την κατανομή χι-τετράγωνο με βαθμούς ελευθερίας df ίσους με τη διαφορά του αριθμού των παραμέτρων ανάμεσα στα δύο μοντέλα. Η μηδενική υπόθεση απορρίπτεται για μεγάλες τιμές της στατιστικής συνάρτησης.

Μία έννοια που χρησιμοποιείται συχνά στη λογιστική παλινδρόμηση είναι αυτή της *απόκλισης (deviance)*. Πρόκειται για τη γενίκευση της έννοιας του αθροίσματος των τετραγώνων των καταλοίπων και προκύπτει από τον έλεγχο του λόγου πιθανοφανειών. Έστω ότι έχουμε δύο εμφωλευμένα μοντέλα M_1 και M_2, όπου το σύνολο των επεξηγηματικών μεταβλητών του M_1 είναι υποσύνολο αυτών του M_2. Ως απόκλιση του μοντέλου M_1 ορίζεται η τιμή της ποσότητας της στατιστικής συνάρτησης

$$-2(\log L(M_1) - \log(M_2))$$

όταν το μοντέλο M_2 είναι το κορεσμένο, δηλαδή ένα μοντέλο που έχει τόσες παραμέτρους όσα και τα δεδομένα. Η απόκλιση θα μπορούσε να χρησιμοποιηθεί σαν μία στατιστική συνάρτηση ελέγχου για την υπόθεση

H_0 : Το προσαρμοσμένο μοντέλο δε διαφέρει σημαντικά από το κορεσμένο μοντέλο

H_1 : Το προσαρμοσμένο μοντέλο διαφέρει σημαντικά από το κορεσμένο μοντέλο

Διαισθητικά, όσο πιο μικρή είναι η απόκλιση ενός μοντέλου, τόσο πιο κοντά είναι στο κορεσμένο μοντέλο και αυτό παρέχει ένδειξη καλής προσαρμογής.

β) Έλεγχος Wald

Από την θεωρία εκτιμητικής, οι ε.μ.π. για τις παραμέτρους ενός γενικευμένου γραμμικού μοντέλου ακολουθούν ασυμπτωτικά την κανονική κατανομή. Συνεπώς, για μεγάλο αριθμό παρατηρήσεων μπορούμε να χρησιμοποιήσουμε την κανονική κατανομή για να εξετάσουμε αν κάποια παράμετρος β_i διαφέρει σημαντικά από το μηδέν. Ο έλεγχος Wald γίνεται με τη βοήθεια της στατιστικής συνάρτησης

$$z = \frac{\hat{\beta}_i}{s(\hat{\beta}_i)}$$

η οποία ακολουθεί (ασυμπτωτικά) την κατανομή N(0,1). Η μηδενική υπόθεση απορρίπτεται σε ε.σ a αν

$$|z| > z_{\alpha/2}.$$

Ο έλεγχος του Wald και ο έλεγχος του λόγο πιθανοφανειών συνήθως δίνουν τα ίδια αποτελέσματα. Όταν πρόκειται όμως για μικρά δείγματα, τα αποτελέσματα πιθανόν και να διαφέρουν. Στην περίπτωση αυτή, προτιμάται η χρήση του ελέγχου με το λόγο των πιθανοφανειών.

2.3 Χρονοσειρές

Ο όρος χρονοσειρά σημαίνει μία σειρά από παρατηρήσεις που λαμβάνονται σε ορισμένες χρονικές στιγμές ή περιόδους και ισαπέχουν μεταξύ τους. Αν με Y συμβολίζεται η μεταβλητή που μελετάται και χρησιμοποιηθεί ο δείκτης t για να δηλώσει τη χρονική περίοδο που ελήφθη η τιμή αυτή, τότε η ακολουθία των τιμών αυτών αποτελεί ένα δείγμα N παρατηρήσεων της χρονολογικής σειράς. Βασικό χαρακτηριστικό κάθε χρονολογικής σειράς είναι η εξάρτηση μεταξύ των διαδοχικών τιμών της. Οι πιο δημοφιλείς μέθοδοι χρονοσειρών για την πρόβλεψη των εκλογικών αποτελεσμάτων είναι εκείνες του απλού κινητού μέσου και της απλής εκθετικής εξομάλυνσης.

2.3.1 Η μέθοδος του απλού κινητού μέσου

Είναι μία μέθοδος προβλέψεων που χρησιμοποιεί ως πρόβλεψη την τιμή του αριθμητικού μέσου όρου των m πλέον πρόσφατων παρατηρήσεων της διαθέσιμης χρονοσειράς των ερευνών. Ειδικότερα, οι προβλέψεις μίας χρονοσειράς Y_t για $t = 1,2,...,n$ δημιουργούνται με τη μέθοδο αυτή ως εξής:

$$\hat{Y}_{t+1} = \frac{1}{m}(Y_t + Y_{t-1} + ... + Y_{t-m+1}),$$

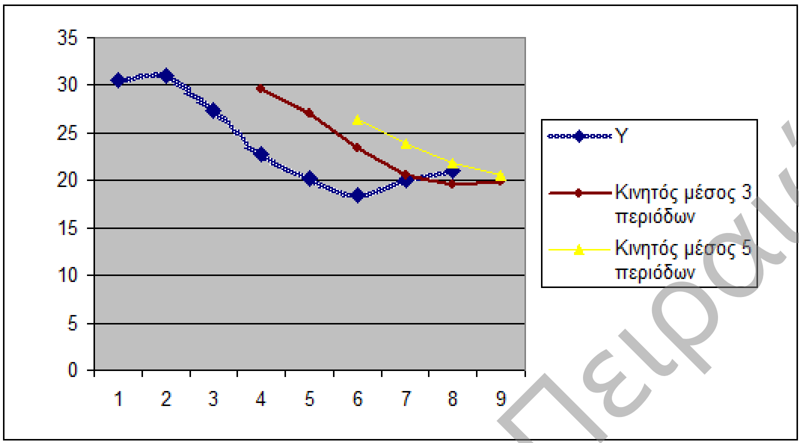
όπου $\hat{Y}_{t+1}$ είναι η πρόβλεψη για την περίοδο $t+1$ και m ο αριθμός των περιόδων που χρησιμοποιούνται για τον υπολογισμό της τιμής του μέσου όρου. Στον Πίνακα 2.1, παρουσιάζεται η διευκρινισμένη πρόθεση ψήφου της ΝΔ για τους τελευταίους οκτώ μήνες πριν από τις βουλευτικές εκλογές του Ιουνίου του 2012 (Οκτώβριος-Μάιος 2012), οι προβλέψεις για τον ένατο μήνα (Ιούνιος 2012) καθώς και οι αντίστοιχες τιμές των σφαλμάτων της πρόβλεψης με τη μέθοδο του απλού κινητού μέσου τριών και πέντε περιόδων. Επειδή η τιμή του MSE για m=3 είναι μικρότερη από την τιμή του MSE για m=5, η μέθοδος του απλού κινητού μέσου τριών περιόδων είναι πιο κατάλληλη για τη δημιουργία προβλέψεων της συγκεκριμένης χρονοσειράς από τη μέθοδο του απλού κινητού μέσου πέντε περιόδων. Άρα, καλύτερη πρόβλεψη θα είναι το 19.87%.

Πίνακας 2.1: Δεδομένα και προβλέψεις με τη μέθοδο του απλού κινητού μέσου για m =3 και m =5

		m=3		m=5	
t	Y	Y^	Τετραγωνικό σφάλμα πρόβλεψης e^2	Y^	Τετραγωνικό σφάλμα πρόβλεψης e^2
1	30,5				
2	31				
3	27,3				
4	22,7	29,60	47,61		
5	20,2	27,00	46,24		
6	18,5	23,40	24,01	26,34	61,47
7	20,1	20,47	0,13	23,94	14,75
8	21	19,60	1,96	21,76	0,58
9		**19,87**	**119,95**	**20,50**	**76,79**
		MSE(m=3)	23,99	MSE(m=5)	25,60

Στο Διάγραμμα 2.1 παρουσιάζεται η εξομάλυνση των τιμών της χρονοσειράς με τη μέθοδο του απλού κινητού μέσου τριών και πέντε περιόδων για τους μήνες Ιανουάριο έως Ιούνιο και Μάρτιο έως Ιούνιο αντίστοιχα καθώς και η πρόβλεψη του ποσοστού της ΝΔ για τον μήνα Ιούνιο.

Διάγραμμα 2.1



2.3.2 Η μέθοδος της απλής εκθετικής εξομάλυνσης

Οι προβλέψεις δημιουργούνται με βάση έναν σταθμισμένο μέσο όρο έτσι ώστε να δίνεται διαφορετική βαρύτητα σε κάθε παρατήρηση. Με τη μέθοδο αυτή, δίνεται πολύ μεγαλύτερη βαρύτητα στις πιο πρόσφατες παρατηρήσεις από αυτή που δίνεται στις πιο απομακρυσμένες. Οι προβλέψεις μίας χρονοσειράς Y_t για $t = 1,2,...,n$ δημιουργούνται με τη μέθοδο αυτή ως εξής:

$$\hat{Y}_{t+1} = aY_t + a(1-a)Y_{t-1} + a(1-a)^2 Y_{t-2} + \dots,$$

όπου a είναι η σταθερά εξομάλυνσης η οποία λαμβάνει τιμές μεταξύ 0 και 1 και για την οποία ισχύουν τα εξής:

- όσο πιο μεγάλη είναι η τιμή της παραμέτρου a τόσο μεγαλύτερη βαρύτητα δίνεται στις πιο πρόσφατες παρατηρήσεις και πολύ μικρή έως μηδαμινή βαρύτητα στις απομακρυσμένες,
- όσο μεγαλύτερη είναι η τιμή του a τόσο μεγαλύτερη βαρύτητα δίνεται στο σφάλμα της πρόβλεψης,

- Για $a=1$ η πρόβλεψη της περιόδου $t+1$ είναι η πραγματική τιμή της περιόδου t κάτι που ισχύει και για τη μέθοδο του απλού κινητού μέσου για $m=1$,
- Για $a=0$ η πρόβλεψη της περιόδου $t+1$ είναι ίση με την πρόβλεψη της περιόδου t.

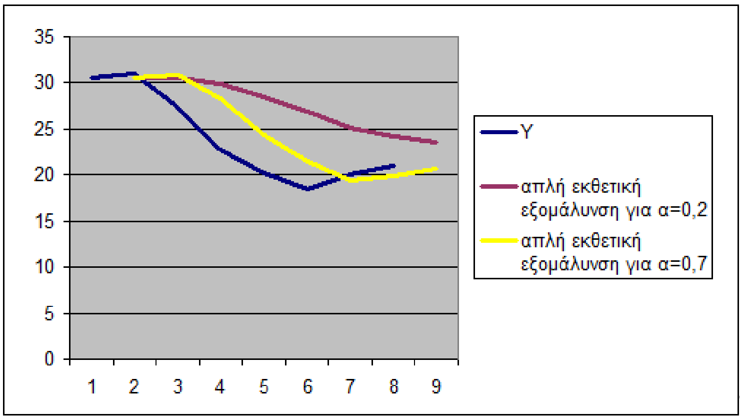
Τέλος, αξίζει να σημειωθεί ότι οι παρατηρήσεις της χρονοσειράς εξομαλύνονται περισσότερο από μικρές τιμές της παραμέτρου a παρά από μεγάλες. Αντίθετα, για μεγάλες τιμές της παραμέτρου η μέθοδος αυτή αντιδρά πιο γρήγορα στις πραγματικές τιμές των παρατηρήσεων της χρονοσειρας δημιουργώντας μεγάλη αστάθεια στις προβλέψεις. Στον Πίνακα 2.2, παρουσιάζονται τα δεδομένα και οι προβλέψεις των ποσοστών της ΝΔ για τις βουλευτικές εκλογές του Ιουνίου του 2012 με την μέθοδο της απλής εκθετικής εξομάλυνσης καθώς και οι αντίστοιχες τιμές των σφαλμάτων της πρόβλεψης για $a=0{,}2$ και για $a=0{,}7$. Επειδή η τιμή του MSE για $a=0{,}7$ είναι μικρότερη από την τιμή του MSE για $a=0{,}2$, η μέθοδος της απλής εκθετικής εξομάλυνσης αναμένεται να δώσει καλύτερες προβλέψεις για $a=0{,}7$, με την προϋπόθεση ότι θα ισχύουν στο μέλλον οι ίδιες συνθήκες που ίσχυαν και στο παρελθόν.

Πίνακας 2.2: Δεδομένα και προβλέψεις με τη μέθοδο της απλής εκθετικής εξομάλυνσης για $a=0{,}2$ και $a=0{,}7$

		0,2		0,7	
t	Y	Y^	Τετραγωνικό σφάλμα πρόβλεψης e^2	Y^	Τετραγωνικό σφάλμα πρόβλεψης e^2
1	30,5				
2	31	30,500	0,250	30,500	0,250
3	27,3	30,600	10,890	30,850	12,603
4	22,7	29,940	52,418	28,365	32,092
5	20,2	28,492	68,757	24,400	17,636
6	18,5	26,834	69,449	21,460	8,761
7	20,1	25,167	25,673	19,388	0,507
8	21	24,154	9,945	19,886	1,240
9		**23,523**	**237,382**	**20,666**	**73,088**
		MSE (α=0,2)	**33,912**	**MSE (α=0,8)**	**10,441**

Στο Διάγραμμα 2.2 παρουσιάζεται η εξομάλυνση των τιμών της χρονοσειράς με τη μέθοδο της απλής εκθετικής εξομάλυνσης για α=0,2, για α=0,7 για τους μήνες Ιανουάριο έως Ιούνιο και Μάρτιο έως Ιούνιο αντίστοιχα καθώς και η πρόβλεψη του ποσοστού της ΝΔ για τον μήνα Ιούνιο.

Διάγραμμα 2.2



2.4 Πρόβλεψη της εκλογικής επιρροής

Θεωρείται μία από τις πιο συνηθισμένες μεθοδολογίες ανά τον κόσμο, ενώ για πρώτη φορά εφαρμόστηκε στην Ελλάδα στις εκλογές του 2004 από την εταιρεία δημοσκοπήσεων Public Issue. Πιο αναλυτικά, με τον όρο πρόβλεψη της εκλογικής επιρροής εννοούμε τον επανυπολογισμό των ποσοστών των κομμάτων, έτσι ώστε τα αποτελέσματα της έρευνας να μην συμπεριλαμβάνουν τις "αδιευκρίνιστες" απαντήσεις των ερωτώμενων (πχ. «αναποφάσιστος/η», «δεν απαντώ») και να είναι συγκρίσιμα με τα πραγματικά εκλογικά αποτελέσματα. Αυτό συμβαίνει διότι η λεγόμενη "αδιευκρίνιστη" ψήφος αποτελεί απλά μία ιδεολογική κατασκευή των δημοσκοπήσεων και δεν καταγράφεται ποτέ στην κάλπη των εκλογών. Επιπλέον, η κατανομή της "αδιευκρίνιστης" ψήφου που προκύπτει δεν είναι αναλογική (απλή απαλοιφή), αλλά πραγματοποιείται με τη χρήση στατιστικών υποδειγμάτων. Κύριο χαρακτηριστικό της μεθοδολογίας αυτής, αποτελεί η κατάργηση της στάθμισης του δείγματος με την προηγούμενη ψήφο (το δείγμα σταθμίζεται συγκρίνοντας τη δειγματοληπτική προηγούμενη ψήφο με το 100% της πραγματικής προηγούμενης ψήφου ή με κάποιο κλάσμα της) και η αντικατάστασή της με τη μεθοδολογία της ανάλυσης χρονολογικών σειρών της διαθέσιμης χρονοσειράς των ερευνών. Η πολιτική στάθμιση έχει εγκαταλειφθεί εδώ και χρόνια από τις περισσότερες μεγάλες εταιρείες παγκοσμίως (Gallup, MORI και η Communicate Research) καθώς όπως χαρακτηριστικά δήλωσε ο διευθύνων σύμβουλος της εταιρείας ICM της Μ.Βρετανίας N.Sparrow, όλο και λιγότεροι ψηφοφόροι δηλώνουν τι ψήφισαν στις προηγούμενες εκλογές. Τέλος, πρέπει να τονιστεί ότι η εκτίμηση της εκλογικής επιρροής που δίνεται σε κάθε έρευνα δεν αποτελεί μια «εκ των προτέρων πρόβλεψη του τελικού εκλογικού αποτελέσματος», αλλά εκτίμηση της εκλογικής επιρροής των κομμάτων, κατά τη χρονική στιγμή διεξαγωγής της έρευνας.

Μία από τις βασικότερες μεθόδους πρόβλεψης της εκλογικής επιρροής, αποτελεί ο συνδυασμός της ανάλυσης των βραχυχρόνιων και των μακροχρόνιων τάσεων της πρόθεσης ψήφου. Η μεθοδολογία αυτή, την οποία θα αναλύσουμε στη συνέχεια, βασίζεται στο μέσο όρο των βραχυχρόνιων και των μακροχρόνιων εκτιμήσεων.

Πριν ξεκινήσουμε την ανάλυση, κρίνεται απαραίτητο να γίνει μία αναφορά σε δύο βασικά εργαλεία της μεθοδολογίας. Τη μεθοδολογία *Box-Jenkins* και την τεχνική *Kalman*.

2.4.1 Η μεθοδολογία Box-Jenkins (ARIMA(p,d,q))

Η μεθοδολογία Box-Jenkins είναι μία δυναμική και αποτελεσματική μέθοδος στην ανάλυση χρονολογικών σειρών. Το γενικό υπόδειγμα παρουσιάστηκε από τους Box και Jenkins το 1976 και περιλαμβάνει τρεις τύπους παραμέτρων:

- τις αυτοπαλίνδρομες παραμέτρους (p),
- τον αριθμό των “διαφορών” (d),
- τις παραμέτρους κινητού μέσου (q).

Τα υποδείγματα ARIMA παρουσιάζονται στη μορφή ARIMA(p,d,q). Για παράδειγμα, ένα υπόδειγμα της μορφής (0,1,2) δηλώνει ότι δεν περιλαμβάνει αυτοπαλίνδρομη παράμετρο, περιλαμβάνει δύο παραμέτρους κινητών μέσων, ενώ έχουμε πάρει και τις πρώτες διαφορές της σειράς. Η μεθοδολογία Box-Jenkins απαιτεί οι χρονολογικές σειρές να είναι στάσιμες, δηλαδή να έχουν σταθερό μέσο όρο, σταθερή διακύμανση και σταθερή αυτοσυσχέτιση στη διάρκεια του χρόνου. Σε διαφορετική περίπτωση κάνουμε χρήση των διαφορών d τάξης (συνήθως πρώτης ή δεύτερης).

Οι Box-Jenkins παρουσιάζουν διάφορα υποδείγματα τόσο για αναλυτικούς όσο και για προβλεπτικούς σκοπούς προτείνοντας τρία πρωταρχικά στάδια για την κατασκευή των υποδειγμάτων αυτών. Το πρώτο στάδιο είναι το στάδιο της ταυτοποίησης, όπου σκοπός του ερευνητή είναι να εξειδικεύσει τη δομή του υποδείγματος αφού πρώτα καθοριστεί εάν η χρονολογική σειρά είναι στάσιμη, εάν δηλαδή είναι αμετάβλητη αναφορικά με το χρόνο. Στο δεύτερο στάδιο, γίνεται η εκτίμηση των συντελεστών του υποδείγματος που προέκυψε από το στάδιο της ταυτοποίησης ενώ στο τρίτο στάδιο, γίνεται ο διαγνωστικός έλεγχος του υποδείγματος κατά τον οποίο ο ερευνητής πρέπει να επιλέξει τους στατιστικά σημαντικούς εκτιμητές, να αποδεχτεί τη μορφή του υποδείγματος και να εξετάσει την προβλεπτική ικανότητά του.

2.4.2 Η τεχνική Kalman

Είναι μία τεχνική που μειώνει την επίδραση του δειγματοληπτικού σφάλματος σε κάθε έρευνα και επιτρέπει τη διάκριση της πραγματικής μεταβολής από το δειγματοληπτικό σφάλμα στη διαχρονική εξέλιξη των ποσοστών ενός κόμματος. Η τεχνική Kalman, ακολουθεί τρία βήματα:

α) Αν το αληθινό ποσοστό ενός κόμματος είναι ξ και το παρατηρούμενο ποσοστό μέσω μίας δημοσκόπησης είναι X, τότε

$$X = \xi + e,$$

όπου e το δειγματοληπτικό σφάλμα. Η αληθινή μεταβολή του ξ θεωρείται ότι περιγράφεται από ένα αυτοπαλίνδρομο AR(1) υπόδειγμα[(1)] της μορφής

$$\xi_t = a + \gamma \xi_{t-1} + u_t .$$

Αρχικά, εκτιμώνται με τη μέθοδο της μέγιστης πιθανοφάνειας η παράμετρος a που εκφράζει την τάση της χρονοσειράς και η παράμετρος γ που εκφράζει το κατά πόσο συσχετισμένη είναι η τωρινή με την προηγούμενη τιμή. Μετά την πρόβλεψη των παραμέτρων, τα κενά της σειράς (missing values) συμπληρώνονται με τις προβλέψεις που δίνει το παραπάνω υπόδειγμα.

β) Kalman filtering. Στο βήμα αυτό οι τιμές της σειράς φιλτράρονται έτσι ώστε να μειωθεί η επίδραση του δειγματοληπτικού σφάλματος. Το φιλτράρισμα των τιμών της σειράς γίνεται

[(1)]Η γενική μορφή ενός αυτοπαλίνδρομου υποδείγματος AR(p) είναι $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} ... + \beta_p Y_{t-p} + u_t$. Πρόκειται για ένα υπόδειγμα παλινδρόμησης όπου η εξαρτημένη μεταβλητή Y_t εξαρτάται αποκλειστικά και μόνο από τις τιμές που παίρνει η ίδια η μεταβλητή Y_t στις περιόδους από $t-1$ έως $t-p$. Γι' αυτό και ονομάζεται αυτοπαλίνδρομο υπόδειγμα τάξης p και συμβολίζεται με AR(p).

δημιουργώντας μια νέα σειρά με αφετηρία την πρώτη δημοσκόπηση. Οι τιμές της φιλτραρισμένης σειράς συνιστούν ένα σταθμισμένο άθροισμα της προηγούμενης φιλτραρισμένης τιμής και της παρούσας δημοσκόπησης. Το βάρος της παρούσας δημοσκόπησης γίνεται όλο και μικρότερο όσο μεγαλώνει το δειγματοληπτικό της σφάλμα. Αντίστοιχα, το βάρος της παρούσας δημοσκόπησης γίνεται όλο και μεγαλύτερο, όσο μικραίνει το δειγματοληπτικό της σφάλμα. Έτσι, δημιουργείται διαδοχικά η σειρά των φιλτραρισμένων ποσοστών. Τα κενά της σειράς (missing values) που συμπληρώθηκαν, θεωρούνται ότι προέρχονται από έρευνες με πολύ μικρό μέγεθος δείγματος, ώστε να έχουν πολύ μεγάλο δειγματοληπτικό σφάλμα και κατ'αυτόν τον τρόπο να έχουν πολύ μικρό βάρος στη δημιουργία των φιλτραρισμένων ποσοστών.

γ) Kalman smoothing. Στο βήμα αυτό ακολουθείται η διαδικασία της εξομάλυνσης η οποία στηρίζεται στις φιλτραρισμένες τιμές. Η διαδικασία αυτή είναι αναδρομική, ξεκινά δηλαδή από την τελευταία φιλτραρισμένη τιμή και πηγαίνοντας προς τα πίσω εξομαλύνει τις φιλτραρισμένες τιμές. Η εξομάλυνση γίνεται με τέτοιο τρόπο, ώστε όσο χειρότερα μια φιλτραρισμένη τιμή προβλέπει την επόμενή της, τόσο μεγαλύτερη να είναι και η εξομάλυνση που δέχεται. Στην περίπτωσή μας, υποθέτοντας ότι το αληθινό ποσοστό ξ ακολουθεί το υπόδειγμα του τυχαίου περίπατου χωρίς τάση(2), η παραπάνω τεχνική εφαρμόζεται συνήθως θέτοντας a=0 και γ=1.

2.4.3 Η ανάλυση των βραχυχρόνιων τάσεων

Ο όρος βραχυχρόνιες τάσεις στην διαχρονική εξέλιξη της εκλογικής επιρροής ενός κόμματος, περιγράφει τις τάσεις που διαμορφώνονται κατά τη διάρκεια ενός εκλογικού κύκλου. Για τη μελέτη των βραχυχρόνιων τάσεων της επιρροής κάθε κόμματος χρησιμοποιείται η χρονολογική σειρά των διευκρινισμένων απαντήσεων στην ερώτηση πρόθεσης ψήφου όπως αυτές καταγράφηκαν στις μηνιαίες έρευνες του τρέχοντος εκλογικού κύκλου (π.χ για την πρόβλεψη του 2014, από τις εκλογές του 2012 και ύστερα).

(2)Τυχαίος περίπατος είναι η απλούστερη μορφή μίας μη στάσιμης ως προς το μέσο χρονολογικής σειράς. Προκύπτει από τη μορφή των ARIMA υποδειγμάτων για $p = q = 0$. Αν $c \neq 0$, τότε ο τυχαίος περίπατος είναι χωρίς τάση δηλαδή, η επόμενη τιμή της είναι η προηγούμενη συν έναν τυχαίο παράγοντα. Αν $c = 0$, τότε ο τυχαίος περίπατος είναι με τάση δηλαδή, η επόμενη τιμή της έχει μία σταθερή αύξηση ή μείωση σε σχέση με την προηγούμενη.

Η ανάλυση των βραχυχρόνιων τάσεων γίνεται ως εξής: Αρχικά, εφαρμόζεται στη σειρά πρόθεσης ψήφου κάθε κόμματος η τεχνική εξομάλυνσης Kalman και στη συνέχεια, οι σειρές των εξομαλυσμένων πλέον ποσοστών αναλύονται με τη μεθοδολογία Box-Jenkins. Τα υποδείγματα χρονολογικών σειρών που κατασκευάζονται είναι συνδυασμός ενός ARIMA υποδείγματος και ψευδομεταβλητών (dummy variables) οι οποίες, για την Ελλάδα, περιγράφουν τις εκλογικές αναμετρήσεις από το 2004 και ύστερα και είναι της μορφής

$$(1-B)^d X_t = c + (1-B)^d dummy\_variables + \frac{(1-\theta_1 B - ... - \theta_q B^q)\varepsilon_t}{(1-\phi_1 B - ... - \phi_p B^p)}.$$

2.4.4 Η ανάλυση των μακροχρόνιων τάσεων

Ο όρος μακροχρόνιες τάσεις στη διαχρονική εξέλιξη της εκλογικής επιρροής ενός κόμματος, περιγράφει τις τάσεις που διαμορφώνονται σε ένα χρονικό διάστημα μεγαλύτερο του ενός εκλογικού κύκλου. Για τη μελέτη των μακροχρόνιων τάσεων της επιρροής κάθε κόμματος χρησιμοποιείται η χρονολογική σειρά των διευκρινισμένων απαντήσεων στην ερώτηση πρόθεσης ψήφου σε όλες τις έρευνες της εταιρείας μέχρι και σήμερα.

Οι σειρές της πρόθεσης ψήφου κατασκευάζονται σε τριμηνιαία βάση, παίρνοντας ως τιμή για κάθε τρίμηνο το μέσο όρο των ερευνών του τριμήνου. Στη συνέχεια, αναλύονται με τη μεθοδολογία Box-Jenkins. Τα υποδείγματα χρονολογικών σειρών που κατασκευάζονται είναι και πάλι της μορφής

$$(1-B)^d X_t = c + (1-B)^d dummy\_variables + \frac{(1-\theta_1 B - ... - \theta_q B^q)\varepsilon_t}{(1-\phi_1 B - ... - \phi_p B^p)}.$$

δηλαδή, συνδυασμός ενός ARIMA υποδείγματος και ψευδομεταβλητών (dummy variables), με την μόνη διαφορά ότι τώρα οι ψευδομεταβλητές θα περιγράφουν τα έκτακτα γεγονότα που επηρέασαν την εξέλιξη της επιρροής των κομμάτων σε ολόκληρη την υπό μελέτη περίοδο δηλαδή, από το 1996 έως σήμερα. Να σημειωθεί ότι ανάμεσα στα έκτακτα γεγονότα συμπεριλαμβάνονται και όλες οι εκλογικές αναμετρήσεις από το 1996 και ύστερα.

Κλείνοντας, να σημειώσουμε ότι το περιθώριο σφάλματος της πρόβλεψης προκύπτει από τα υποδείγματα χρονολογικών σειρών με τα οποία γίνεται η εκτίμηση (ARIMA υποδείγματα και εξομάλυνση Kalman) και σε καμία περίπτωση δεν ταυτίζεται με το δειγματοληπτικό σφάλμα. Το περιθώριο σφάλματος της πρόβλεψης επηρεάζεται από πολλούς παράγοντες, όπως το μέγεθος του δείγματος, το ποσοστό του κόμματος στη διευκρινισμένη ψήφο, την καλή προσαρμογή των υποδειγμάτων στη χρονοσειρά των δεδομένων καθώς και το συγκεκριμένο ARIMA υπόδειγμα που επιλέγεται.

2.5 Πρόβλεψη με βάση τα εκλογικά αποτελέσματα

Εφαρμόστηκε για πρώτη φορά στην Ελλάδα στις βουλευτικές εκλογές του 2012 για την πρόβλεψη του εκλογικού αποτελέσματος και της αποχής από την εκλογική διαδικασία. Η συγκεκριμένη μέθοδος πρόβλεψης στηρίζεται στα πρώτα εκλογικά αποτελέσματα, όπως αυτά ανακοινώνονται ανά εκλογικό τμήμα. Το ακριβές δειγματοληπτικό σχήμα που χρησιμοποιείται, είναι μια *δισταδιακή κατά συστάδες δειγματοληψία*(3), συνδυασμένη με *στρωματοποίηση*(4). Συστάδες αποτελούν οι 1034 Καποδιστριακοί δήμοι ενώ υποσυστάδες αποτελούν τα εκλογικά τμήματα εντός του κάθε δήμου. Το σύνολο των δήμων χωρίζεται σε τρία στρώματα, με βάση τον πληθυσμό τους. Η μεθοδολογία που ακολουθείται είναι η εξής:

- το σύνολο της επικράτειας χωρίζεται σε 1034 συστάδες που αποτελούνται από τους 1034 Καποδιστριακούς δήμους
- αντιστοιχώντας τα εκλογικά τμήματα στους δήμους που ανήκουν, δημιουργείται ένα δείγμα από το σύνολο των Καποδιστριακών δήμων της χώρας
- γνωρίζοντας τα αποτελέσματα των παραπάνω δήμων στις προηγούμενες εκλογές μπορούν να προβλεφθούν και τα αποτελέσματα των νέων εκλογών με τη βοήθεια υποδειγμάτων γραμμικής παλινδρόμησης

Η μέθοδος δίνει πολύ ακριβή αποτελέσματα ακόμα και όταν έχουν ανακοινωθεί τα αποτελέσματα μόλις του 5% του συνόλου των εκλογικών τμημάτων. Οι προβλέψεις που δόθηκαν τόσο τον Μάιο όσο και τον Ιούνιο του 2012, αποδείχθηκαν απολύτως επιτυχείς και παρουσιάζονται στον Πίνακα 2.3

(3)Μία μέθοδος δειγματοληψίας η οποία επιλέγει το δείγμα της από το σύνολο των συστάδων στις οποίες χωρίζεται ο πληθυσμός. Η κάθε μία από τις συστάδες που επιλέξαμε, ιδίως αν το μέγεθός της είναι μεγάλο, μπορεί να χωριστεί σε υποσυστάδες και από το σύνολό τους να επιλεγεί ένα τυχαίο δείγμα υποσυστάδων για κάθε συστάδα που επιλέξαμε κ.ο.κ. Όταν η διαδικασία επαναλαμβάνεται δύο φορές, τότε μιλάμε για δισταδιακή δειγματοληψία.

(4)Μία μέθοδος δειγματοληψίας που επιλέγει το δείγμα της από υποπληθυσμούς–στρώματα στα οποία διαμερίζεται ο πληθυσμός, ομοιογενή ως προς κάποιο χαρακτηριστικό.

Πίνακας 2.3: Πρόβλεψη την ημέρα των εκλογών με βάση τη ροή των εκλογικών αποτελεσμάτων

	Ιούνιος (20:40, 16% ΕΚΛ.ΤΜ)			Μάιος (21:10, 6% ΕΚΛ.ΤΜ)		
Κόμμα	Πρόβλεψη PI με βάση εκλογικά αποτελέσματα	Τυπικό σφάλμα	Εκλογικό αποτέλεσμα	Πρόβλεψη PI με βάση εκλογικά αποτελέσματα	Τυπικό σφάλμα	Εκλογικό αποτέλεσμα
ΝΔ	29,5	1	29,7	20,4	0,5	18,9
ΣΥΡΙΖΑ	27	1	26,9	16,2	0,5	16,8
ΠΑΣΟΚ	12,4	1	12,3	14,2	0,5	13,2
ΚΚΕ	4,5	0,5	4,5	8,6	0,2	8,5
ΑΝΕΛ	7,4	0,5	7,5	10,1	0,2	10,6
ΔΗΜΑΡ	6,2	0,5	6,3	5,8	0,2	6,1
ΧΑ	6,9	0,5	6,9	7	0,2	7
ΟΙΚΟΛΟΓΟΙ	0,9	0,5	0,9	2,8	0,2	2,9
ΛΑΟΣ	1,6	0,5	1,6	2,9	0,2	2,9
ΔΗΞΑΝ	1,5	0,5	1,5	-	-	-
ΛΟΙΠΑ	2,1		1,9	12		12,5
ΑΠΟΧΗ	38,7		37,5	38		34,9

ΚΕΦΑΛΑΙΟ 3

Εκλογές και προσωμοίωση

3.1 Εισαγωγή στην προσομοίωση

Ένας πειραματικός τρόπος μελέτης ενός στοχαστικού φαινόμενου είναι η παρακολούθηση πολλών πραγματοποιήσεών του, η καταγραφή των χαρακτηριστικών που μας ενδιαφέρουν και η εξαγωγή συμπερασμάτων. Είναι λογικό ότι όσες περισσότερες φορές αναπαριστούμε το φαινόμενο, τόσο καλύτερη και πιο αξιόπιστη θα είναι και η μελέτη μέσω αυτής της πειραματικής μεθόδου. Στην πράξη όμως κάτι τέτοιο κρίνεται, αν όχι ανέφικτο, ιδιαίτερα δαπανηρό και χρονοβόρο. Χαρακτηριστικό παράδειγμα αποτελεί η διαδικασία του exit poll για την πρόβλεψη των τελικών εκλογικών αποτελεσμάτων. Το exit poll δεν είναι δυνατόν να επαναληφθεί περισσότερες από μία φορές, αφενός λόγω του χρονικού ορίου που περιορίζει τον ερευνητή να συγκεντρώσει το δείγμα του στο διάστημα που οι κάλπες παραμένουν ανοιχτές (7π.μ – 7μ.μ) και αφετέρου λόγω του μεγάλου κόστους που θα συνεπάγονταν η συνεχής επανάληψη μίας τέτοιας διαδικασίας. Παρόλα αυτά, στις περισσότερες των περιπτώσεων, η επιστημονική ακρίβειά τους ως προς το τελικό αποτέλεσμα κρίνεται εντυπωσιακή.

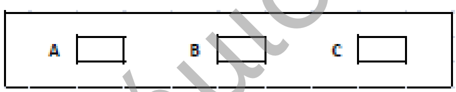
Στο κεφάλαιο αυτό, θα επιδιώξουμε να προβλέψουμε τα τελικά εκλογικά αποτελέσματα με μία τεχνική διαφορετική από εκείνη του exit poll και λιγότερο διαδεδομένη στον κύκλο των εταιρειών δημοσκοπήσεων, την τεχνική της *προσομοίωσης*. Μία εικονική πειραματική μέθοδο μέσω Η/Υ της οποίας η αποτελεσματικότητα βασίζεται στο νόμο των μεγάλων αριθμών. Με τη βοήθεια της προσομοίωσης ένας αναλυτής δεν έχει παρά να μοντελοποιήσει μία διαδικασία με συγκεκριμένες παραμέτρους και στη συνέχεια να την επαναλάβει με ψευδοτυχαίο τρόπο στην εικονική πραγματικότητα ενός Η/Υ χιλιάδες ή εκατομμύρια φορές σε ελάχιστο χρόνο και με ελάχιστο κόστος. Βάσει του νόμου των μεγάλων αριθμών, η οριακή σχετική συχνότητα εκδήλωσης του κάθε ενδεχομένου θα είναι ασυμπτωτικά ίση με την πραγματική πιθανότητα εκδήλωσής του.

Προκειμένου να γίνει κατανοητό το πώς λειτουργεί η τεχνική της προσομοίωσης και να τονιστεί η τεράστια χρησιμότητά της στην πράξη, θα κατασκευάσουμε στη συνέχεια έναν αλγόριθμο ο οποίος θα μπορεί να προβλέπει με όσο το δυνατόν μεγαλύτερη ακρίβεια τα τελικά αποτελέσματα μίας εκλογικής αναμέτρησης στην οποία εφαρμόστηκε η ταξινομική μέθοδος (Μέθοδος Condorcet) για τρεις υποψηφίους. Η διαδικασία γενικεύεται εύκολα και για περισσότερους από τρεις υποψηφίους.

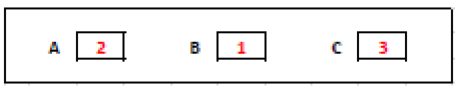
Το παράδειγμα εφαρμογής της μεθοδολογίας του Condorcet υλοποιήθηκε στο λογισμικό Wolfram Mathematica® και αποτελείται από δύο βασικά στάδια.

- Τη δημιουργία ενός εικονικού πληθυσμού 100000 ψηφοφόρων (αυτό αντιπροσωπεύει το συνολικό αριθμό των πολιτών που ψήφισαν στις εκλογές), για κάθε έναν από τους οποίους γνωρίζουμε τις προτιμήσεις του ως προς τους τρεις υποψηφίους A, B και C.
- Την εκτίμηση των αναμενόμενων ποσοστών επικράτησης του κάθε υποψηφίου στις ανά δύο αναμετρήσεις τους καθώς επίσης και την κατασκευή 95% διαστημάτων εμπιστοσύνης κατόπιν λήψης ενός τυχαίου δείγματος 4000 ψηφοφόρων από το συνολικό πληθυσμό.

Αρχικά, υποθέτουμε ότι κάθε ψηφοδέλτιο αποτελείται από τους υποψηφίους A, B και C δίπλα από τους οποίους υπάρχει ένα κενό κουτάκι:



Κατά την ψηφοφορία, ο κάθε ψηφοφόρος συμπληρώνει μέσα στα κουτάκια τους αριθμούς 1,2 και 3 ανάλογα με τη σειρά στην οποία επιθυμεί να κατατάξει τους τρεις υποψηφίους. Για παράδειγμα, αν ένας ψηφοφόρος θεωρεί ως καλύτερη επιλογή τον υποψήφιο B, ως δεύτερη επιλογή τον A και ως τελευταία επιλογή τον C τότε το ψηφοδέλτιο που θα ρίξει στην κάλπη θα είναι της μορφής:



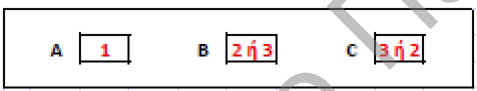
Ο συγκεκριμένος ψηφοφόρος (και το ψηφοδέλτιό του) περιγράφεται από το διάνυσμα {2,1,3}.

3.2 Περίπτωση τυχαίας ψηφοφορίας

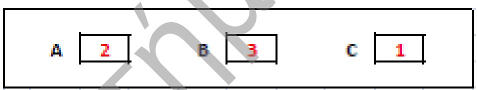
Υποθέτουμε ότι ο κάθε ψηφοφόρος επιλέγει τη σειρά των υποψηφίων με τυχαίο τρόπο. Για το λόγο αυτό κατασκευάζουμε 100000 διανύσματα με τους αριθμούς 1, 2 και 3 σε τυχαίες θέσεις. Για να το επιτύχουμε αυτό, σταθμίζουμε με τον ίδιο συντελεστή (1/3) το ενδεχόμενο εμφάνισης οποιουδήποτε αριθμού σε μία συγκεκριμένη θέση. Με άλλα λόγια, σε κάθε ψηφοδέλτιο και οι τρεις υποψήφιοι μοιράζονται μεταξύ τους την ίδια πιθανότητα να λάβουν τον αριθμό 1, 2 ή 3. Τα ψηφοδέλτια των 100000 ψηφοφόρων κατασκευάζονται και καταχωρούνται στον πίνακα *voter*:

```
population = 100 000;
voter = Table[RandomSample[{1 / 3, 1 / 3, 1 / 3} → {1, 2, 3}], {population}];
```

Στη συνέχεια, μπορούμε εύκολα να υπολογίσουμε τα πραγματικά ποσοστά επικράτησης του κάθε υποψήφιου στις ανά δύο αναμετρήσεις σύμφωνα με τις προτιμήσεις του πληθυσμού των 100000 ψηφοφόρων. Ειδικότερα, για να υπολογίσουμε το ποσοστό επικράτησης του υποψήφιου Α έναντι του Β θα πρέπει πρώτα να μετρήσουμε όλα εκείνα τα ψηφοδέλτια της μορφής:



και



Με άλλα λόγια, καταμετρούμε το πλήθος των διανυσμάτων της μορφής {1,_, _} και {2, 3, _} μέσα στον πίνακα *voter* και στη συνέχεια διαιρούμε με το συνολικό πλήθος των ψηφοφόρων *population* για να προκύψει το ποσοστό επικράτησης του A επί του B, ως μεταβλητή $A\,wins\,B$.

- Ποσοστό επικράτησης του A επί του B: $P(A=1)+P(A=2\,|\,B=3)$

Με τον ίδιο τρόπο υπολογίζουμε και τα ποσοστά επικράτησης του Α επί του C και του Β επί του C:

- Ποσοστό επικράτησης του A επί του C: $P(A=1)+P(A=2\,|\,C=3)$
- Ποσοστό επικράτησης του B επί του C: $P(B=1)+P(B=2\,|\,C=3)$

```
AwinsB = Count[voter, {1, _, _}] + Count[voter, {2, 3, _}];
AwinsC = Count[voter, {1, _, _}] + Count[voter, {2, _, 3}];
BwinsC = Count[voter, {_, 1, _}] + Count[voter, {_, 2, 3}];
```

```
results =
 MatrixForm[{{AwinsB / population // N}, {AwinsC / population // N},
   {BwinsC / population // N}},
  TableHeadings → {{"A wins B", "A wins C", "B wins C"}, None}]
```

A wins B	0.49944
A wins C	0.50205
B wins C	0.50167

Παρατηρούμε ότι (οριακά) η πλειοψηφία προτιμά τον υποψήφιο B έναντι του A, τον A έναντι του C και τον B έναντι του C. Επομένως, σύμφωνα με το κριτήριο του Condorcet νικητής των εκλογών ανακηρύσσεται ο υποψήφιος B από τη στιγμή που επικρατεί (έστω και οριακά) στις ανά δύο αναμετρήσεις με τον καθένα από τους υποψηφίους A και C.

Ένας τρόπος να αναπαραστήσουμε τη διαδικασία του exit poll που θα προβλέπει τα τελικά εκλογικά αποτελέσματα, είναι να επιλέξουμε από το συνολικό πληθυσμό του πίνακα *voter* ένα τυχαίο δείγμα n διανυσμάτων-ψηφοδελτίων (έστω n=4000) και να τα καταχωρήσουμε προσωρινά σε έναν νέο πίνακα *sample* για να τα επεξεργαστούμε. Ενεργώντας όπως προηγουμένως, υπολογίζουμε τα ποσοστά επικράτησης του κάθε υποψήφιου στις ανά δύο αναμετρήσεις αυτή τη φορά βάσει των 4000 ψηφοδελτίων που έχουμε στα χέρια μας και όχι βάσει ολόκληρου του πληθυσμού. Τα ποσοστά που υπολογίζουμε καταχωρούνται ως στοιχεία στους πίνακες AB, AC και BC αντίστοιχα. Αφού επανατοποθετήσουμε τα 4000 ψηφοδέλτια στον πληθυσμό, η παραπάνω διαδικασία επαναλαμβάνεται στην εικονική πραγματικότητα του Η/Υ χίλιες φορές (ενδεικτικά) έτσι ώστε να προκύψει ένας ικανός αριθμός πειραματικών αποτελεσμάτων. Για να έχει νόημα κάτι τέτοιο θα πρέπει τα εκτιμώμενα ποσοστά που θα προκύψουν από τη λήψη του δείγματος n να είναι όσο το δυνατόν πιο "κοντά" στα πραγματικά ποσοστά.

```
samp = 4000;
n = samp;
iterations = 1000;
AB = Table[0, {iterations}];
AC = Table[0, {iterations}];
BC = Table[0, {iterations}];
```

```
Do[
 sample = RandomSample[voter, n];
 AwinsB = Count[sample, {1, _, _}] + Count[sample, {2, 3, _}];
 AwinsC = Count[sample, {1, _, _}] + Count[sample, {2, _, 3}];
 BwinsC = Count[sample, {_, 1, _}] + Count[sample, {_, 2, 3}];
 AB[[j]] = AwinsB / n // N; AC[[j]] = AwinsC / n // N; BC[[j]] = BwinsC / n // N
 , {j, 1, iterations}]
```

Τα αναμενόμενα ποσοστά επικράτησης των ανά δύο αναμετρήσεων εκτιμώνται από τη μέση τιμή των δεδομένων που καταχωρήθηκαν σε κάθε πίνακα AB, AC και BC. Για να εκτιμηθούν τα όρια των 95% δ.ε θα πρέπει πρώτα να γίνει μία ταξινόμηση των δεδομένων των πινάκων σε αύξουσα σειρά προκειμένου να εντοπιστούν τα στοιχεία της 25ης (= 0.025 *iterations) και της 97.5ης (= 0.975 * iterations) θέσης[(5)].

```
AB = Sort[AB]; AC = Sort[AC]; BC = Sort[BC];
MatrixForm[
 {{AB[[Floor[0.025*iterations]]], Mean[AB], AB[[Ceiling[0.975*iterations]]],
   AB[[Ceiling[0.975*iterations]]] - AB[[Floor[0.025*iterations]]]},
  {AC[[Floor[0.025*iterations]]], Mean[AC], AC[[Ceiling[0.975*iterations]]],
   AC[[Ceiling[0.975*iterations]]] - AC[[Floor[0.025*iterations]]]},
  {BC[[Floor[0.025*iterations]]], Mean[BC], BC[[Ceiling[0.975*iterations]]],
   BC[[Ceiling[0.975*iterations]]] - BC[[Floor[0.025*iterations]]]}},
 TableHeadings →
  {{"95% Confidence Interval: A wins B", "95% Confidence Interval: A wins C", "95% Confidence Interval: B wins C"},
   {"Lower", "Mean", "Upper", "Range"}}]
Histogram[{AB, AC, BC}, Automatic, "Probability", ChartLabels → Placed[{"A wins B", "A wins C", "B wins C"}, Above]]
```

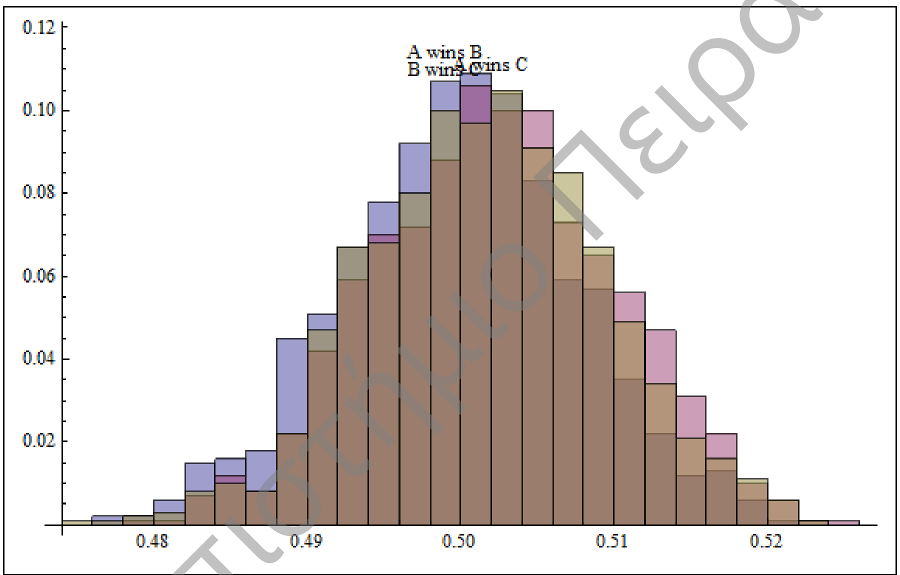
Τέλος, στον Πίνακα 3.1 παρουσιάζονται τα αναμενόμενα ποσοστά επικράτησης όλων των ανά δύο αναμετρήσεων καθώς και τα όρια του 95% δ.ε ενώ, στο Διάγραμμα 3.1 παρουσιάζονται τα ιστογράμματα των πινάκων AB, AC και BC.

[(5)]Ενδέχεται τα άνω και κάτω ποσοστημόρια του επιθυμητού διαστήματος εμπιστοσύνης πολλαπλασιαζόμενα με τον αριθμό των επαναλήψεων του πειράματος να μην οδηγήσουν σε ακέραιους αριθμούς. Για παράδειγμα, αν κάποιος αναλυτής θέσει *iterations*=1035 και διάστημα εμπιστοσύνης 95%, τότε τα αποτελέσματα 25,875 και 1009,13 δεν αντιστοιχούν σε θέσεις στοιχείων εντός των πινάκων. Στην περίπτωση αυτή, η υπολογιστική διαδικασία θα βρεθεί σε αδιέξοδο. Για την εξάλειψη αυτού του ενδεχόμενου χρησιμοποιήθηκαν οι εντολές *Floor* και *Ceiling* έτσι ώστε να προκύπτει πάντα ακέραιος αριθμός (από την πλευρά ασφαλείας του δ.ε.), και ο οποίος αντιστοιχεί σε συγκεκριμένη θέση εντός των πινάκων. Έτσι, στην περίπτωση των 1035 επαναλήψεων, το διάστημα εμπιστοσύνης 95% θα οριοθετηθεί από τα στοιχεία της *Floor*(25,875) = 25ης και της *Ceiling*(1009,13) = 1010ης θέσης.

Πίνακας 3.1: Αποτελέσματα των αναμενόμενων ποσοστών επικράτησης όλων των ανά δύο αναμετρήσεων

Exit Poll	Κάτω Φράγμα	Άνω Φράγμα	Εκτίμηση (Μέση Τιμή)	Επίσημα Αποτελέσματα
A-B	**0.48375**	**0.515**	0.499705	**0.49944**
A-C	**0.4865**	**0.51725**	0.502162	**0.50205**
B-C	**0.48575**	**0.51675**	0.501486	**0.50167**

Διάγραμμα 3.1
Ιστόγραμμα των κατανομών των εκτιμώμενων ποσοστών



Από τα παραπάνω προκύπτει ότι οι μέσες τιμές των ποσοστών επικράτησης όπως αυτές προέκυψαν από την επιλογή ενός δείγματος 4000 ψηφοφόρων, κατορθώνουν να κατατάξουν τους υποψήφιους σύμφωνα με τα τελικά αποτελέσματα των εκλογών. Οποιαδήποτε όμως πρόβλεψη για την τελική κατάταξη των υποψηφίων θα ήταν παρακινδυνευμένη από τη στιγμή που τα διαστήματα εμπιστοσύνης σχεδόν αλληλεπικαλύπτονται. Για παράδειγμα, το ελάχιστο «προβάδισμα» κατά 0,06% του B έναντι του A δεν προεξοφλεί σε καμιά περίπτωση τη νίκη του B, γεγονός που αφήνει ανοιχτά όλα τα ενδεχόμενα. Στο ίδιο συμπέρασμα καταλήγουμε και αν κοιτάξουμε το ιστόγραμμα του διαγράμματος 3.1 το οποίο αναπαριστά τις κατανομές των εκτιμώμενων ποσοστών. Παρατηρούμε ότι η μία κατανομή πέφτει σχεδόν πάνω στην άλλη και επομένως δε μπορούμε να προβλέψουμε με ακρίβεια ποιος υποψήφιος κερδίζει ποιόν σε κάθε “μονομαχία” και κατά συνέπεια ποιος θα είναι ο τελικός νικητής. Τέλος, είναι σκόπιμο να αναφέρουμε ότι αν κάποιος αναλυτής επιλέξει να λάβει ένα μεγαλύτερο δείγμα ψηφοφόρων, το εύρος του 95% διαστήματος εμπιστοσύνης των εκτιμήσεων θα μειωθεί. Επιπλέον, όσο αυξάνεται ο αριθμός επαναλήψεων του πειράματος τόσο τα ιστογράμματα θα δίνουν μία «ομαλότερη» εικόνα κατανομών.

3.3 Έλεγχος κανονικότητας των αποτελεσμάτων

Βάσει του κριτηρίου Kolmogorov-Smirnov, η μηδενική υπόθεση ότι τα στοιχεία των πινάκων AB, AC και BC ακολουθούν την κανονική κατανομή δεν απορρίπτεται, καθώς το *p-value* του ελέγχου προκύπτει πολύ υψηλό (*p-value*=0.5). Ο έλεγχος K-S για τα στοιχεία του πίνακα AB υλοποιείται ως εξής:

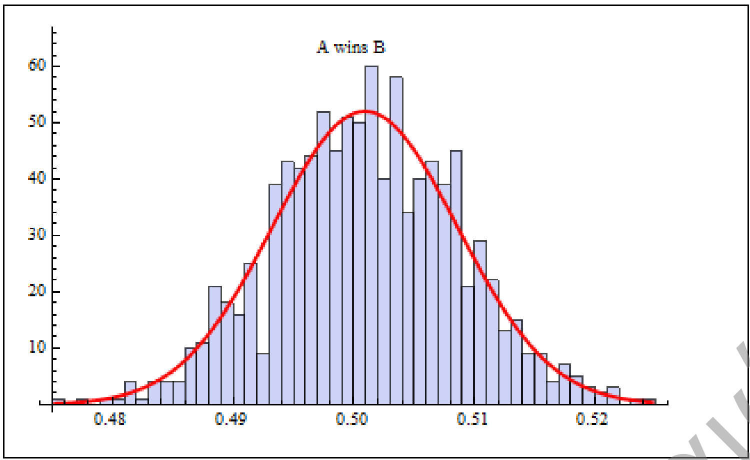
```
m = iterations;
μ = Mean[AB]; σ = StandardDeviation[AB]; norm0 = NormalDistribution[μ, σ];
d1 = Max[Table[i / m - CDF[norm0, AB[[i]]], {i, 1, m}]];
d2 = Max[Table[CDF[norm0, AB[[i]]] - (i - 1) / m, {i, 1, m}]];
d = Max[d1, d2];

k1 = 100; s = 0;
Do[
  X = Table[RandomReal[norm0], {m}]; X = Sort[X];
  μ1 = Mean[X]; σ1 = StandardDeviation[X];
  norm1 = NormalDistribution[μ1, σ1];
  d1 = Max[Table[i / m - CDF[norm1, X[[i]]], {i, 1, m}]];
  d2 = Max[Table[CDF[norm1, X[[i]]] - (i - 1) / m, {i, 1, m}]];
  If[Max[d1, d2] ≥ d, s = s + 1], {j, 1, k1}];
pvalue = s / k1 // N
```

Επιπλέον, η κανονική κατανομή με παραμέτρους $\hat{\mu}$ και $\hat{\sigma}^2$ οι οποίες προκύπτουν από τη μέση τιμή και τη διακύμανση του πίνακα AB, φαίνεται να προσαρμόζεται πολύ ικανοποιητικά στο ιστόγραμμα των στοιχείων του πίνακα (Διάγραμμα 3.2):

```
R1 := Plot[PDF[NormalDistribution[μ, σ], x], {x, AB[[1]], AB[[m]]}, PlotStyle → {Red, Thick}]
R2 := Histogram[AB, 50, ChartLabels → Placed[{"A wins B"}, Above]]
Show[{R2, R1}]
```

Διάγραμμα 3.2
Ιστόγραμμα που αναπαριστά την κατανομή των στοιχείων του πίνακα AB



3.4 Περίπτωση μεροληπτικής ψηφοφορίας

Προκειμένου να γίνει πιο ξεκάθαρο το πώς ο συγκεκριμένος αλγόριθμος θα μπορούσε να προβλέψει τα τελικά ποσοστά των εκλογών, θα εργαστούμε με τον ίδιο ακριβώς τρόπο μόνο που αυτή τη φορά θα δημιουργήσουμε έναν εικονικό πληθυσμό με τις προτιμήσεις που εμείς θέλουμε. Με άλλα λόγια θα θεωρήσουμε ότι καθένας από τους 100000 ψηφοφόρους δεν ψήφισε με τυχαίο τρόπο, αλλά ότι επέλεξε τον υποψήφιο A με πιθανότητα 0.5 για να πάρει τη σειρά {1,2,3}, τον υποψήφιο B με πιθανότητα 0.3 και τον υποψήφιο C με πιθανότητα 0.2.

```
population = 100 000;
voter = Table[RandomSample[{0.5, 0.3, 0.2} → {1, 2, 3}], {population}];
```

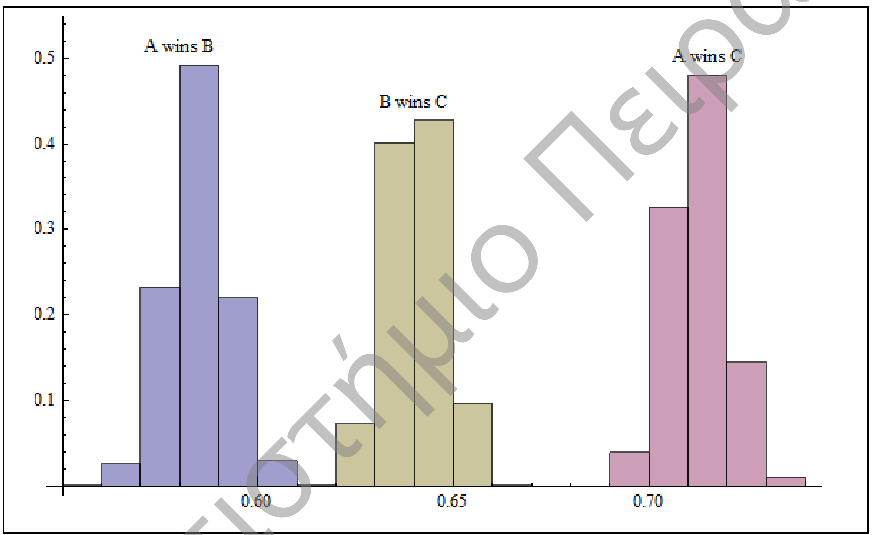
A wins B	0.58454
A wins C	0.71237
B wins C	0.64039

Παρατηρούμε ότι η πλειοψηφία προτιμά τον υποψήφιο A έναντι του B, τον A έναντι του C και τον B έναντι του C. Επομένως, σύμφωνα με το κριτήριο του Condorcet νικητής των εκλογών ανακηρύσσεται ο υποψήφιος A. Τέλος, στον Πίνακα 3.2 παρουσιάζονται τα αναμενόμενα ποσοστά επικράτησης των ανά δύο αναμετρήσεων καθώς και τα όρια του 95% δ.ε ενώ, στο Διάγραμμα 3.3 παρουσιάζονται τα ιστογράμματα των πινάκων AB, AC και BC.

Πίνακας 3.2: Αποτελέσματα των αναμενόμενων ποσοστών επικράτησης όλων των ανά δύο αναμετρήσεων

Exit Poll	95% Διάστημα Εμπιστοσύνης	Εκτίμηση (Μέση Τιμή)	Επίσημα Αποτελέσματα	Απόκλιση
A-B	**[56.98% , 60.03]**	58.50%	**58.45%**	**0.05%**
A-C	**[69.78% , 72.70%]**	71.24%	**71.24%**	**0%**
B-C	**[62.63% , 65.50%]**	64.06%	**64.04%**	**0.2%**

Διάγραμμα 3.3
Ιστόγραμμα των κατανομών των εκτιμώμενων ποσοστών



Από το ιστόγραμμα του διαγράμματος 3.3 παρατηρούμε ότι είναι ξεκάθαρο πλέον ποιος υποψήφιος κερδίζει ποιόν και κατά συνέπεια ποιος θα είναι ο τελικός νικητής. Στο ίδιο συμπέρασμα καταλήγουμε και αν κοιτάξουμε τα 95% διαστήματα εμπιστοσύνης του πίνακα 3.2.

3.5 Εκτίμηση του απαιτούμενου μεγέθους δείγματος

Πριν από κάθε exit poll οι εταιρίες δημοσκοπήσεων υπολογίζουν ποιο θα πρέπει να είναι το μικρότερο δυνατό μέγεθος του δείγματος n που θα τους επιτρέψει να προβλέψουν με όσο το δυνατόν μεγαλύτερη ακρίβεια τα τελικά εκλογικά αποτελέσματα, δεδομένου ότι όσο αυξάνεται το μέγεθος του δείγματος τόσο αυξάνεται το κόστος αλλά και ο χρόνος που απαιτείται για την υλοποίηση της διαδικασίας. Με μία μικρή τροποποίηση της υπολογιστικής διαδικασίας που περιγράψαμε στις προηγούμενες παραγράφους, θα εκτιμήσουμε το απαιτούμενο μέγεθος δείγματος βάσει ενός συγκεκριμένου εύρους εκτίμησης και για ένα συγκεκριμένο περιθώριο σφάλματος που εμείς εκ των προτέρων έχουμε ορίσει.

Έστω ότι ένας αναλυτής προτίθεται να πραγματοποιήσει ένα exit poll σε έναν πληθυσμό 100000 ατόμων θέλοντας να προσδιορίσει εκείνο το μέγεθος δείγματος ψηφοφόρων που θα του εξασφαλίσει ένα εύρος το πολύ ±1.5 ποσοστιαίας μονάδας γύρω από τα αναμενόμενα ποσοστά επικράτησης όλων των ανά δύο αναμετρήσεων, διατηρώντας ταυτόχρονα ένα επίπεδο εμπιστοσύνης 90%. Σε πρώτο στάδιο, θα κατασκευάσουμε έναν πληθυσμό ψηφοφόρων, οι οποίοι έχουν ως πρώτη επιλογή τον υποψήφιο A σε ποσοστό 50%, ως δεύτερη επιλογή τον υποψήφιο B σε ποσοστό 30% και ως τρίτη επιλογή τον υποψήφιο C σε ποσοστό 20%:

```
population = 100 000;
voter = Table[RandomSample[{0.5, 0.3, 0.2} → {1, 2, 3}], {population}];
AwinsB = Count[voter, {1, _, _}] + Count[voter, {2, 3, _}];
AwinsC = Count[voter, {1, _, _}] + Count[voter, {2, _, 3}];
BwinsC = Count[voter, {_, 1, _}] + Count[voter, {_, 2, 3}];
results = MatrixForm[{{AwinsB / population // N}, {AwinsC / population // N}, {BwinsC / population // N}},
  TableHeadings → {{"A wins B", "A wins C", "B wins C"}, None}]
```

Στη συνέχεια, θα θεωρήσουμε ότι οι εκτιμήσεις των ποσοστών επικράτησης βάσει του δείγματος ακολουθούν την κανονική κατανομή με παραμέτρους τη μέση τιμή και τη διακύμανση των πινάκων AB, AC και BC. Στην προκειμένη περίπτωση η εκτίμηση των διαστημάτων εμπιστοσύνης γίνεται παραμετρικά, σε αντίθεση με το αρχικό παράδειγμα όπου εκτιμήθηκαν μη παραμετρικά διαστήματα εμπιστοσύνης (δεν γνωρίζαμε ακόμη ότι επρόκειτο για κανονική κατανομή). Επίσης, θα ξεκινήσουμε με ένα αρχικό μέγεθος δείγματος 4000 ψηφοφόρων, το οποίο θα αυξάνουμε σταδιακά κατά 100, έως ότου επιτύχουμε το επιθυμητό εύρος των τριών ποσοστιαίων μονάδων για επίπεδο εμπιστοσύνης 90%:

```
signif = 0.1;
limitrange = 0.03;
iterations = 1000;
samp = 4000;
step = 100;
AB = Table[0, {iterations}]; AC = Table[0, {iterations}]; BC = Table[0, {iterations}];
repeat = True;
While[repeat, samp = samp + step; n = samp;
  Do[
   sample = RandomSample[voter, n];
   AwinsB = Count[sample, {1, _, _}] + Count[sample, {2, 3, _}];
   AwinsC = Count[sample, {1, _, _}] + Count[sample, {2, _, 3}];
   BwinsC = Count[sample, {_, 1, _}] + Count[sample, {_, 2, 3}];
   AB[[j]] = AwinsB / n // N; AC[[j]] = AwinsC / n // N; BC[[j]] = BwinsC / n // N, {j, 1, iterations}];
  μAB = Mean[AB]; σAB = StandardDeviation[AB];
  μAC = Mean[AC]; σAC = StandardDeviation[AC];
  μBC = Mean[BC]; σBC = StandardDeviation[BC];
  lowerAB = InverseCDF[NormalDistribution[μAB, σAB], signif / 2];
  upperAB = InverseCDF[NormalDistribution[μAB, σAB], 1 - signif / 2];
  lowerAC = InverseCDF[NormalDistribution[μAC, σAC], signif / 2];
  upperAC = InverseCDF[NormalDistribution[μAC, σAC], 1 - signif / 2];
  lowerBC = InverseCDF[NormalDistribution[μBC, σBC], signif / 2];
  upperBC = InverseCDF[NormalDistribution[μBC, σBC], 1 - signif / 2];
  rangeAB = upperAB - lowerAB; rangeAC = upperAC - lowerAC; rangeBC = upperBC - lowerBC;
  If[Max[rangeAB, rangeAC, rangeBC] < limitrange, repeat = False]];
MatrixForm[{{lowerAB, μAB, upperAB, rangeAB, 1 - signif}, {lowerAC, μAC, upperAC, rangeAC, 1 - signif},
  {lowerBC, μBC, upperBC, rangeBC, 1 - signif}},
 TableHeadings → {{"A wins B", "A wins C", "B wins C"}, {"Lower", "Mean", "Upper", "Range", "Conf. level"}}]
Print["The required sample to achieve all confidence intervals of the pairwise encounters less than ", limitrange,
  " at the ", 1 - signif, " confidence level in a population of N=", population, " is: n=", samp];
```

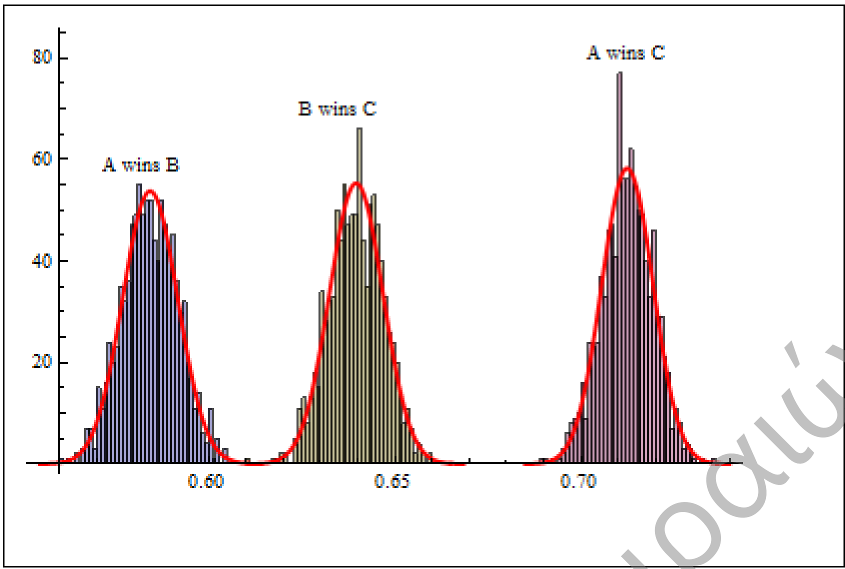
	Lower	Mean	Upper	Range	Conf. level
A wins B	0.572103	0.584314	0.596526	0.0244231	0.9
A wins C	0.701166	0.712434	0.723702	0.0225359	0.9
B wins C	0.62771	0.639575	0.65144	0.0237298	0.9

The required sample to achieve all confidence intervals of the pairwise encounters less than 0.03 at the 0.9 confidence level in a population of N=100 000 is: n=4100

Ακολουθεί η γραφική απεικόνιση των παραπάνω:

```
R11 := Plot[PDF[NormalDistribution[μAB, σAB], x], {x, μAB - 4 * σAB, μAB + 4 * σAB}, PlotStyle → {Thick, Red}]
R22 := Plot[PDF[NormalDistribution[μAC, σAC], x], {x, μAC - 4 * σAC, μAC + 4 * σAC}, PlotStyle → {Thick, Red}]
R33 := Plot[PDF[NormalDistribution[μBC, σBC], x], {x, μBC - 4 * σBC, μBC + 4 * σBC}, PlotStyle → {Thick, Red}]
R44 := Histogram[{AB, AC, BC}, 200, ChartLabels → Placed[{"A wins B", "A wins C", "B wins C"}, Above]]
Show[{R44, R11, R22, R33}, PlotRange → All]
```

Διάγραμμα 3.4
Ιστόγραμμα για δεδομένο εύρος 3% και επίπεδο εμπιστοσύνης 90%



Παρατηρούμε ότι για την κατασκευή 90% διαστημάτων εμπιστοσύνης και για δεδομένο εύρος της τάξεως του 3% , το μέγεθος δείγματος που απαιτείται για την πρόβλεψη είναι $n = 4100$ ψηφοφόροι ενώ η ερμηνεία των αποτελεσμάτων είναι όμοια με εκείνη των προηγούμενων παραγραφών. Όμοια, μπορούμε να κατασκευάσουμε μικρότερα ή μεγαλύτερα διαστήματα εμπιστοσύνης με διαφορετικό εύρος στα ποσοστά των προβλέψεων. Ενδεικτικά αναφέρουμε ότι:

- το απαιτούμενο δείγμα για 95% διαστήματα εμπιστοσύνης και εύρος 3% είναι $n = 4200$ ψηφοφόροι
- το απαιτούμενο δείγμα για 95% διαστήματα εμπιστοσύνης και εύρος 2% είναι $n = 8700$ ψηφοφόροι
- το απαιτούμενο δείγμα για 99% διαστήματα εμπιστοσύνης και εύρος 3% είναι $n = 6800$ ψηφοφόροι

ΚΕΦΑΛΑΙΟ 4

Exit poll και τελικά εκλογικά αποτελέσματα

4.1 Εισαγωγή

Στο τελευταίο κεφάλαιο, με τη βοήθεια των πινάκων συνάφειας, θα επιδιώξουμε σε πρώτη φάση να εξετάσουμε κατά πόσο ορισμένες μεταβλητές (π.χ φύλο, ηλικία, επάγγελμα) είναι ικανές να επηρεάσουν στατιστικά σημαντικά την ψήφο του i-ψηφοφόρου. Στη συνέχεια, θα επιδιώξουμε να κατασκευάσουμε ένα στατιστικό μοντέλο λογιστικής παλινδρόμησης που θα προβλέπει τα τελικά εκλογικά αποτελέσματα με βάση τα δεδομένα του exit poll. Επειδή λόγω απορρήτου δε μας δόθηκε η δυνατότητα να έχουμε στη διάθεσή μας τη βάση δεδομένων με τα στοιχεία ενός πραγματικού exit poll, αυτό που θα κάνουμε είναι να ακολουθήσουμε την αντίστροφη πορεία. Συγκεκριμένα, θα χρησιμοποιήσουμε τα αποτελέσματα του exit poll των ευρωεκλογών του 2014 και με τη βοήθεια της προσομοίωσης θα κατασκευάσουμε μία εικονική βάση δεδομένων για ένα δείγμα ψηφοφόρων. Εν συνεχεία, θα μεταφέρουμε τα δεδομένα στο στατιστικό πακέτο IBM SPSS Statistics 22.0 και θα προβούμε στην ανάλυση των αποτελεσμάτων.

4.2 Κατασκευή των δεδομένων

Από τη στιγμή που δεν έχουμε στη διάθεσή μας τη βάση δεδομένων με τα στοιχεία του πραγματικού exit poll, θα χρησιμοποιήσουμε τα έτοιμα αποτελέσματα του exit poll των ευρωεκλογών του 2014 και με τη βοήθεια της προσομοίωσης θα κατασκευάσουμε έναν αλγόριθμο ο οποίος θα παράγει με βάση αυτά τα στοιχεία μία εικονική βάση δεδομένων για ένα δείγμα 8000 ψηφοφόρων, όσο ακριβώς δηλαδή ήταν και το μέγεθος του δείγματος του κοινού exit poll πέντε εταιριών δημοσκοπήσεων που πραγματοποιήθηκε στις 25/5/2014. Ένας τέτοιος αλγόριθμος θα επιλέγει τυχαία ένα δείγμα 8000 ψηφοφόρων και σε κάθε έναν από αυτούς θα αντιστοιχεί πέντε χαρακτηριστικά. Το κόμμα που ψήφισε σήμερα, την ηλικία του, το φύλο του, τη θέση του στην απασχόληση καθώς και το κόμμα που ψήφισε στις βουλευτικές εκλογές του 2012 χωρίς όμως να λαμβάνει υπόψη τις συσχετίσεις μεταξύ των πέντε χαρακτηριστικών.

Οι ανεξάρτητες μεταβλητές που χρησιμοποιήσαμε και που μπορεί να επηρεάσουν την ψήφο του i-ψηφοφόρου είναι οι εξής:

- **Age**, η οποία θα παίρνει την τιμή 0 εάν ο i-ψηφοφόρος ήταν από 18 έως 24, την τιμή 1 εάν ήταν από 25 έως 34, την τιμή 2 εάν ήταν από 35 έως 54 και την τιμή 3 εάν ήταν από 55 και πάνω.
- **Gender**, η οποία θα παίρνει την τιμή 0 εάν ο i-ψηφοφόρος ήταν άντρας και την τιμή 1 εάν ήταν γυναίκα.
- **Occupation**, η οποία θα παίρνει την τιμή 0 εάν ο i-ψηφοφόρος ήταν εργαζόμενος (ιδιωτικού ή δημοσίου τομέα), την τιμή 1 εάν ήταν φοιτητής, την τιμή 2 εάν ήταν άνεργος και την τιμή 3 εάν ήταν συνταξιούχος.
- **Previous Vote**,η οποία θα παίρνει την τιμή 0 εάν στις βουλευτικές εκλογές του 2012 ο i-ψηφοφόρος ψήφισε ΝΔ, την τιμή 1 εάν ψήφισε ΣΥ.ΡΙΖ.Α, την τιμή 2 εάν ψήφισε ΠΑΣΟΚ, την τιμή 3 εάν ψήφισε ΚΚΕ, την τιμή 4 εάν ψήφισε Χ.Α και την τιμή 5 εάν ψήφισε κάποιο άλλο κόμμα.

Οι περιορισμοί που χρησιμοποιήθηκαν είναι οι εξής:

- Ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 0 ή 1 δε μπορεί να είναι συνταξιούχος,
- Ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 2 ή 3 δε μπορεί να είναι φοιτητής,
- Ένας ψηφοφόρος που δηλώνει φοιτητής δε μπορεί να ανήκει στην ηλικιακή ομάδα 2 ή 3,
- Ένας ψηφοφόρος που είναι συνταξιούχος δε μπορεί ανήκει στην ηλικιακή ομάδα 0 ή 1.

Με βάση τα παραπάνω, ο αλγόριθμος που προέκυψε είναι, σε κώδικα του Mathematica ο εξής:

```
SeedRandom[2];
exitpoll = 8000;

"κόμματα ψηφοφόρων του exitpoll";
parties = Table[RandomSample[{24, 27, 8.5, 6, 9.5, 25} → {0, 1, 2, 3, 4, 5}][[1]], {exitpoll}];

"ΝΕΑ ΔΗΜΟΚΡΑΤΙΑ";
NDpartie = Count[parties, 0];

"κατασκευή φύλου και συσπείρωσης ψηφοφόρων";
NDgender = Table[RandomSample[{23.1, 25.1} → {0, 1}][[1]], {NDpartie}];
NDprevious = Table[RandomSample[{66, 6, 2, 1, 7, 18} → {0, 1, 2, 3, 4, 5}][[1]], {NDpartie}];
NDoccup = Table[RandomSample[{18.6, 16.5, 17.5, 35} → {0, 1, 2, 3}][[1]], {NDpartie}];

"κατασκευή ψηφοφόρων ΣΟΡΤΑΡΙΣΜΕΝΩΝ ανά ηλικιακή ομάδα";
NDage = Sort[Table[RandomSample[{16.1, 17.5, 20.5, 32.3} → {0, 1, 2, 3}][[1]], {NDpartie}]];

"πλήθος ανά ηλικιακή ομάδα";
age0 = Count[NDage, 0];
age1 = Count[NDage, 1];
age2 = Count[NDage, 2];
age3 = Count[NDage, 3];

"πλήθος ηλικιακών ομάδων 0+1 και 2+3";
age01 = age0 + age1;
age23 = age2 + age3;

"πλήθος ανά επάγγελμα";
occup0 = Count[NDoccup, 0];
occup1 = Count[NDoccup, 1];
occup2 = Count[NDoccup, 2];
occup3 = Count[NDoccup, 3];

"κατασκευή επαγγελμάτων για τις ηλικιακές ομάδες 1+2 (δεν έχει συνταξιούχους)";
occupage01 = Table[RandomSample[{occup0, occup1, occup2} → {0, 1, 2}][[1]], {age01}];

"κατασκευή επαγγελμάτων για τις ηλικιακές ομάδες 2+3 (δεν έχει φοιτητές)";
occupage23 = Table[RandomSample[{occup0, occup2, occup3} → {0, 2, 3}][[1]], {age23}];

"τελική κατασκευή όλων των επαγγελμάτων για τους ΣΟΡΤΑΡΙΣΜΕΝΟΥΣ ανά ηλικιακή ομάδα ψηφοφόρους";
NDoccup = Join[occupage01, occupage23];

"τελικό προφιλ ψηφοφόρων της ΝΔ, ΣΟΡΤΑΡΙΣΜΕΝΩΝ ανά ηλικιακή ομάδα";
NDvoter = Table[{0, NDage[[i]], NDoccup[[i]], NDgender[[i]], NDprevious[[i]]}, {i, 1, NDpartie}];
```

```
"ΣΥ.ΡΙΖ.Α";
SYRpartie = Count[parties, 1];

"κατασκευή φύλου και συσπείρωσης ψηφοφόρων";
SYRgender = Table[RandomSample[{26, 28.2} → {0, 1}][[1]], {SYRpartie}];
SYRprevious = Table[RandomSample[{2, 76, 2, 4, 2, 14} → {0, 1, 2, 3, 4, 5}][[1]], {SYRpartie}];
SYRoccup = Table[RandomSample[{29.6, 29.3, 30, 23.3} → {0, 1, 2, 3}][[1]], {SYRpartie}];

"κατασκευή ψηφοφόρων ΣΟΡΤΑΡΙΣΜΕΝΩΝ ανά ηλικιακή ομάδα";
SYRage = Sort[Table[RandomSample[{25.5, 26.5, 29.1, 25.3} → {0, 1, 2, 3}][[1]], {SYRpartie}]];

"πλήθος ανά ηλικιακή ομάδα";
age0 = Count[SYRage, 0];
age1 = Count[SYRage, 1];
age2 = Count[SYRage, 2];
age3 = Count[SYRage, 3];

"πλήθος ηλικιακών ομάδων 0+1 και 2+3";
age01 = age0 + age1;
age23 = age2 + age3;

"πλήθος ανά επάγγελμα";
occup0 = Count[SYRoccup, 0];
occup1 = Count[SYRoccup, 1];
occup2 = Count[SYRoccup, 2];
occup3 = Count[SYRoccup, 3];
"κατασκευή επαγγελμάτων για τις ηλικιακές ομάδες 1+2 (δεν έχει συνταξιούχους)";
occupage01 = Table[RandomSample[{occup0, occup1, occup2} → {0, 1, 2}][[1]], {age01}];

"κατασκευή επαγγελμάτων για τις ηλικιακές ομάδες 2+3 (δεν έχει φοιτητές)";
occupage23 = Table[RandomSample[{occup0, occup2, occup3} → {0, 2, 3}][[1]], {age23}];

"τελική κατασκευή όλων των επαγγελμάτων για τους ΣΟΡΤΑΡΙΣΜΕΝΟΥΣ ανά ηλικιακή ομάδα ψηφοφόρους";
SYRoccup = Join[occupage01, occupage23];

"τελικό προφιλ ψηφοφόρων του ΣΥ.ΡΙΖ.Α, ΣΟΡΤΑΡΙΣΜΕΝΩΝ ανά ηλικιακή ομάδα";
SYRvoter = Table[{1, SYRage[[i]], SYRoccup[[i]], SYRgender[[i]], SYRprevious[[i]]}, {i, 1, SYRpartie}];

"ΕΛΙΑ";
ELIApartie = Count[parties, 2];

"κατασκευή φύλου και συσπείρωσης ψηφοφόρων";
ELIAgender = Table[RandomSample[{8.9, 8} → {0, 1}][[1]], {ELIApartie}];
ELIAprevious = Table[RandomSample[{6, 24, 43, 2, 3, 22} → {0, 1, 2, 3, 4, 5}][[1]], {ELIApartie}];
ELIAoccup = Table[RandomSample[{6.5, 3.6, 5.9, 14} → {0, 1, 2, 3}][[1]], {ELIApartie}];

"κατασκευή ψηφοφόρων ΣΟΡΤΑΡΙΣΜΕΝΩΝ ανά ηλικιακή ομάδα";
ELIAage = Sort[Table[RandomSample[{5.7, 4.7, 6.7, 12.6} → {0, 1, 2, 3}][[1]], {ELIApartie}]];
```

```
"πλήθος ανά ηλικιακή ομάδα";
age0 = Count[ELIAage, 0];
age1 = Count[ELIAage, 1];
age2 = Count[ELIAage, 2];
age3 = Count[ELIAage, 3];

"πλήθος ηλικιακών ομάδων 0+1 και 2+3";
age01 = age0 + age1;
age23 = age2 + age3;

"πλήθος ανά επάγγελμα";
occup0 = Count[ELIAoccup, 0];
occup1 = Count[ELIAoccup, 1];
occup2 = Count[ELIAoccup, 2];
occup3 = Count[ELIAoccup, 3];

"κατασκευή επαγγελμάτων για τις ηλικιακές ομάδες 1+2 (δεν έχει συνταξιούχους)";
occupage01 = Table[RandomSample[{occup0, occup1, occup2} → {0, 1, 2}][[1]], {age01}];

"κατασκευή επαγγελμάτων για τις ηλικιακές ομάδες 2+3 (δεν έχει φοιτητές)";
occupage23 = Table[RandomSample[{occup0, occup2, occup3} → {0, 2, 3}][[1]], {age23}];

"τελική κατασκευή όλων των επαγγελμάτων για τους ΣΟΡΤΑΡΙΣΜΕΝΟΥΣ ανά ηλικιακή ομάδα ψηφοφόρους";
ELIAoccup = Join[occupage01, occupage23];

"τελικό προφιλ ψηφοφόρων της ΕΛΙΑΣ, ΣΟΡΤΑΡΙΣΜΕΝΩΝ ανά ηλικιακή ομάδα";
ELIAvoter = Table[{2, ELIAage[[i]], ELIAoccup[[i]], ELIAgender[[i]], ELIAprevious[[i]]}, {i, 1, ELIApartie}];

"KKE";
KKEpartie = Count[parties, 3];

"κατασκευή φύλου και συσπείρωσης ψηφοφόρων";
KKEgender = Table[RandomSample[{6.3, 5.6} → {0, 1}][[1]], {KKEpartie}];
KKEprevious = Table[RandomSample[{1, 14, 0, 73, 3, 9} → {0, 1, 2, 3, 4, 5}][[1]], {KKEpartie}];
KKEoccup = Table[RandomSample[{5.1, 7.9, 6, 6.1} → {0, 1, 2, 3}][[1]], {KKEpartie}];

"κατασκευή ψηφοφόρων ΣΟΡΤΑΡΙΣΜΕΝΩΝ ανά ηλικιακή ομάδα";
KKEage = Sort[Table[RandomSample[{6.5, 5.8, 5.2, 6.8} → {0, 1, 2, 3}][[1]], {KKEpartie}]];

"πλήθος ανά ηλικιακή ομάδα";
age0 = Count[KKEage, 0];
age1 = Count[KKEage, 1];
age2 = Count[KKEage, 2];
age3 = Count[KKEage, 3];

"πλήθος ηλικιακών ομάδων 0+1 και 2+3";
age01 = age0 + age1;
age23 = age2 + age3;

"πλήθος ανά επάγγελμα";
occup0 = Count[KKEoccup, 0];
occup1 = Count[KKEoccup, 1];
occup2 = Count[KKEoccup, 2];
occup3 = Count[KKEoccup, 3];
```

```
"κατασκευή επαγγελμάτων για τις ηλικιακές ομάδες 1+2 (δεν έχει συνταξιούχους)";
occupage01 = Table[RandomSample[{occup0, occup1, occup2} → {0, 1, 2}][[1]], {age01}];

"κατασκευή επαγγελμάτων για τις ηλικιακές ομάδες 2+3 (δεν έχει φοιτητές)";
occupage23 = Table[RandomSample[{occup0, occup2, occup3} → {0, 2, 3}][[1]], {age23}];

"τελική κατασκευή όλων των επαγγελμάτων για τους ΣΟΡΤΑΡΙΣΜΕΝΟΥΣ ανά ηλικιακή ομάδα ψηφοφόρους";
KKEoccup = Join[occupage01, occupage23];

"τελικό προφιλ ψηφοφόρων του ΚΚΕ, ΣΟΡΤΑΡΙΣΜΕΝΩΝ ανά ηλικιακή ομάδα";
KKEvoter = Table[{3, KKEage[[i]], KKEoccup[[i]], KKEgender[[i]], KKEprevious[[i]]}, {i, 1, KKEpartie}];

"ΧΑ";
XApartie = Count[parties, 4];

"κατασκευή φύλου και συσπείρωσης ψηφοφόρων";
XAgender = Table[RandomSample[{12.5, 5.7} → {0, 1}][[1]], {XApartie}];
XAprevious = Table[RandomSample[{6, 4, 2, 1, 73, 14} → {0, 1, 2, 3, 4, 5}][[1]], {XApartie}];
XAoccup = Table[RandomSample[{9.8, 8.2, 12.2, 6.6} → {0, 1, 2, 3}][[1]], {XApartie}];

"κατασκευή ψηφοφόρων ΣΟΡΤΑΡΙΣΜΕΝΩΝ ανά ηλικιακή ομάδα";
XAage = Sort[Table[RandomSample[{11.2, 11, 11.1, 6.8} → {0, 1, 2, 3}][[1]], {XApartie}]];

"πλήθος ανά ηλικιακή ομάδα";
age0 = Count[XAage, 0];
age1 = Count[XAage, 1];
age2 = Count[XAage, 2];
age3 = Count[XAage, 3];

"πλήθος ηλικιακών ομάδων 0+1 και 2+3";
age01 = age0 + age1;
age23 = age2 + age3;

"πλήθος ανά επάγγελμα";
occup0 = Count[XAoccup, 0];
occup1 = Count[XAoccup, 1];
occup2 = Count[XAoccup, 2];
occup3 = Count[XAoccup, 3];

"κατασκευή επαγγελμάτων για τις ηλικιακές ομάδες 1+2 (δεν έχει συνταξιούχους)";
occupage01 = Table[RandomSample[{occup0, occup1, occup2} → {0, 1, 2}][[1]], {age01}];

"κατασκευή επαγγελμάτων για τις ηλικιακές ομάδες 2+3 (δεν έχει φοιτητές)";
occupage23 = Table[RandomSample[{occup0, occup2, occup3} → {0, 2, 3}][[1]], {age23}];

"τελική κατασκευή όλων των επαγγελμάτων για τους ΣΟΡΤΑΡΙΣΜΕΝΟΥΣ ανά ηλικιακή ομάδα ψηφοφόρους";
XAoccup = Join[occupage01, occupage23];

"τελικό προφιλ ψηφοφόρων της ΧΑ, ΣΟΡΤΑΡΙΣΜΕΝΩΝ ανά ηλικιακή ομάδα";
XAvoter = Table[{4, XAage[[i]], XAoccup[[i]], XAgender[[i]], XAprevious[[i]]}, {i, 1, XApartie}];
```

```
"ΑΛΛΟ ΚΟΜΜΑ";
ALLOpartie = Count[parties, 5];

"κατασκευή φύλου και συσπείρωσης ψηφοφόρων";
ALLOgender = Table[RandomSample[{23.2, 27.4} → {0, 1}][[1]], {ALLOpartie}];
ALLOprevious = Table[RandomSample[{6, 12, 2, 2, 7, 71} → {0, 1, 2, 3, 4, 5}][[1]], {ALLOpartie}];
ALLOoccup = Table[RandomSample[{30.4, 34.5, 28.4, 15} → {0, 1, 2, 3}][[1]], {ALLOpartie}];


"κατασκευή ψηφοφόρων ΣΟΡΤΑΡΙΣΜΕΝΩΝ ανά ηλικιακή ομάδα";
ALLOage = Sort[Table[RandomSample[{35, 34.5, 27.4, 16.2} → {0, 1, 2, 3}][[1]], {ALLOpartie}]];

"πλήθος ανά ηλικιακή ομάδα";
age0 = Count[ALLOage, 0];
age1 = Count[ALLOage, 1];
age2 = Count[ALLOage, 2];
age3 = Count[ALLOage, 3];

"πλήθος ηλικιακών ομάδων 0+1 και 2+3";
age01 = age0 + age1;
age23 = age2 + age3;

"πλήθος ανά επάγγελμα";
occup0 = Count[ALLOoccup, 0];
occup1 = Count[ALLOoccup, 1];
occup2 = Count[ALLOoccup, 2];
occup3 = Count[ALLOoccup, 3];

"κατασκευή επαγγελμάτων για τις ηλικιακές ομάδες 1+2 (δεν έχει συνταξιούχους)";
occupage01 = Table[RandomSample[{occup0, occup1, occup2} → {0, 1, 2}][[1]], {age01}];

"κατασκευή επαγγελμάτων για τις ηλικιακές ομάδες 2+3 (δεν έχει φοιτητές)";
occupage23 = Table[RandomSample[{occup0, occup2, occup3} → {0, 2, 3}][[1]], {age23}];

"τελική κατασκευή όλων των επαγγελμάτων για τους ΣΟΡΤΑΡΙΣΜΕΝΟΥΣ ανά ηλικιακή ομάδα ψηφοφόρους";
ALLOoccup = Join[occupage01, occupage23];

"τελικό προφιλ ψηφοφόρων ΑΛΛΟΥ ΚΟΜΜΑΤΟΣ, ΣΟΡΤΑΡΙΣΜΕΝΩΝ ανά ηλικιακή ομάδα";
ALLOvoter = Table[{5, ALLOage[[i]], ALLOoccup[[i]], ALLOgender[[i]], ALLOprevious[[i]]}, {i, 1, ALLOpartie}];

"ένωση όλων των ψηφοφόρων από κάθε κόμμα";
voter = Join[NDvoter, SYRvoter, ELIAvoter, KKEvoter, XAvoter, ALLOvoter];

PrependTo[voter, {"Partie", "Age", "Occup", "Gender", "Previous"}];

"έλεγχος αν έγινε λάθος";
Do[
  If[voter[[i, 2]] == 0 && voter[[i, 3]] == 3, Print["λάθος"]];
  If[voter[[i, 2]] == 1 && voter[[i, 3]] == 3, Print["λάθος"]]; If[voter[[i, 2]] == 3 && voter[[i, 3]] == 1, Print["λάθος"]];
  If[voter[[i, 2]] == 4 && voter[[i, 3]] == 1, Print["λάθος"]];
  , {i, 1, exitpoll}];

"Εξαγωγή σε εξελόφυλλο στο φάκελο ΤΑ ΕΓΓΡΑΦΑ ΜΟΥ";
Export["voter.xls", voter];
```

4.3 Πίνακες συνάφειας

Στόχος της έρευνάς μας είναι να εξετάσει κατά πόσο οι μεταβλητές Age, Gender, Occupation και Previous Vote επηρεάζουν την ψήφο του i-ψηφοφόρου. Συγκεκριμένα, θα εξετάσουμε εάν το κόμμα που υποστηρίζουν οι 8000 ψηφοφόροι του δείγματος είναι ανεξάρτητο

α) του φύλου τους,

β) του επαγγέλματός τους,

γ) της ηλικίας τους,

δ) της προηγούμενης ψήφου τους.

Θα ξεκινήσουμε την ανάλυση ελέγχοντας κατά πόσο το κόμμα που υποστήριξαν οι ψηφοφόροι είναι ανεξάρτητο από το φύλο τους. Στον Πίνακα 4.1 παρουσιάζεται ο 2×6 πίνακας συνάφειας στον οποίο εμφανίζονται τα περιθώρια αθροίσματα γραμμών και στηλών καθώς και οι σχετικές συχνότητες.

Πίνακας 4.1: Αποτελέσματα του πίνακα συνάφειας

Gender * Party Crosstabulation

			Party						
			ND	SYRIZA	ELIA	KKE	XA	ALLO	Total
Gender	antras	Count	903	1016	354	240	529	934	3976
		Expected Count	916,0	1086,9	371,8	224,6	366,8	1009,9	3976,0
		% within Gender	22,7%	25,6%	8,9%	6,0%	13,3%	23,5%	100,0%
		% within Party	49,0%	46,5%	47,3%	53,1%	71,7%	46,0%	49,7%
		% of Total	11,3%	12,7%	4,4%	3,0%	6,6%	11,7%	49,7%
	gunaika	Count	940	1171	394	212	209	1098	4024
		Expected Count	927,0	1100,1	376,2	227,4	371,2	1022,1	4024,0
		% within Gender	23,4%	29,1%	9,8%	5,3%	5,2%	27,3%	100,0%
		% within Party	51,0%	53,5%	52,7%	46,9%	28,3%	54,0%	50,3%
		% of Total	11,8%	14,6%	4,9%	2,7%	2,6%	13,7%	50,3%
Total		Count	1843	2187	748	452	738	2032	8000
		Expected Count	1843,0	2187,0	748,0	452,0	738,0	2032,0	8000,0
		% within Gender	23,0%	27,3%	9,4%	5,7%	9,2%	25,4%	100,0%
		% within Party	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
		% of Total	23,0%	27,3%	9,4%	5,7%	9,2%	25,4%	100,0%

Η ερμηνεία των ποσοστών που παρουσιάζονται στα γραμμοσκιασμένα κελιά είναι η εξής:

- Το 25.6% των αντρών ψηφίζουν ΣΥ.ΡΙΖ.Α,
- Το 46,5% των ψηφοφόρων που ψηφίζουν ΣΥ.ΡΙΖ.Α είναι άντρες,
- Το 12,7% του δείγματος είναι άντρες που ψηφίζουν ΣΥ.ΡΙΖ.Α,

- Το 50,3% του δείγματος αποτελείται από γυναίκες,
- Το 5,7% του δείγματος ψηφίζει ΚΚΕ.

Για να ελέγξουμε εάν το κόμμα που υποστήριξαν οι ψηφοφόροι είναι ανεξάρτητο από το φύλο τους θα πρέπει να εφαρμόσομε τον έλεγχο ανεξαρτησίας χ^2. Από τον Πίνακα 4.2 με τα αποτελέσματα του ελέγχου παρατηρούμε ότι η τιμή της στατιστικής συνάρτησης ισούται με 167,309 ενώ το αντίστοιχο p-value είναι μικρότερο του 0,001. Επομένως, σε επίπεδο σημαντικότητας $a = 0.05$, απορρίπτουμε τη μηδενική υπόθεση της ανεξαρτησίας των δύο μεταβλητών και συμπεραίνουμε ότι το ποιο κόμμα ψηφίζουν τα άτομα επηρεάζεται στατιστικά σημαντικά από το φύλο τους.

Πίνακας 4.2: Αποτελέσματα του ελέγχου ανεξαρτησίας χ^2

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	167,309[a]	5	,000
Likelihood Ratio	172,041	5	,000
Linear-by-Linear Association	5,693	1	,017
N of Valid Cases	8000		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 224,64.

Στον Πίνακα 4.3 παρουσιάζονται τόσο οι παρατηρούμενες τιμές, όσο και οι αναμενόμενες. Για παράδειγμα, εάν το κόμμα και το φύλο ήταν ανεξάρτητα μεταξύ τους, θα αναμέναμε 916 άντρες να έχουν ψηφίσει ΝΔ (αντί για 903 που έχουν παρατηρηθεί στο δείγμα). Ομοίως, αντί για 394, θα αναμέναμε 376,2 γυναίκες να έχουν ψηφίσει ΕΛΙΑ.

Πίνακας 4.3: Παρατηρούμενες και αναμενόμενες τιμές

Gender * Party Crosstabulation

			Party						
			ND	SYRIZA	ELIA	KKE	XA	ALLO	Total
Gender	antras	Count	903	1016	354	240	529	934	3976
		Expected Count	916,0	1086,9	371,8	224,6	366,8	1009,9	3976,0
		% of Total	11,3%	12,7%	4,4%	3,0%	6,6%	11,7%	49,7%
	gunaika	Count	940	1171	394	212	209	1098	4024
		Expected Count	927,0	1100,1	376,2	227,4	371,2	1022,1	4024,0
		% of Total	11,8%	14,6%	4,9%	2,7%	2,6%	13,7%	50,3%
Total		Count	1843	2187	748	452	738	2032	8000
		Expected Count	1843,0	2187,0	748,0	452,0	738,0	2032,0	8000,0
		% of Total	23,0%	27,3%	9,4%	5,7%	9,2%	25,4%	100,0%

Τέλος, στον Πίνακα 4.4 παρουσιάζονται τα τυποποιημένα κατάλοιπα. Παρατηρούμε ότι τα τυποποιημένα κατάλοιπα των γυναικών ψηφοφόρων που ψηφίζουν ΣΥ.ΡΙΖ.Α είναι θετικά. Αυτό σημαίνει ότι οι γυναίκες ψηφοφόροι ψηφίζουν πιο συχνά ΣΥ.ΡΙΖ.Α σε σχέση με τους άντρες. Αντίστοιχα, οι άντρες ψηφοφόροι ψηφίζουν πιο συχνά ΧΑ σε σχέση με τις γυναίκες.

Πίνακας 4.4: Πίνακας τυποποιημένων καταλοίπων

Gender * Party Crosstabulation

			Party						Total
			ND	SYRIZA	ELIA	KKE	XA	ALLO	
Gender	antras	Count	903	1016	354	240	529	934	3976
		Expected Count	916,0	1086,9	371,8	224,6	366,8	1009,9	3976,0
		Std. Residual	-,4	-2,2	-,9	1,0	8,5	-2,4	
		Adjusted Residual	-,7	-3,6	-1,4	1,5	12,5	-3,9	
	gunaika	Count	940	1171	394	212	209	1098	4024
		Expected Count	927,0	1100,1	376,2	227,4	371,2	1022,1	4024,0
		Std. Residual	,4	2,1	,9	-1,0	-8,4	2,4	
		Adjusted Residual	,7	3,6	1,4	-1,5	-12,5	3,9	
Total		Count	1843	2187	748	452	738	2032	8000
		Expected Count	1843,0	2187,0	748,0	452,0	738,0	2032,0	8000,0

Στη συνέχεια, θα εξετάσουμε κατά πόσο το κόμμα που υποστήριξαν οι ψηφοφόροι είναι ανεξάρτητο από τη θέση τους στην εργασία. Από τον Πίνακα 4.5 συμπεραίνουμε τα εξής:

- Το 25.2% των φοιτητών ψηφίζει ΣΥ.ΡΙΖ.Α ενώ, μόλις το 3,8% των φοιτητών ψηφίζει ΕΛΙΑ,
- Το 10,6% των ψηφοφόρων που ψηφίζουν ΧΑ είναι άνεργοι,
- Το 7% του δείγματος είναι συνταξιούχοι που ψηφίζουν ΝΔ,
- Το 34,8% του δείγματος αποτελείται από εργαζόμενους του ιδιωτικού ή του δημόσιου τομέα.

Πίνακας 4.5: Αποτελέσματα του πίνακα συνάφειας

Occupation * Party Crosstabulation

			Party						Total
			ND	SYRIZA	ELIA	KKE	XA	ALLO	
Occupation	ergazomenos	Count	575	802	256	142	263	747	2785
		% within Occupation	20,6%	28,8%	9,2%	5,1%	9,4%	26,8%	100,0%
		% within Party	31,2%	36,7%	34,2%	31,4%	35,6%	36,8%	34,8%
		% of Total	7,2%	10,0%	3,2%	1,8%	3,3%	9,3%	34,8%
	foithths	Count	216	332	50	112	126	482	1318
		% within Occupation	16,4%	25,2%	3,8%	8,5%	9,6%	36,6%	100,0%
		% within Party	11,7%	15,2%	6,7%	24,8%	17,1%	23,7%	16,5%
		% of Total	2,7%	4,2%	0,6%	1,4%	1,6%	6,0%	16,5%
	anergos	Count	490	759	197	125	261	637	2469
		% within Occupation	19,8%	30,7%	8,0%	5,1%	10,6%	25,8%	100,0%
		% within Party	26,6%	34,7%	26,3%	27,7%	35,4%	31,3%	30,9%
		% of Total	6,1%	9,5%	2,5%	1,6%	3,3%	8,0%	30,9%
	suntaxiouxos	Count	562	294	245	73	88	166	1428
		% within Occupation	39,4%	20,6%	17,2%	5,1%	6,2%	11,6%	100,0%
		% within Party	30,5%	13,4%	32,8%	16,2%	11,9%	8,2%	17,9%
		% of Total	7,0%	3,7%	3,1%	0,9%	1,1%	2,1%	17,9%
Total		Count	1843	2187	748	452	738	2032	8000
		% within Occupation	23,0%	27,3%	9,4%	5,7%	9,2%	25,4%	100,0%
		% within Party	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
		% of Total	23,0%	27,3%	9,4%	5,7%	9,2%	25,4%	100,0%

Από τον Πίνακα 4.6 με τα αποτελέσματα του ελέγχου παρατηρούμε ότι η τιμή της στατιστικής συνάρτησης ισούται με 604,606 ενώ το αντίστοιχο *p*-value είναι μικρότερο του 0,001. Επομένως, σε επίπεδο σημαντικότητας $a = 0.05$, απορρίπτουμε τη μηδενική υπόθεση της ανεξαρτησίας των δύο μεταβλητών και συμπεραίνουμε ότι το ποιο κόμμα ψηφίζουν τα άτομα επηρεάζεται στατιστικά σημαντικά από τη θέση τους στην απασχόληση.

Πίνακας 4.6: Αποτελέσματα του ελέγχου ανεξαρτησίας χ^2

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	604,606[a]	15	,000
Likelihood Ratio	593,987	15	,000
Linear-by-Linear Association	120,088	1	,000
N of Valid Cases	8000		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 74,47.

Η ερμηνεία των Πινάκων 4.7 και 4.8 είναι παρόμοια με εκείνη των Πινάκων 4.3 και 4.4. Ενδεικτικά αναφέρουμε ότι εάν το κόμμα και το επάγγελμα ήταν ανεξάρτητα μεταξύ τους, θα αναμέναμε 329 συνταξιούχους να έχουν ψηφίσει ΝΔ αντί για 562 που έχουν παρατηρηθεί στο δείγμα. Αντίστοιχα, τα τυποποιημένα κατάλοιπα των συνταξιούχων ψηφοφόρων που ψηφίζουν ΕΛΙΑ είναι θετικά. Αυτό σημαίνει ότι οι συνταξιούχοι ψηφίζουν πιο συχνά ΕΛΙΑ συγκριτικά με τους υπόλοιπους.

Πίνακας 4.7: Παρατηρούμενες και αναμενόμενες τιμές

Occupation * Party Crosstabulation

			Party						Total
			ND	SYRIZA	ELIA	KKE	XA	ALLO	
Occupation	ergazomenos	Count	575	802	256	142	263	747	2785
		Expected Count	641,6	761,3	260,4	157,4	256,9	707,4	2785,0
	foithths	Count	216	332	50	112	126	482	1318
		Expected Count	303,6	360,3	123,2	74,5	121,6	334,8	1318,0
	anergos	Count	490	759	197	125	261	637	2469
		Expected Count	568,8	675,0	230,9	139,5	227,8	627,1	2469,0
	suntaxiouxos	Count	562	294	245	73	88	166	1428
		Expected Count	329,0	390,4	133,5	80,7	131,7	362,7	1428,0
Total		Count	1843	2187	748	452	738	2032	8000
		Expected Count	1843,0	2187,0	748,0	452,0	738,0	2032,0	8000,0

Πίνακας 4.8: Πίνακας τυποποιημένων καταλοίπων

Occupation * Party Crosstabulation

			Party						Total
			ND	SYRIZA	ELIA	KKE	XA	ALLO	
Occupation	ergazomenos	Count	575	802	256	142	263	747	2785
		Expected Count	641,6	761,3	260,4	157,4	256,9	707,4	2785,0
		Std. Residual	-2,6	1,5	-,3	-1,2	,4	1,5	
		Adjusted Residual	-3,7	2,1	-,4	-1,6	,5	2,1	
	foithths	Count	216	332	50	112	126	482	1318
		Expected Count	303,6	360,3	123,2	74,5	121,6	334,8	1318,0
		Std. Residual	-5,0	-1,5	-6,6	4,3	,4	8,0	
		Adjusted Residual	-6,3	-1,9	-7,6	4,9	,5	10,2	
	anergos	Count	490	759	197	125	261	637	2469
		Expected Count	568,8	675,0	230,9	139,5	227,8	627,1	2469,0
		Std. Residual	-3,3	3,2	-2,2	-1,2	2,2	,4	
		Adjusted Residual	-4,5	4,6	-2,8	-1,5	2,8	,5	
	suntaxiouxos	Count	562	294	245	73	88	166	1428
		Expected Count	329,0	390,4	133,5	80,7	131,7	362,7	1428,0
		Std. Residual	12,8	-4,9	9,6	-,9	-3,8	-10,3	
		Adjusted Residual	16,2	-6,3	11,2	-1,0	-4,4	-13,2	
Total		Count	1843	2187	748	452	738	2032	8000
		Expected Count	1843,0	2187,0	748,0	452,0	738,0	2032,0	8000,0

Ο επόμενος έλεγχος θα έχει να κάνει με το κατά πόσο το κόμμα που υποστήριξαν οι ψηφοφόροι στις ευρωεκλογές του 2014 είναι ανεξάρτητο από το κόμμα που είχαν ψηφίσει στις βουλευτικές εκλογές του 2012. Από τον Πίνακα 4.9 συμπεραίνουμε τα εξής:

- Το 82,4% των ψηφοφόρων που είχαν ψηφίσει ΝΔ το 2012, την ξαναψήφισαν και στις ευρωεκλογές του 2014,
- Το 67,2% των ψηφοφόρων που ψήφισαν ΝΔ στις ευρωεκλογές του 2014, την ξαναψηφίσει και το 2012,
- Το 15,5% του δείγματος αποτελείται από άτομα τα οποία ψήφισαν και στις δύο εκλογές ΝΔ,
- Το 10,6% του δείγματος αποτελείται από άτομα τα οποία στις βουλευτικές εκλογές είχαν ψηφίσει ΧΑ.

Πίνακας 4.9: Αποτελέσματα του πίνακα συνάφειας

Previous * Party Crosstabulation

			Party						
			ND	SYRIZA	ELIA	KKE	XA	ALLO	Total
Previous	ND	Count	1238	37	44	6	60	118	1503
		Expected Count	346,3	410,9	140,5	84,9	138,7	381,8	1503,0
		% within Previous	82,4%	2,5%	2,9%	0,4%	4,0%	7,9%	100,0%
		% within Party	67,2%	1,7%	5,9%	1,3%	8,1%	5,8%	18,8%
		% of Total	15,5%	0,5%	0,6%	0,1%	0,8%	1,5%	18,8%
	SYRIZA	Count	116	1655	197	53	26	238	2285
		Expected Count	526,4	624,7	213,6	129,1	210,8	580,4	2285,0
		% within Previous	5,1%	72,4%	8,6%	2,3%	1,1%	10,4%	100,0%
		% within Party	6,3%	75,7%	26,3%	11,7%	3,5%	11,7%	28,6%
		% of Total	1,5%	20,7%	2,5%	0,7%	0,3%	3,0%	28,6%
	PASOK	Count	44	47	314	0	15	47	467
		Expected Count	107,6	127,7	43,7	26,4	43,1	118,6	467,0
		% within Previous	9,4%	10,1%	67,2%	0,0%	3,2%	10,1%	100,0%
		% within Party	2,4%	2,1%	42,0%	0,0%	2,0%	2,3%	5,8%
		% of Total	0,6%	0,6%	3,9%	0,0%	0,2%	0,6%	5,8%
	KKE	Count	18	89	11	327	7	35	487
		Expected Count	112,2	133,1	45,5	27,5	44,9	123,7	487,0
		% within Previous	3,7%	18,3%	2,3%	67,1%	1,4%	7,2%	100,0%
		% within Party	1,0%	4,1%	1,5%	72,3%	0,9%	1,7%	6,1%
		% of Total	0,2%	1,1%	0,1%	4,1%	0,1%	0,4%	6,1%
	XA	Count	104	46	20	14	522	145	851
		Expected Count	196,0	232,6	79,6	48,1	78,5	216,2	851,0
		% within Previous	12,2%	5,4%	2,4%	1,6%	61,3%	17,0%	100,0%
		% within Party	5,6%	2,1%	2,7%	3,1%	70,7%	7,1%	10,6%
		% of Total	1,3%	0,6%	0,3%	0,2%	6,5%	1,8%	10,6%
	ALLO	Count	323	313	162	52	108	1449	2407
		Expected Count	554,5	658,0	225,1	136,0	222,0	611,4	2407,0
		% within Previous	13,4%	13,0%	6,7%	2,2%	4,5%	60,2%	100,0%
		% within Party	17,5%	14,3%	21,7%	11,5%	14,6%	71,3%	30,1%
		% of Total	4,0%	3,9%	2,0%	0,7%	1,4%	18,1%	30,1%
Total		Count	1843	2187	748	452	738	2032	8000
		Expected Count	1843,0	2187,0	748,0	452,0	738,0	2032,0	8000,0
		% within Previous	23,0%	27,3%	9,4%	5,7%	9,2%	25,4%	100,0%
		% within Party	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
		% of Total	23,0%	27,3%	9,4%	5,7%	9,2%	25,4%	100,0%

Από τον Πίνακα 4.10 με τα αποτελέσματα του ελέγχου παρατηρούμε ότι η τιμή της στατιστικής συνάρτησης ισούται με 15102,097 ενώ το αντίστοιχο p-value είναι μικρότερο του 0,001. Επομένως, σε επίπεδο σημαντικότητας $a = 0.05$, απορρίπτουμε τη μηδενική υπόθεση της ανεξαρτησίας των δύο μεταβλητών και συμπεραίνουμε ότι το ποιο κόμμα ψηφίζουν τα άτομα σήμερα επηρεάζεται στατιστικά σημαντικά από το κόμμα που είχαν ψηφίσει στις προηγούμενες εκλογές.

Πίνακας 4.10: Αποτελέσματα του ελέγχου ανεξαρτησίας χ^2

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	15102,097[a]	25	,000
Likelihood Ratio	9778,440	25	,000
Linear-by-Linear Association	2602,161	1	,000
N of Valid Cases	8000		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 26,39.

Όσον αφορά την ερμηνεία των Πινάκων 4.11 και 4.12, ενδεικτικά αναφέρουμε ότι εάν το κόμμα που ψήφισε σήμερα ο ψηφοφόρος και το κόμμα που είχε ψηφίσει στις προηγούμενες εκλογές ήταν ανεξάρτητα μεταξύ τους, θα αναμέναμε 410,9 άτομα, που στις εκλογές του 2012 είχαν ψηφίσει ΝΔ, να ψηφίσουν ΣΥ.ΡΙΖ.Α αντί για 37 που έχουν παρατηρηθεί στο δείγμα. Αντίστοιχα, τα άτομα που στις εκλογές του 2012 είχαν ψηφίσει ΝΔ ψηφίζουν πιο συχνά ΝΔ συγκριτικά με εκείνους που είχαν ψηφίσει κάποιο από τα υπόλοιπα κόμματα.

Πίνακας 4.11: Παρατηρούμενες και αναμενόμενες τιμές

Previous * Party Crosstabulation

			Party						Total
			ND	SYRIZA	ELIA	KKE	XA	ALLO	
Previous	ND	Count	1238	37	44	6	60	118	1503
		Expected Count	346,3	410,9	140,5	84,9	138,7	381,8	1503,0
	SYRIZA	Count	116	1655	197	53	26	238	2285
		Expected Count	526,4	624,7	213,6	129,1	210,8	580,4	2285,0
	PASOK	Count	44	47	314	0	15	47	467
		Expected Count	107,6	127,7	43,7	26,4	43,1	118,6	467,0
	KKE	Count	18	89	11	327	7	35	487
		Expected Count	112,2	133,1	45,5	27,5	44,9	123,7	487,0
	XA	Count	104	46	20	14	522	145	851
		Expected Count	196,0	232,6	79,6	48,1	78,5	216,2	851,0
	ALLO	Count	323	313	162	52	108	1449	2407
		Expected Count	554,5	658,0	225,1	136,0	222,0	611,4	2407,0
Total		Count	1843	2187	748	452	738	2032	8000
		Expected Count	1843,0	2187,0	748,0	452,0	738,0	2032,0	8000,0

Πίνακας 4.12: Πίνακας τυποποιημένων καταλοίπων

Previous * Party Crosstabulation

			Party						Total
			ND	SYRIZA	ELIA	KKE	XA	ALLO	
Previous	ND	Count	1238	37	44	6	60	118	1503
		Std. Residual	47,9	-18,4	-8,1	-8,6	-6,7	-13,5	
		Adjusted Residual	60,6	-24,0	-9,5	-9,8	-7,8	-17,3	
	SYRIZA	Count	116	1655	197	53	26	238	2285
		Std. Residual	-17,9	41,2	-1,1	-6,7	-12,7	-14,2	
		Adjusted Residual	-24,1	57,2	-1,4	-8,2	-15,8	-19,5	
	PASOK	Count	44	47	314	0	15	47	467
		Std. Residual	-6,1	-7,1	40,9	-5,1	-4,3	-6,6	
		Adjusted Residual	-7,2	-8,6	44,3	-5,4	-4,6	-7,8	
	KKE	Count	18	89	11	327	7	35	487
		Std. Residual	-8,9	-3,8	-5,1	57,1	-5,7	-8,0	
		Adjusted Residual	-10,5	-4,6	-5,5	60,7	-6,1	-9,5	
	XA	Count	104	46	20	14	522	145	851
		Std. Residual	-6,6	-12,2	-6,7	-4,9	50,1	-4,8	
		Adjusted Residual	-7,9	-15,2	-7,4	-5,4	55,6	-5,9	
	ALLO	Count	323	313	162	52	108	1449	2407
		Std. Residual	-9,8	-13,4	-4,2	-7,2	-7,7	33,9	
		Adjusted Residual	-13,4	-18,9	-5,3	-8,9	-9,6	46,9	
Total		Count	1843	2187	748	452	738	2032	8000

Τέλος, θέλουμε να ελέγξουμε εάν το κόμμα που υποστήριξαν οι ψηφοφόροι είναι ανεξάρτητο από την ηλικιακή ομάδα στην οποία ανήκουν. Επειδή η μία μεταβλητή είναι ονομαστικής κλίμακας (Party) και η άλλη διατάξιμης κλίμακας (Age) θα πρέπει να εφαρμόσουμε τον έλεγχο χ^2 για γραμμική τάση. Από τον Πίνακα 4.13 προκύπτουν τα ακόλουθα συμπεράσματα:

- Το 4,1% των ψηφοφόρων που ανήκουν στην ηλικιακή ομάδα 35-54 ψηφίζουν ΚΚΕ,
- Το 17,5% των ψηφοφόρων που ψηφίζουν ΚΚΕ είναι μεταξύ 35 και 54 ετών,
- Το 1% του δείγματος αποτελείται από άτομα ηλικίας 35 εώς 54 ετών που ψηφίζουν ΚΚΕ,
- Το 24,3% του δείγματος αποτελείται από ψηφοφόρους που ανήκουν στην ηλικιακή ομάδα 35-54.

Πίνακας 4.13: Αποτελέσματα του πίνακα συνάφειας

Age * Party Crosstabulation

			Party						
			ND	SYRIZA	ELIA	KKE	XA	ALLO	Total
Age	18-24	Count	344	493	154	126	213	612	1942
		% within Age	17,7%	25,4%	7,9%	6,5%	11,0%	31,5%	100,0%
		% within Party	18,7%	22,5%	20,6%	27,9%	28,9%	30,1%	24,3%
		% of Total	4,3%	6,2%	1,9%	1,6%	2,7%	7,7%	24,3%
	25-34	Count	380	533	106	105	199	656	1979
		% within Age	19,2%	26,9%	5,4%	5,3%	10,1%	33,1%	100,0%
		% within Party	20,6%	24,4%	14,2%	23,2%	27,0%	32,3%	24,7%
		% of Total	4,8%	6,7%	1,3%	1,3%	2,5%	8,2%	24,7%
	35-54	Count	404	602	173	79	195	494	1947
		% within Age	20,7%	30,9%	8,9%	4,1%	10,0%	25,4%	100,0%
		% within Party	21,9%	27,5%	23,1%	17,5%	26,4%	24,3%	24,3%
		% of Total	5,1%	7,5%	2,2%	1,0%	2,4%	6,2%	24,3%
	55+	Count	715	559	315	142	131	270	2132
		% within Age	33,5%	26,2%	14,8%	6,7%	6,1%	12,7%	100,0%
		% within Party	38,8%	25,6%	42,1%	31,4%	17,8%	13,3%	26,7%
		% of Total	8,9%	7,0%	3,9%	1,8%	1,6%	3,4%	26,7%
Total		Count	1843	2187	748	452	738	2032	8000
		% within Age	23,0%	27,3%	9,4%	5,7%	9,2%	25,4%	100,0%
		% within Party	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
		% of Total	23,0%	27,3%	9,4%	5,7%	9,2%	25,4%	100,0%

Στη συνέχεια, προκύπτει ο Πίνακας 4.14 με το αποτέλεσμα του ελέγχου στον οποίο αναζητούμε την τιμή Linear-by-Linear Association που μετράει την γραμμική τάση. Παρατηρούμε ότι η τιμή της στατιστικής συνάρτησης ισούται με 519,453 ενώ το αντίστοιχο p-value είναι μικρότερο του 0,001 οπότε σε επίπεδο σημαντικότητας $a = 0.05$ απορρίπτουμε τη μηδενική υπόθεση και συμπεραίνουμε ότι η ψήφος εξαρτάται από την ηλικία των ψηφοφόρων.

Πίνακας 4.14: Αποτελέσματα του ελέγχου ανεξαρτησίας χ^2

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	519,453[a]	15	,000
Likelihood Ratio	533,783	15	,000
Linear-by-Linear Association	291,044	1	,000
N of Valid Cases	8000		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 109,72.

Κλείνοντας, από τους Πίνακες 4.15 και 4.16 ενδεικτικά αναφέρουμε τα εξής:

- Εάν το κόμμα και η ηλικία ήταν ανεξάρτητα μεταξύ τους, αντί για 493, θα αναμέναμε 530,9 άτομα ηλικίας 18-24 να έχουν ψηφίσει ΣΥ.ΡΙΖ.Α,
- Οι ψηφοφόροι που είναι μεγαλύτεροι των 55 ετών τείνουν να ψηφίζουν πιο συχνά ΝΔ σε σχέση με τους υπόλοιπους.

Πίνακας 4.15: Παρατηρούμενες και αναμενόμενες τιμές

Age * Party Crosstabulation

			Party						Total
			ND	SYRIZA	ELIA	KKE	XA	ALLO	
Age	18-24	Count	344	493	154	126	213	612	1942
		Expected Count	447,4	530,9	181,6	109,7	179,1	493,3	1942,0
	25-34	Count	380	533	106	105	199	656	1979
		Expected Count	455,9	541,0	185,0	111,8	182,6	502,7	1979,0
	35-54	Count	404	602	173	79	195	494	1947
		Expected Count	448,5	532,3	182,0	110,0	179,6	494,5	1947,0
	55+	Count	715	559	315	142	131	270	2132
		Expected Count	491,2	582,8	199,3	120,5	196,7	541,5	2132,0
Total		Count	1843	2187	748	452	738	2032	8000
		Expected Count	1843,0	2187,0	748,0	452,0	738,0	2032,0	8000,0

Πίνακας 4.16: Πίνακας τυποποιημένων καταλοίπων

Age * Party Crosstabulation

			Party						Total
			ND	SYRIZA	ELIA	KKE	XA	ALLO	
Age	18-24	Count	344	493	154	126	213	612	1942
		Std. Residual	-4,9	-1,6	-2,0	1,6	2,5	5,3	
		Adjusted Residual	-6,4	-2,2	-2,5	1,8	3,1	7,1	
	25-34	Count	380	533	106	105	199	656	1979
		Std. Residual	-3,6	-,3	-5,8	-,6	1,2	6,8	
		Adjusted Residual	-4,7	-,5	-7,0	-,8	1,5	9,1	
	35-54	Count	404	602	173	79	195	494	1947
		Std. Residual	-2,1	3,0	-,7	-3,0	1,1	,0	
		Adjusted Residual	-2,8	4,1	-,8	-3,5	1,4	,0	
	55+	Count	715	559	315	142	131	270	2132
		Std. Residual	10,1	-1,0	8,2	2,0	-4,7	-11,7	
		Adjusted Residual	13,4	-1,4	10,0	2,4	-5,7	-15,8	
Total		Count	1843	2187	748	452	738	2032	8000

4.3.1 Το παράδοξο του Simpson

Στην προηγούμενη παράγραφο, διαπιστώσαμε ότι η ψήφος κάθε ατόμου εξαρτάται από τη θέση του στην απασχόληση. Υπάρχει όμως το ενδεχόμενο, εάν εξετάσουμε τους πίνακες συνάφειας ξεχωριστά για άντρες και γυναίκες, να παρατηρήσουμε ότι δεν υπάρχει καμία εξάρτηση μεταξύ ψήφου και θέσης στην απασχόληση. Δηλαδή, η κατεύθυνση της περιθωριακής συνάφειας να είναι αντίστροφη από αυτή των δεσμευμένων συναφειών. Αυτό αναφέρεται ως *παράδοξο του Simpson* και οφείλεται πιθανότερα στο γεγονός ότι η μεταβλητή "φύλο" που προσθέσαμε παρουσιάζει ισχυρή συνάφεια με τις άλλες δύο μεταβλητές (ψήφος και θέση στην απασχόληση).

Το γεγονός αυτό μπορεί να δικαιολογηθεί και στατιστικά εάν κάνουμε τον έλεγχο υποθέσεων για ανεξαρτησία μεταξύ των μεταβλητών Gender- Party και Gender – Occupation από τον οποίο θα παρατηρήσουμε ότι όντως υπάρχει ισχυρή συνάφεια μεταξύ του φύλου και των άλλων δύο μεταβλητών (p-value < 0.05). Εάν πάμε τώρα να εφαρμόσουμε ξανά τον έλεγχο ανεξαρτησίας των Mantel – Haenszel ο οποίος, ελέγχει τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος κάθε υποπίνακα είναι ίσος με 1 (εναλλακτικά, η ψήφος δε σχετίζεται με τη θέση στην απασχόληση λαμβάνοντας υπόψη το φύλο) θα καταλήξουμε τελικά στο συμπέρασμα ότι η ψήφος και η θέση στην απασχόληση είναι υπό συνθήκη ανεξάρτητες δοθέντος του επιπέδου του φύλου (p-value > 0.05).

Κάτι αντίστοιχο μπορεί να ισχύσει και για οποιαδήποτε άλλη μεταβλητή η οποία παρουσιάζει ισχυρή συνάφεια με κάποιες από τις υπόλοιπες.

4.4 Σχετικοί λόγοι πιθανοτήτων (odds ratio)

Θα συνεχίσουμε την ανάλυσή μας υπολογίζοντας, ξεχωριστά για κάθε κόμμα αυτή τη φορά, το σχετικό λόγο πιθανοτήτων μεταξύ μίας σειράς ενδεχομένων. Στις παραγράφους που ακολουθούν, θα περιγράψουμε αναλυτικά τον υπολογισμό και την ερμηνεία των σχετικών λόγων πιθανοτήτων στην περίπτωση της ΝΔ και του ΣΥ.ΡΙΖ.Α, με τη διαδικασία να είναι η ίδια και για όλα τα υπόλοιπα κόμματα.

4.4.1 Η περίπτωση της ΝΔ

Ξεκινώντας από την περίπτωση της ΝΔ, κατασκευάζουμε αρχικά μέσω του SPSS μία νέα κατηγορική μεταβλητή (**Action_ND**) ακολουθώντας τη διαδικασία

Transform – Recode into different variables

η οποία θα είναι η εξαρτημένη μεταβλητή και θα παίρνει την τιμή 1 εάν ο i-ψηφοφόρος "ψήφισε" ΝΔ και την τιμή 0 διαφορετικά.

Σε πρώτη φάση θα υπολογίσουμε το σχετικό λόγο πιθανοτήτων για το ζεύγος Action_ND – Gender. Από τα αποτελέσματα του Πίνακα 4.17 παρατηρούμε ότι:

- *odds ratio* = 1,037,
- το 95% δ.ε για το *odds ratio* είναι [0,935 , 1,151].

Αυτό σημαίνει ότι η σχετική πιθανότητα να ψηφίσει ΝΔ μία γυναίκα ψηφοφόρος είναι 0,037 (1,037 - 1) φορές μεγαλύτερη από τη σχετική πιθανότητα να ψηφίσει ΝΔ ένας άντρας ψηφοφόρος. Επιπλέον, επειδή το 95% δ.ε περιλαμβάνει την τιμή 1, δε μπορούμε να απορρίψουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στη ΝΔ δεν είναι διαφορετικά στις δύο ομάδες. Το συμπέρασμα αυτό ισοδυναμεί πλήρως με την μη-απόρριψη της μηδενικής υπόθεσης της ισότητας των επιμέρους σχετικών λόγων με τη μονάδα μέσω του ελέγχου Mantel-Haenszel (Πίνακας 4.18).

Πίνακας 4.17: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Gender (antras / gunaika)	1,037	,935	1,151
For cohort Action_ND = den pshfise ND	1,008	,985	1,033
For cohort Action_ND = pshfise ND	,972	,897	1,053
N of Valid Cases	8000		

Πίνακας 4.18: Αποτελέσματα του ελέγχου των Mantel-Haenszel

Mantel-Haenszel Common Odds Ratio Estimate

Estimate			1,037
ln(Estimate)			,037
Std. Error of ln(Estimate)			,053
Asymp. Sig. (2-sided)			,491
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound	,935
		Upper Bound	1,151
	ln(Common Odds Ratio)	Lower Bound	-,068
		Upper Bound	,141

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1,000 assumption. So is the natural log of the estimate.

Tests of Conditional Independence

	Chi-Squared	df	Asymp. Sig. (2-sided)
Cochran's	,474	1	,491
Mantel-Haenszel	,439	1	,508

Under the conditional independence assumption, Cochran's statistic is asymptotically distributed as a 1 df chi-squared distribution, only if the number of strata is fixed, while the Mantel-Haenszel statistic is always asymptotically distributed as a 1 df chi-squared distribution. Note that the continuity correction is removed from the Mantel-Haenszel statistic when the sum of the differences between the observed and the expected is 0.

Στη συνέχεια, με τον ίδιο τρόπο, θα υπολογίσουμε το σχετικό λόγο πιθανοτήτων για τα ζεύγη

- Action_ND – Age
- Action_ND – Occupation
- Action_ND – Previous

Η ερμηνεία των Πινάκων 4.19 – 4.33 είναι παρόμοια με εκείνη του Πίνακα 4.17. Πιο αναλυτικά, από τα αποτελέσματα που προκύπτουν, καταλήγουμε στα ακόλουθα συμπεράσματα:

Πίνακας 4.19: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Age (18-24 / 25-34)	1,104	,939	1,297
For cohort Action_ND = den pshfise ND	1,018	,989	1,049
For cohort Action_ND = pshfise ND	,923	,809	1,052
N of Valid Cases	3921		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 25-34 είναι 0,104 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 18-24.
- ✓ Επειδή το 95% δ.ε περιλαμβάνει την τιμή 1, δε μπορούμε να απορρίψουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στη ΝΔ δεν είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.20: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Age (18-24 / 35-54)	1,216	1,037	1,427
For cohort Action_ND = den pshfise ND	1,038	1,007	1,071
For cohort Action_ND = pshfise ND	,854	,750	,972
N of Valid Cases	3889		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 35-54 είναι 0,216 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 18-24.
- ✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στη ΝΔ είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.21: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Age (18-24 / 55+)	2,344	2,023	2,716
For cohort Action_ND = den pshfise ND	1,238	1,194	1,284
For cohort Action_ND = pshfise ND	,528	,472	,591
N of Valid Cases	4074		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που είναι μεγαλύτερος από 55 ετών είναι 1,344 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 18-24.
- ✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στη ΝΔ είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.22: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Age (25-34 / 35-54)	1,102	,942	1,288
For cohort Action_ND = den pshfise ND	1,020	,988	1,052
For cohort Action_ND = pshfise ND	,925	,816	1,049
N of Valid Cases	3926		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 35-54 είναι 0,102 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 25-34.
- ✓ Επειδή το 95% δ.ε περιλαμβάνει την τιμή 1, δε μπορούμε να απορρίψουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στη ΝΔ δεν είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.23: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Age (25-34 / 55+)	2,123	1,839	2,451
For cohort Action_ND = den pshfise ND	1,216	1,172	1,262
For cohort Action_ND = pshfise ND	,573	,514	,638
N of Valid Cases	4111		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που είναι μεγαλύτερος από 55 ετών είναι 1,123 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 25-34.
- ✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στη ΝΔ είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.24: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Age (35-54 / 55+)	1,927	1,673	2,221
For cohort Action_ND = den pshfise ND	1,192	1,148	1,238
For cohort Action_ND = pshfise ND	,619	,557	,687
N of Valid Cases	4079		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που είναι μεγαλύτερος από 55 ετών είναι 0,927 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 35-54.
- ✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στη ΝΔ είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.25: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Occupation (ergazomenos / foithths)	,753	,634	,895
For cohort Action_ND = den pshfise ND	,949	,921	,978
For cohort Action_ND = pshfise ND	1,260	1,093	1,452
N of Valid Cases	4103		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΝΔ ένας φοιτητής είναι 0,247 φορές μικρότερη από την σχετική πιθανότητα να ψηφίσει ΝΔ ένας εργαζόμενος.
- ✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στη ΝΔ είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.26: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Occupation (ergazomenos / anergos)	,952	,832	1,089
For cohort Action_ND = den pshfise ND	,990	,963	1,017
For cohort Action_ND = pshfise ND	1,040	,934	1,159
N of Valid Cases	5254		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΝΔ ένας άνεργος είναι 0,048 φορές μικρότερη από την σχετική πιθανότητα να ψηφίσει ΝΔ ένας εργαζόμενος.
- ✓ Επειδή το 95% δ.ε περιλαμβάνει την τιμή 1, δε μπορούμε να απορρίψουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στη ΝΔ δεν είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.27: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Occupation (ergazomenos / suntaxiouxos)	2,494	2,168	2,870
For cohort Action_ND = den pshfise ND	1,309	1,250	1,370
For cohort Action_ND = pshfise ND	,525	,476	,578
N of Valid Cases	4213		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΝΔ ένας συνταξιούχος είναι 1,494 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΝΔ ένας εργαζόμενος.
- ✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στη ΝΔ είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.28: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Occupation (foithths / anergos)	1,263	1,059	1,507
For cohort Action_ND = den pshfise ND	1,043	1,011	1,076
For cohort Action_ND = pshfise ND	,826	,714	,955
N of Valid Cases	3787		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΝΔ ένας άνεργος είναι 0,263 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΝΔ ένας φοιτητής.
- ✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στη ΝΔ είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.29: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Occupation (foithths / suntaxiouxos)	3,311	2,764	3,965
For cohort Action_ND = den pshfise ND	1,379	1,314	1,447
For cohort Action_ND = pshfise ND	,416	,363	,478
N of Valid Cases	2746		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΝΔ ένας συνταξιούχος είναι 2,311 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΝΔ ένας φοιτητής.
- ✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στη ΝΔ είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.30: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Occupation (anergos / suntaxiouxos)	2,621	2,267	3,030
For cohort Action_ND = den pshfise ND	1,322	1,262	1,384
For cohort Action_ND = pshfise ND	,504	,455	,558
N of Valid Cases	3897		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΝΔ ένας συνταξιούχος είναι 1,621 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΝΔ ένας άνεργος.
- ✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στη ΝΔ είναι διαφορετικά στις δύο ομάδες.

Όσον αφορά το σχετικό λόγο πιθανοτήτων για το ζεύγος Action_ND – Previous ενδεικτικά αναφέρουμε ότι:

- ✓ Η σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που στις προηγούμενες εκλογές είχε ψηφίσει ΣΥ.ΡΙΖ.Α είναι 0,89 φορές μικρότερη από την σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που στις προηγούμενες εκλογές είχε ψηφίσει ΝΔ (Πίνακας 4.31).
- ✓ Η σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που στις προηγούμενες εκλογές είχε ψηφίσει ΠΑΣΟΚ είναι 0,945 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που στις προηγούμενες εκλογές είχε ψηφίσει ΣΥ.ΡΙΖ.Α (Πίνακας 4.32).
- ✓ Η σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που στις προηγούμενες εκλογές είχε ψηφίσει ΧΑ είναι 0,338 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΝΔ ένας ψηφοφόρος που στις προηγούμενες εκλογές είχε ψηφίσει ΠΑΣΟΚ (Πίνακας 4.33).

Πίνακας 4.31: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Previous (ND / SYRIZA)	,011	,009	,014
For cohort Action_ND = den pshfise ND	,186	,166	,207
For cohort Action_ND = pshfise ND	16,225	13,568	19,402
N of Valid Cases	3788		

Πίνακας 4.32: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Previous (SYRIZA / PASOK)	1,945	1,354	2,794
For cohort Action_ND = den pshfise ND	1,048	1,016	1,081
For cohort Action_ND = pshfise ND	,539	,386	,751
N of Valid Cases	2752		

Πίνακας 4.33: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Previous (PASOK / XA)	1,338	,923	1,942
For cohort Action_ND = den pshfise ND	1,032	,993	1,072
For cohort Action_ND = pshfise ND	,771	,552	1,077
N of Valid Cases	1318		

4.4.2 Η περίπτωση του ΣΥ.ΡΙΖ.Α

Ακολουθώντας την ίδια διαδικασία και στην περίπτωση του ΣΥ.ΡΙΖ.Α, παίρνουμε τους Πίνακες 4.34 – 4.48 από τους οποίους προκύπτουν τα ακόλουθα συμπεράσματα:

Πίνακας 4.34: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Gender (antras / gunaika)	1,196	1,084	1,320
For cohort Action_SYRIZA = den pshfise SYRIZA	1,050	1,022	1,079
For cohort Action_SYRIZA = pshfise SYRIZA	,878	,817	,943
N of Valid Cases	8000		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α μία γυναίκα ψηφοφόρος είναι 0,196 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας άντρας ψηφοφόρος.
- ✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στον ΣΥ.ΡΙΖ.Α είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.35: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Age (18-24 / 25-34)	1,083	,939	1,249
For cohort Action_SYRIZA = den pshfise SYRIZA	1,021	,984	1,060
For cohort Action_SYRIZA = pshfise SYRIZA	,943	,848	1,047
N of Valid Cases	3921		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 25-34 είναι 0,083 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 18-24.
- ✓ Επειδή το 95% δ.ε περιλαμβάνει την τιμή 1, δε μπορούμε να απορρίψουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στον ΣΥ.ΡΙΖ.Α δεν είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.36: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Age (18-24 / 35-54)	1,316	1,143	1,514
For cohort Action_SYRIZA = den pshfise SYRIZA	1,080	1,038	1,124
For cohort Action_SYRIZA = pshfise SYRIZA	,821	,742	,908
N of Valid Cases	3889		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 35-54 είναι 0,316 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 18-24.
- ✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στον ΣΥ.ΡΙΖ.Α είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.37: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Age (18-24 / 55+)	1,044	,908	1,202
For cohort Action_SYRIZA = den pshfise SYRIZA	1,011	,975	1,049
For cohort Action_SYRIZA = pshfise SYRIZA	,968	,872	1,075
N of Valid Cases	4074		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 55+ είναι 0,044 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 18-24.
- ✓ Επειδή το 95% δ.ε περιλαμβάνει την τιμή 1, δε μπορούμε να απορρίψουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στον ΣΥ.ΡΙΖ.Α δεν είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.38: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Age (25-34 / 35-54)	1,214	1,058	1,394
For cohort Action_SYRIZA = den pshfise SYRIZA	1,058	1,016	1,101
For cohort Action_SYRIZA = pshfise SYRIZA	,871	,789	,961
N of Valid Cases	3926		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 35-54 είναι 0,214 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 25-34.
- ✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στον ΣΥ.ΡΙΖ.Α είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.39: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Age (25-34 / 55+)	,964	,839	1,107
For cohort Action_SYRIZA = den pshfise SYRIZA	,990	,955	1,027
For cohort Action_SYRIZA = pshfise SYRIZA	1,027	,928	1,137
N of Valid Cases	4111		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 55+ είναι 0,036 φορές μικρότερη από την σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 25-34.
- ✓ Επειδή το 95% δ.ε περιλαμβάνει την τιμή 1, δε μπορούμε να απορρίψουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στον ΣΥ.ΡΙΖ.Α δεν είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.40: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Age (35-54 / 55+)	,794	,693	,910
For cohort Action_SYRIZA = den pshfise SYRIZA	,936	,900	,974
For cohort Action_SYRIZA = pshfise SYRIZA	1,179	1,070	1,300
N of Valid Cases	4079		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 55+ είναι 0,206 φορές μικρότερη από την σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που ανήκει στην ηλικιακή ομάδα 35-54.

- ✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στον ΣΥ.ΡΙΖ.Α είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.41: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval Lower	95% Confidence Interval Upper
Odds Ratio for Occupation (ergazomenos / foithths)	,833	,717	,966
For cohort Action_SYRIZA = den pshfise SYRIZA	,952	,915	,990
For cohort Action_SYRIZA = pshfise SYRIZA	1,143	1,024	1,276
N of Valid Cases	4103		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας φοιτητής είναι 0,167 φορές μικρότερη από την σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας εργαζόμενος.
- ✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στον ΣΥ.ΡΙΖ.Α είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.42: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval Lower	95% Confidence Interval Upper
Odds Ratio for Occupation (ergazomenos / anergos)	1,097	,975	1,236
For cohort Action_SYRIZA = den pshfise SYRIZA	1,028	,992	1,065
For cohort Action_SYRIZA = pshfise SYRIZA	,937	,862	1,018
N of Valid Cases	5254		

- ✓ Η σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας άνεργος είναι 0,097 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας εργαζόμενος.

✓ Επειδή το 95% δ.ε περιλαμβάνει την τιμή 1, δε μπορούμε να απορρίψουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στον ΣΥ.ΡΙΖ.Α δεν είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.43: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Occupation (ergazomenos / suntaxiouxos)	,641	,551	,746
For cohort Action_SYRIZA = den pshfise SYRIZA	,897	,865	,929
For cohort Action_SYRIZA = pshfise SYRIZA	1,399	1,244	1,573
N of Valid Cases	4213		

✓ Η σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας συνταξιούχος είναι 0,359 φορές μικρότερη από την σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας εργαζόμενος.

✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στον ΣΥ.ΡΙΖ.Α είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.44: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Occupation (foithths / anergos)	1,318	1,134	1,533
For cohort Action_SYRIZA = den pshfise SYRIZA	1,080	1,037	1,125
For cohort Action_SYRIZA = pshfise SYRIZA	,819	,734	,915
N of Valid Cases	3787		

✓ Η σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας άνεργος είναι 0,318 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας φοιτητής.

✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στον ΣΥ.ΡΙΖ.Α είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.45: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Occupation (foithths / suntaxiouxos)	,770	,644	,921
For cohort Action_SYRIZA = den pshfise SYRIZA	,942	,904	,981
For cohort Action_SYRIZA = pshfise SYRIZA	1,223	1,066	1,404
N of Valid Cases	2746		

✓ Η σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας συνταξιούχος είναι 0,23 φορές μικρότερη από την σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας φοιτητής.

✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στον ΣΥ.ΡΙΖ.Α είναι διαφορετικά στις δύο ομάδες.

Πίνακας 4.46: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Occupation (anergos / suntaxiouxos)	,584	,501	,681
For cohort Action_SYRIZA = den pshfise SYRIZA	,872	,840	,905
For cohort Action_SYRIZA = pshfise SYRIZA	1,493	1,327	1,680
N of Valid Cases	3897		

✓ Η σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας συνταξιούχος είναι 0,416 φορές μικρότερη από την σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας άνεργος.

- ✓ Επειδή το 95% δ.ε δεν περιλαμβάνει την τιμή 1, απορρίπτουμε τη μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος ισούται με τη μονάδα και επομένως τα πραγματικά odds για την ψήφο στον ΣΥ.ΡΙΖ.Α είναι διαφορετικά στις δύο ομάδες.

Όσον αφορά το σχετικό λόγο πιθανοτήτων για το ζεύγος Action_SYRIZA – Previous ενδεικτικά αναφέρουμε ότι:

- ✓ Η σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που στις προηγούμενες εκλογές είχε ψηφίσει ΠΑΣΟΚ είναι 0,957 φορές μικρότερη από την σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που στις προηγούμενες εκλογές είχε ψηφίσει ΣΥ.ΡΙΖ.Α (Πίνακας 4.47).
- ✓ Η σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που στις προηγούμενες εκλογές είχε ψηφίσει ΚΚΕ είναι 0,998 φορές μεγαλύτερη από την σχετική πιθανότητα να ψηφίσει ΣΥ.ΡΙΖ.Α ένας ψηφοφόρος που στις προηγούμενες εκλογές είχε ψηφίσει ΠΑΣΟΚ (Πίνακας 4.48).

Πίνακας 4.47: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Previous (SYRIZA / PASOK)	,043	,031	,058
For cohort Action_SYRIZA = den pshfise SYRIZA	,307	,285	,330
For cohort Action_SYRIZA = pshfise SYRIZA	7,197	5,481	9,449
N of Valid Cases	2752		

Πίνακας 4.48: Αποτελέσματα του σχετικού λόγου πιθανοτήτων

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Previous (PASOK / KKE)	1,998	1,368	2,919
For cohort Action_SYRIZA = den pshfise SYRIZA	1,100	1,045	1,159
For cohort Action_SYRIZA = pshfise SYRIZA	,551	,396	,766
N of Valid Cases	954		

4.5 Το μοντέλο πρόβλεψης

Στόχος της παρούσας ενότητας είναι να περιγράψει τη διαδικασία κατασκευής ενός μοντέλου πολλαπλής λογιστικής παλινδρόμησης που μελετά την επίδραση μερικών χαρακτηριστικών των ψηφοφόρων στην ψήφο τους την ημέρα των εκλογών. Η διαδικασία κατασκευής του μοντέλου θα παρουσιαστεί αναλυτικά στην περίπτωση της ΝΔ και θα είναι ίδια για όλα τα υπόλοιπα κόμματα.

Ξεκινώντας από την περίπτωση της ΝΔ, ενδιαφερόμαστε να εκτιμήσουμε το ποσοστό που θα λάβει το κόμμα στις εκλογές καθώς επίσης και να διερευνήσουμε το κατά πόσο οι ερμηνευτικές μεταβλητές που αναφέρθηκαν προηγουμένως επιδρούν στην ψήφο του i -ψηφοφόρου.

Για να προσαρμόσουμε το μοντέλο της πολλαπλής λογιστικής παλινδρόμησης, επιλέγουμε

Analyze – Regression – Binary Logistic.

Η εξαρτημένη μεταβλητή είναι η μεταβλητή Action_ND ενώ οι ανεξάρτητες (κατηγορικές) μεταβλητές είναι οι μεταβλητές Age, Occupation, Gender και Previous. Η μέθοδος που χρησιμοποιήθηκε για την εισαγωγή των ανεξάρτητων μεταβλητών στον σχηματισμό της εξίσωσης είναι η μέθοδος της προοδευτικής κατά στάδια επιλογής μεταβλητών (forward stepwise method) και με κριτήριο απομάκρυνσης τη στατιστική συνάρτηση του Wald. Στο Πρώτο βήμα (Step 0) συμμετέχει στην εξίσωση μόνο ο σταθερός όρος.

Αρχικά, παίρνουμε τον Πίνακα 4.49 ο οποίος απεικονίζει τις ψευδομεταβλητές που δημιουργήθηκαν για την εισαγωγή των κατηγορικών μεταβλητών στο μοντέλο. Κατά τη διαδικασία δημιουργίας των ψευδομεταβλητών δημιουργούνται p-1 μεταβλητές, όπου p ο αριθμός των κατηγοριών της αρχικής μεταβλητής. Βάσει αυτής της κωδικοποίησης μπορούμε να συμπεράνουμε την επίδραση, στην εξαρτημένη μεταβλητή, κάποιου συγκεκριμένου επιπέδου της κατηγορικής μεταβλητής σε σύγκριση με μία άλλη κατηγορία που έχουμε ορίσει ως επίπεδο αναφοράς. Στην περίπτωση της μεταβλητής Previous, η οποία έχει p=6 επίπεδα, το επίπεδο αναφοράς είναι το “άλλο” για το οποίο οι τιμές των p-1=5 ψευδομεταβλητών έχουν τιμή 0. Όμοια, στην περίπτωση της μεταβλητής Occupation το επίπεδο αναφοράς είναι το επίπεδο “συνταξιούχος”, στην περίπτωση της μεταβλητής Age το επίπεδο αναφοράς είναι το “55+” και τέλος στην περίπτωση της μεταβλητής Gender το επίπεδο αναφοράς είναι το επίπεδο “άντρας”.

Πίνακας 4.49: Απεικόνιση των ψευδομεταβλητών που δημιουργήθηκαν για την εισαγωγή των κατηγορικών μεταβλητών στο μοντέλο

Categorical Variables Codings

			Parameter coding				
		Frequency	(1)	(2)	(3)	(4)	(5)
Previous	ND	1503	1,000	,000	,000	,000	,000
	SYRIZA	2285	,000	1,000	,000	,000	,000
	PASOK	467	,000	,000	1,000	,000	,000
	KKE	487	,000	,000	,000	1,000	,000
	XA	851	,000	,000	,000	,000	1,000
	ALLO	2407	,000	,000	,000	,000	,000
Occupation	ergazomenos	2785	1,000	,000	,000		
	foithths	1318	,000	1,000	,000		
	anergos	2469	,000	,000	1,000		
	suntaxiouxos	1428	,000	,000	,000		
Age	18-24	1942	1,000	,000	,000		
	25-34	1979	,000	1,000	,000		
	35-54	1947	,000	,000	1,000		
	55+	2132	,000	,000	,000		
Gender	antras	3976	1,000				
	gunaika	4024	,000				

Στη συνέχεια, παίρνουμε τον πίνακα ορθής ταξινόμησης (Πίνακας 4.50). Είναι ένας πίνακας μέσω του οποίου μπορούμε να εκτιμήσουμε τις πιθανότητες σωστής πρόβλεψης των τιμών της εξαρτημένης μεταβλητής από την εξίσωση που δημιουργήθηκε με βάση το μοντέλο της λογιστικής παλινδρόμησης που διαμορφώσαμε. Στις γραμμές του πίνακα απεικονίζονται οι πραγματικές τιμές της εξαρτημένης μεταβλητής, βάσει του αρχείου δεδομένων. Στις στήλες απεικονίζονται οι τιμές της εξαρτημένης μεταβλητής όπως αυτές προβλέπονται βάσει της εξίσωσης.

Πίνακας 4.50: Πίνακας ορθής ταξινόμησης

Classification Table[a]

			Predicted		
			Action_ND		
	Observed		den pshfise ND	pshfise ND	Percentage Correct
Step 1	Action_ND	den pshfise ND	5892	265	95,7
		pshfise ND	605	1238	67,2
	Overall Percentage				89,1
Step 2	Action_ND	den pshfise ND	5892	265	95,7
		pshfise ND	605	1238	67,2
	Overall Percentage				89,1
Step 3	Action_ND	den pshfise ND	5892	265	95,7
		pshfise ND	605	1238	67,2
	Overall Percentage				89,1
Step 4	Action_ND	den pshfise ND	5892	265	95,7
		pshfise ND	605	1238	67,2
	Overall Percentage				89,1

a. The cut value is ,500

Σύμφωνα με τον πίνακα ορθής ταξινόμησης διαπιστώνουμε ότι 5892 ψηφοφόροι που "δεν ψήφισαν ΝΔ" (95,7%) και 1238 ψηφοφόροι που "ψήφισαν ΝΔ" (67,2%) προβλέφθηκαν σωστά από το υπόδειγμα. Το γεγονός ότι μόνο 870 ψηφοφόροι (10,9%) ταξινομήθηκαν λανθασμένα, οδηγεί στο συμπέρασμα ότι το μοντέλο που διαμορφώσαμε είναι πολύ καλά προσαρμοσμένο στα δεδομένα.

Ο Πίνακας 4.51 περιγράφει το μοντέλο της λογιστικής παλινδρόμησης και μας δίνει πληροφορίες για τις μεταβλητές που συνδέονται στατιστικά σημαντικά με την εξαρτημένη μεταβλητή. Πιο αναλυτικά, παρουσιάζονται οι συντελεστές του μοντέλου, οι τιμές της στατιστικής συνάρτησης Wald, τα odds ratios (Exp(B)) καθώς και τα αντίστοιχα 95% δ.ε για τα odds ratio. Στο τέταρτο βήμα, όπου και σταματούν οι επαναλήψεις, παρατηρούμε ότι η μέθοδος θεωρεί στατιστικά σημαντικές και εισάγει στο μοντέλο όλες τις κατηγορικές μεταβλητές και επομένως το μοντέλο λογιστικής παλινδρόμησης για να μπορεί να προβλέψει με ακρίβεια τα τελικά ποσοστά των εκλογών απαιτεί τη γνώση των τιμών όλων των ανεξάρτητων μεταβλητών.

Πίνακας 4.51: Αποτελέσματα για τους συντελεστές του μοντέλου

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Previous			2258,754	5	,000			
	Previous(1)	3,406	,090	1422,095	1	,000	30,142	25,252	35,979
	Previous(2)	-1,064	,113	89,447	1	,000	,345	,277	,430
	Previous(3)	-,399	,169	5,548	1	,019	,671	,482	,935
	Previous(4)	-1,396	,248	31,803	1	,000	,248	,152	,402
	Previous(5)	-,107	,121	,792	1	,373	,898	,709	1,138
	Constant	-1,864	,060	972,072	1	,000	,155		
Step 2[b]	Occupation			183,229	3	,000			
	Occupation(1)	-1,155	,100	132,127	1	,000	,315	,259	,384
	Occupation(2)	-1,408	,126	124,255	1	,000	,245	,191	,313
	Occupation(3)	-1,049	,102	105,754	1	,000	,350	,287	,428
	Previous			2169,547	5	,000			
	Previous(1)	3,396	,092	1353,882	1	,000	29,836	24,899	35,752
	Previous(2)	-1,158	,114	102,820	1	,000	,314	,251	,393
	Previous(3)	-,576	,173	11,132	1	,001	,562	,401	,789
	Previous(4)	-1,468	,250	34,616	1	,000	,230	,141	,376
	Previous(5)	-,152	,123	1,535	1	,215	,859	,676	1,092
	Constant	-,915	,090	103,032	1	,000	,400		

Step 3[c]	Age			39,435	3	,000			
	Age(1)	-,568	,120	22,339	1	,000	,567	,448	,717
	Age(2)	-,410	,117	12,309	1	,000	,663	,527	,834
	Age(3)	-,566	,101	31,428	1	,000	,568	,466	,692
	Occupation			94,833	3	,000			
	Occupation(1)	-1,015	,110	84,563	1	,000	,362	,292	,450
	Occupation(2)	-1,152	,157	53,850	1	,000	,316	,232	,430
	Occupation(3)	-,905	,112	64,688	1	,000	,405	,325	,505
	Previous			2142,656	5	,000			
	Previous(1)	3,382	,093	1330,520	1	,000	29,433	24,542	35,299
	Previous(2)	-1,174	,115	104,982	1	,000	,309	,247	,387
	Previous(3)	-,625	,174	12,978	1	,000	,535	,381	,752
	Previous(4)	-1,522	,250	37,008	1	,000	,218	,134	,356
	Previous(5)	-,141	,123	1,308	1	,253	,869	,683	1,106
	Constant	-,674	,099	46,390	1	,000	,510		
Step 4[d]	Age			39,623	3	,000			
	Age(1)	-,572	,120	22,656	1	,000	,564	,446	,714
	Age(2)	-,414	,117	12,524	1	,000	,661	,525	,831
	Age(3)	-,565	,101	31,389	1	,000	,568	,466	,692
	Occupation			93,597	3	,000			
	Occupation(1)	-1,010	,110	83,552	1	,000	,364	,293	,452
	Occupation(2)	-1,149	,157	53,555	1	,000	,317	,233	,431
	Occupation(3)	-,896	,113	63,376	1	,000	,408	,327	,509
	Gender(1)	-,147	,075	3,873	1	,049	,864	,746	,999
	Previous			2140,667	5	,000			
	Previous(1)	3,390	,093	1330,848	1	,000	29,661	24,722	35,586
	Previous(2)	-1,177	,115	105,517	1	,000	,308	,246	,386
	Previous(3)	-,628	,174	13,078	1	,000	,534	,380	,750
	Previous(4)	-1,518	,250	36,807	1	,000	,219	,134	,358
	Previous(5)	-,121	,123	,959	1	,327	,886	,696	1,129
	Constant	-,608	,104	33,984	1	,000	,544		

a. Variable(s) entered on step 1: Previous.

b. Variable(s) entered on step 2: Occupation.

c. Variable(s) entered on step 3: Age.

d. Variable(s) entered on step 4: Gender.

Ενδεικτικά, τα συμπεράσματα που προκύπτουν για τα odds ratio σύμφωνα με τον Πίνακα 4.51 είναι τα εξής:

- Το log odds για να ξαναψηφίσει ΝΔ ένας ψηφοφόρος που είχε ψηφίσει στις προηγούμενες εκλογές ΝΔ είναι 29,661 φορές περισσότερο σε σχέση με το αντίστοιχο log odds ενός ψηφοφόρου που είχε ψηφίσει κάποιο άλλο κόμμα. Το αντίστοιχο 95% δ.ε για το odds ratio είναι το (24,722 , 35,586).
- Το log odds για να ψηφίσει ΝΔ ένας ψηφοφόρος που έχει ηλικία μεταξύ 18 και 24 ετών είναι 0,564 φορές περισσότερο σε σχέση με το αντίστοιχο log odds ενός ψηφοφόρου μεγαλύτερου των 55 ετών. Το αντίστοιχο 95% δ.ε για το odds ratio είναι το (0,446 , 0,714).

- Το log odds για να ψηφίσει ΝΔ ένας ψηφοφόρος που είναι εργαζόμενος είναι 0,364 φορές περισσότερο σε σχέση με το αντίστοιχο log odds ενός συνταξιούχου. Το αντίστοιχο 95% δ.ε για το odds ratio είναι το (0,293 , 0,452).
- Το log odds για να ψηφίσει ΝΔ ένας άντρας ψηφοφόρος είναι 0,864 φορές περισσότερο σε σχέση με το αντίστοιχο log odds μίας γυναίκας ψηφοφόρο. Το αντίστοιχο 95% δ.ε για το odds ratio είναι το (0,746 , 0,999).

Τέλος, το λογιστικό μοντέλο πρόβλεψης που προκύπτει, αντικαθιστώντας σε κάθε περίπτωση την κάθε μεταβλητή με βάση την κωδικοποίηση που έχει γίνει στον Πίνακα 4.49, θα είναι της μορφής

$$\ln\frac{p}{1-p} = -0.608-0.572Age(1)-0.414Age(2)-0.565Age(3)-0.010Occupation(1)-1.149Occupation(2)-0.896Occupation(3)-0.147Gender(1)+3.390Previous(1)-1.177Previous(2)-0.628Previous(3)-1.518Previous(4).$$

ενώ, η εκτίμηση του ποσοστού που θα λάβει το κόμμα θα υπολογίζεται από τον τύπο

$$p=\frac{e^{-0.608-0.572Age(1)-0.414Age(2)-\ldots-1.518\,Previous(4)}}{1+e^{-0.608-0.572Age(1)-0.414Age(2)-\ldots-1.518\,Previous(4)}}\,.$$

Προκειμένου να γίνει κατανοητό το πώς λειτουργεί το μοντέλο, θα εκτιμήσουμε στη συνέχεια την πιθανότητα να ψηφίσει ΝΔ ένας άντρας ψηφοφόρος ο οποίος ανήκει στην ηλικιακή ομάδα 35-54, είναι άνεργος και στις προηγούμενες εκλογές είχε ψηφίσει ΝΔ. Για την περίπτωση αυτή οι τιμές των ανεξάρτητων μεταβλητών, σύμφωνα με την κωδικοποίηση του Πίνακα 4.49 θα είναι 0,0,1,0,0,1,1,1,0,0,0,0 οπότε

$$p=\frac{e^{-0.608-0.565Age(3)-0.896Occupation(3)-0.147Gender(1)+3.390\,Previous(1)}}{1+e^{-0.608-0.565Age(3)-0.896Occupation(3)-0.147Gender(1)+3.390\,Previous(1)}}=\frac{e^{1.174}}{1+e^{1.174}}=0.76=76\%\,.$$

Αντίστοιχα, οι τιμές των ανεξάρτητων μεταβλητών στην περίπτωση που θέλουμε να εκτιμήσουμε την πιθανότητα να ψηφίσει ΝΔ ένας άντρας ψηφοφόρος ο οποίος ανήκει στην ηλικιακή ομάδα 55+, είναι εργαζόμενος και στις προηγούμενες εκλογές είχε ψηφίσει ΚΚΕ θα είναι 0,0,0,1,0,0,1,0,0,0,1,0 και επομένως

$$p = \frac{e^{-0.608-1.010 Occupation(1)-0.147 Gender(1)-1.518 Previous(4)}}{1+e^{-0.608-1.010 Occupation(1)-0.147 Gender(1)-1.518 Previous(4)}} = \frac{e^{-3.283}}{1+e^{-3.283}} = 0.037 = 3.7\% .$$

Ακολουθώντας την ίδια διαδικασία και για τα υπόλοιπα κόμματα καταλήγουμε στα εξής συμπεράσματα:

- Στο λογιστικό μοντέλο με εξαρτημένη μεταβλητή τη μεταβλητή Action_SYRIZA η κατηγορική μεταβλητή Gender δεν προκύπτει στατιστικά σημαντική (*p*-value = 0,483) και δεν εισάγεται στο μοντέλο. Εάν τρέχουμε ξανά την παλινδρόμηση (Πίνακας 4.52), αυτή τη φορά χωρίς να λάβουμε υπόψη την κατηγορική μεταβλητή Gender, το τελικό μοντέλο πρόβλεψης θα είναι της μορφής:

$$\ln\frac{p}{1-p} = \text{-}2.449\text{-}0.268 Age(1)\text{-}{+}0{,}706 Occupation(1){+}0{,}667 Occupation(2){+}0{,}802 Occupation(3)\text{-}1.725 Previous(1){+}2.903 Previous(2){+}0.430 Previous(4)\text{-}0.960 Previous(5).$$

Πίνακας 4.52: Αποτελέσματα για τους συντελεστές του μοντέλου

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Age			11,697	3	,008	
	Age(1)	-,268	,106	6,400	1	,011	,765
	Age(2)	-,184	,105	3,087	1	,079	,832
	Age(3)	,065	,093	,488	1	,485	1,067
	Occupation			56,447	3	,000	
	Occupation(1)	,706	,109	42,120	1	,000	2,026
	Occupation(2)	,667	,143	21,748	1	,000	1,949
	Occupation(3)	,802	,110	52,743	1	,000	2,231
	Previous			2281,454	5	,000	
	Previous(1)	-1,725	,178	94,198	1	,000	,178
	Previous(2)	2,903	,078	1395,363	1	,000	18,233
	Previous(3)	-,259	,166	2,419	1	,120	,772
	Previous(4)	,430	,133	10,457	1	,001	1,537
	Previous(5)	-,960	,164	34,417	1	,000	,383
	Constant	-2,449	,113	469,051	1	,000	,086

a. Variable(s) entered on step 1: Age, Occupation, Previous.

- Στο λογιστικό μοντέλο με εξαρτημένη μεταβλητή τη μεταβλητή Action_ELIA η κατηγορική μεταβλητή Gender δεν προκύπτει στατιστικά σημαντική (p-value = 0,630) και δεν εισάγεται στο μοντέλο. Το τελικό μοντέλο πρόβλεψης (Πίνακας 4.53) θα είναι της μορφής:

$$\ln\frac{p}{1-p}=\text{-}1.449\text{-}0.706\,Age(2)\text{-}0.604\,Age(3)\text{-}0.802\,Occupation(1)\text{-}1.617\,Occupation(2)$$

$$\text{-}1.039\,Occupation(3)\text{-}1.166\,Previous(1)+3.338\,Previous(3)\text{-}1.248\,Previous(4)\text{-}1.144\,Previous(5).$$

Πίνακας 4.53: Αποτελέσματα για τους συντελεστές του μοντέλου

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Age			38,224	3	,000	
	Age(1)	-,242	,141	2,950	1	,086	,785
	Age(2)	-,706	,152	21,488	1	,000	,494
	Age(3)	-,604	,118	26,096	1	,000	,547
	Occupation			80,341	3	,000	
	Occupation(1)	-,802	,126	40,495	1	,000	,448
	Occupation(2)	-1,617	,214	57,102	1	,000	,199
	Occupation(3)	-1,039	,134	60,499	1	,000	,354
	Previous			979,914	5	,000	
	Previous(1)	-1,166	,177	43,549	1	,000	,312
	Previous(2)	,191	,112	2,890	1	,089	1,210
	Previous(3)	3,338	,132	635,079	1	,000	28,162
	Previous(4)	-1,248	,318	15,404	1	,000	,287
	Previous(5)	-1,144	,242	22,313	1	,000	,319
	Constant	-1,449	,118	151,459	1	,000	,235

a. Variable(s) entered on step 1: Age, Occupation, Previous.

- Στο λογιστικό μοντέλο με εξαρτημένη μεταβλητή τη μεταβλητή Action_KKE η κατηγορικές μεταβλητές Gender και Occupation δεν προκύπτουν στατιστικά σημαντικές και επομένως το τελικό μοντέλο πρόβλεψης (Πίνακας 4.54) θα είναι της μορφής:

$$\ln\frac{p}{1-p}=\text{-}3.563\text{-}0.562\,Age(3)\text{-}0.562\,Age(3)\text{-}1.754\,Previous(1)+4.514\,Previous(4).$$

Πίνακας 4.54: Αποτελέσματα για τους συντελεστές του μοντέλου

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Age			8,848	3	,031	
	Age(1)	-,189	,178	1,138	1	,286	,827
	Age(2)	-,298	,182	2,661	1	,103	,743
	Age(3)	-,562	,193	8,499	1	,004	,570
	Previous			1220,555	5	,000	
	Previous(1)	-1,754	,433	16,393	1	,000	,173
	Previous(2)	,071	,198	,128	1	,721	1,073
	Previous(3)	-17,428	1855,352	,000	1	,993	,000
	Previous(4)	4,514	,171	700,006	1	,000	91,260
	Previous(5)	-,268	,304	,777	1	,378	,765
	Constant	-3,563	,177	406,743	1	,000	,028

a. Variable(s) entered on step 1: Age, Previous.

- Στο λογιστικό μοντέλο με εξαρτημένη μεταβλητή την Action_XA όλες οι μεταβλητές προκύπτουν στατιστικά σημαντικές (Πίνακας 4.55) και εισάγονται στο μοντέλο το οποίο θα είναι της μορφής:

$$\ln \frac{p}{1-p} = -4.439+0.648 Age(1)+0.402 Age(2)+0.412 Age(3)+0.397 Occupation(1)$$

$$+0.570 Occupation(3)+1.003 \text{Gender}(1)-1.357 Previous(2)-1.157 Previous(4)+3.581 Previous(5).$$

Πίνακας 4.55: Αποτελέσματα για τους συντελεστές του μοντέλου

Step 4[d]	Age			16,297	3	,001	
	Age(1)	,648	,164	15,528	1	,000	1,911
	Age(2)	,402	,165	5,928	1	,015	1,495
	Age(3)	,412	,151	7,445	1	,006	1,510
	Previous			1415,716	5	,000	
	Previous(1)	-,042	,167	,063	1	,802	,959
	Previous(2)	-1,357	,221	37,604	1	,000	,257
	Previous(3)	-,304	,282	1,158	1	,282	,738
	Previous(4)	-1,157	,395	8,603	1	,003	,314
	Previous(5)	3,581	,126	814,117	1	,000	35,906
	Occupation			15,023	3	,002	
	Occupation(1)	,397	,174	5,190	1	,023	1,488
	Occupation(2)	,167	,221	,572	1	,449	1,182
	Occupation(3)	,570	,176	10,466	1	,001	1,769
	Gender(1)	1,003	,107	88,002	1	,000	2,727
	Constant	-4,439	,203	479,505	1	,000	,012

a. Variable(s) entered on step 1: Previous.
b. Variable(s) entered on step 2: Gender.
c. Variable(s) entered on step 3: Age.
d. Variable(s) entered on step 4: Occupation.

- Στο λογιστικό μοντέλο με εξαρτημένη μεταβλητή την Action_ALLO όλες οι μεταβλητές προκύπτουν στατιστικά σημαντικές (Πίνακας 4.56) και εισάγονται στο μοντέλο το οποίο θα είναι της μορφής:

$$\ln\frac{p}{1-p} = -0.807+0.790\,Age(1)+0.892\,Age(2)+0.843\,Age(3)+0.771\,Occupation(1)+1.071\,Occupation(2)$$

$$+0.644\,Occupation(3)-0.235\,\mathrm{Gender}(1)-2.812\,Previous(1)-2.619\,Previous(2)-2.535\,Previous(3)$$

$$-3.043\,Previous(4)-2.051\,Previous(5).$$

Πίνακας 4.56: Αποτελέσματα για τους συντελεστές του μοντέλου

Step 4[d]	Age			97,503	3	,000	
	Age(1)	,790	,104	57,733	1	,000	2,204
	Age(2)	,892	,103	74,984	1	,000	2,440
	Age(3)	,843	,097	76,015	1	,000	2,322
	Previous			1703,435	5	,000	
	Previous(1)	-2,812	,106	698,394	1	,000	,060
	Previous(2)	-2,619	,082	1013,415	1	,000	,073
	Previous(3)	-2,535	,162	245,753	1	,000	,079
	Previous(4)	-3,043	,183	276,948	1	,000	,048
	Previous(5)	-2,051	,103	397,472	1	,000	,129
	Occupation			63,748	3	,000	
	Occupation(1)	,771	,113	46,340	1	,000	2,162
	Occupation(2)	1,071	,138	59,827	1	,000	2,917
	Occupation(3)	,644	,116	31,047	1	,000	1,905
	Gender(1)	-,235	,062	14,194	1	,000	,791
	Constant	-,807	,111	52,457	1	,000	,446

a. Variable(s) entered on step 1: Previous.

b. Variable(s) entered on step 2: Age.

c. Variable(s) entered on step 3: Occupation.

d. Variable(s) entered on step 4: Gender.

Τέλος, στους Πίνακες 4.57-4.61 παρουσιάζονται οι αντίστοιχοι πίνακες ορθής ταξινόμησης για τα παραπάνω μοντέλα πρόβλεψης σύμφωνα με τους οποίους τα μοντέλα που διαμορφώσαμε προσαρμόζονται πολύ καλά στα δεδομένα.

Πίνακας 4.57: Πίνακας ορθής ταξινόμησης

Classification Table[a]

Observed			Predicted: Action_SYRIZA: den pshfise SYRIZA	Predicted: Action_SYRIZA: pshfise SYRIZA	Percentage Correct
Step 1	Action_SYRIZA	den pshfise SYRIZA	5183	630	89,2
		pshfise SYRIZA	532	1655	75,7
	Overall Percentage				85,5

a. The cut value is ,500

Πίνακας 4.58: Πίνακας ορθής ταξινόμησης

Classification Table[a]

Observed			Predicted: Action_ELIA den pshfise ELIA	Predicted: Action_ELIA pshfise ELIA	Percentage Correct
Step 1	Action_ELIA	den pshfise ELIA	7118	134	98,2
		pshfise ELIA	439	309	41,3
	Overall Percentage				92,8

a. The cut value is ,500

Πίνακας 4.59: Πίνακας ορθής ταξινόμησης

Classification Table[a]

Observed			Predicted: Action_KKE den pshfise KKE	Predicted: Action_KKE pshfise KKE	Percentage Correct
Step 1	Action_KKE	den pshfise KKE	7388	160	97,9
		pshfise KKE	125	327	72,3
	Overall Percentage				96,4

a. The cut value is ,500

Πίνακας 4.60: Πίνακας ορθής ταξινόμησης

Classification Table[a]

Observed			Predicted: Action_XA den pshfise XA	Predicted: Action_XA pshfise XA	Percentage Correct
Step 1	Action_XA	den pshfise XA	7060	202	97,2
		pshfise XA	316	422	57,2
	Overall Percentage				93,5

a. The cut value is ,500

Πίνακας 4.61: Πίνακας ορθής ταξινόμησης

Classification Table[a]

Observed			Predicted: Action_ALLO: den pshfise ALLO	Predicted: Action_ALLO: pshfise ALLO	Predicted: Percentage Correct
Step 1	Action_ALLO	den pshfise ALLO	5357	611	89,8
		pshfise ALLO	808	1224	60,2
	Overall Percentage				82,3

a. The cut value is ,500

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

Νικολακόπουλος Ηλίας (1989), *Εισαγωγή στη Θεωρία και την Πρακτική των εκλογικών Συστημάτων*, Εκδόσεις Σάκκουλα, Αθήνα.

Παντελής Μ. Αντώνης (2007). *Εγχειρίδιο Συνταγματικού Δικαίου*, Εκδόσεις Λιβάνη, Αθήνα, 121-133.

Ράικος Αθανάσιος Γ. (2002). *Συνταγματικό Δίκαιο Εισαγωγή Οργανωτικό Μέρος*, Εκδόσεις Σάκκουλα, Αθήνα, 212-213, 306-311, 356-370, 377-415.

Δημητρόπουλος Ανδρέας (2004). *Γενική Συνταγματική θεωρία*, Αθήνα.

Σπυρόπουλος Φίλιππος (2006). *Εισαγωγή στο Συνταγματικό Δίκαιο*, Εκδόσεις Σάκκουλα, Αθήνα, 186-191.

Διαμαντόπουλος Θανάσης (2008). *Εκλογικά συστήματα θεωρία και πρακτικές εφαρμογές*, Εκδόσεις Πατάκη, Αθήνα.

Δημητρόπουλος Ανδρέας Γ. (2009). *Οργάνωση και Λειτουργία του Κράτους σύστημα Συνταγματικού Δικαίου Τόμος Β'*, Εκδόσεις Σάκκουλα, Αθήνα-Θεσσαλονίκη, 154-176.

Μπούτσικας Μιχ. (2004). Εκπαιδευτικές σημειώσεις του μαθήματος "*Μέθοδοι Προσομοίωσης και Στατιστικές Υπολογιστικές Τεχνικές"*.

Ηλιόπουλος Γ. (2011). Εκπαιδευτικές σημειώσεις του μαθήματος "*Ανάλυση Διακριτών Δεδομένων*".

Πολίτης Κ. (2011). Εκπαιδευτικές σημειώσεις του μαθήματος "Γενικευμένα Γραμμικά Μοντέλα".

Σαχλάς Αθ. (2013-2014). Εργαστηριακές σημειώσεις του μαθήματος "*Ανάλυση Διακριτών Δεδομένων*".

www.hellenicparliament.gr

http://www.mavris.gr/1798/tomos2007-meros-2-3/

http://www.mavris.gr/3300/greek-elections-2012-and-polls-2/

http://www.metronanalysis.gr/exit-poll-euroelections-may-2014-2/

Ξένη

Jeff Gill and Jason Gainous (2002). *Why Does Voting Get So Complicated? A Review of Theories for Analyzing Democratic Participation*, 386-393.

Andrew Gelman, Jonathan N. Katz and Francis Tuerlinckx (2002). *The Mathematics and Statistics of Voting Power*

Sproul, Robyn (22 October 2008). *EXPLAINER: How Exit Polls Work*, ABCNews.com.

www.Wikipedia.org/wiki/Voting_systems

http://ukpollingreport.co.uk/guide/seat-profiles/haltempriceandhowden/

http://www.electoralcommission.org.uk/data/assets/pdf_file/0006/83373/Part-E-Final-Proofs-ok-web.pdf

http://www.legislation.gov.uk/ukpga/2011/1/pdfs/ukpgaen_20110001_en.pdf
http://www.skepdic.gr/entries/ni/lawtryllynum.htm