ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Σχολή Χρηματοοικονομικής και Στατιστικής

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ **ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

ΣΥΓΚΡΙΣΕΙΣ ΕΝΑΛΛΑΚΤΙΚΩΝ ΜΕΘΟΔΩΝ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΣΕ ΟΙΚΟΝΟΜΕΤΡΙΚΗ ΑΝΑΛΥΣΗ ΜΙΚΡΟΔΕΔΟΜΕΝΩΝ: ΓΡΑΜΜΙΚΗ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗ ΠΕΜΠΤΗΜΟΡΙΩΝ

Κωνσταντίνα Α. Καλογερά

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
Νοέμβριος 2023

Π Α Ν Ε Π Ι Σ Τ Η Μ Ι Ο Π Ε Ι Ρ Α Ι Ω Σ

Σχολή Χρηματοοικονομικής και Στατιστικής

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ **ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

ΣΥΓΚΡΙΣΕΙΣ ΕΝΑΛΛΑΚΤΙΚΩΝ ΜΕΘΟΔΩΝ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΣΕ ΟΙΚΟΝΟΜΕΤΡΙΚΗ ΑΝΑΛΥΣΗ ΜΙΚΡΟΔΕΔΟΜΕΝΩΝ: ΓΡΑΜΜΙΚΗ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗ ΠΕΜΠΤΗΜΟΡΙΩΝ

Κωνσταντίνα Α. Καλογερά

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
Νοέμβριος 2023

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:
- Τήνιος Πλάτων (Αναπληρωτής Καθηγητής) (επιβλέπων)
- Βερροπούλου Γεωργία (Καθηγήτρια)
- Ξένος Παναγιώτης (Επίκουρος Καθηγητής)

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

School of Finance and Statistics

Department of Statistics and Insurance Science

POSTGRADUATE PROGRAM IN
APPLIED STATISTICS

COMPARING ALTERNATIVE REGRESSION TECHNIQUES IN THE ECONOMETRIC ANALYSIS OF MICRO DATA: LINEAR AND QUANTILE REGRESSION USING SHARE DATA

By Konstantina A. Kalogera

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
November 2023

Ευχαριστίες

Αρχικά, θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στον επιβλέποντα καθηγητή μου, κ^{ο} Πλάτων Τήνιο, για την άριστη συνεργασία του. Επίσης, θα ήθελα να ευχαριστήσω τον κ^{ο} Μιχάλη Χουζούρη για τις πολύτιμες συμβουλές του και την άμεση ανταπόκρισή του στις απορίες μου, κατά τη διάρκεια εκπόνησης της παρούσας εργασίας.

Ακόμα, θα ήθελα να ευχαριστήσω την μητέρα μου που είναι πάντα δίπλα μου, στηρίζοντας όλες μου τις αποφάσεις, και τον πατέρα μου που πιστεύει σε εμένα. Τέλος, σημαντική ήταν η συμβολή των φίλων μου, των οποίων η υποστήριξη ήταν καθοριστική για την ολοκλήρωση της εργασίας.

Περίληψη

Η Πολλαπλή Γραμμική Παλινδρόμηση αποτελεί την πλέον γνωστή μέθοδο διερεύνησης της μέσης συμπεριφοράς μιας μεταβλητής, δεδομένου ενός συνόλου μεταβλητών που την επηρεάζουν. Απεναντίας, η Παλινδρόμηση Πεμπτημορίων μελετά την συμπεριφορά της υπό εξέτασης μεταβλητής σε όλο το εύρος της κατανομής της. Στην παρούσα εργασία παρουσιάζονται τα μοντέλα, οι ιδιαιτερότητες τους και οι διαφορές τους. Στη συνέχεια, χρησιμοποιούνται τα δεδομένα του 8^{ου} κύματος της έρευνας SHARE 26 προκειμένου να διερευνηθούν ποιοι παράγοντες επηρεάζουν το ετήσιο ύψος της σύνταξης. Γίνεται χρήση του πακέτου SPSS, όπου πραγματοποιείται στατιστική ανάλυση των δεδομένων και διεξάγονται οι παλινδρομήσεις. Απώτερος σκοπός της παρούσας εργασίας είναι η σύγκριση των δύο μεθόδων παλινδρόμησης προκειμένου να διαπιστωθεί ποια από τις δύο έχει καλύτερη ερμηνευτική ικανότητα στο ύψος της σύνταξης. Από την ανάλυση προέκυψε ότι δεν ικανοποιούνται οι προϋποθέσεις της Γραμμικής Παλινδρόμησης ώστε να είναι αξιόπιστα τα συμπεράσματα, ενώ η Παλινδρόμηση Πεμπτημορίων έχει την δυνατότητα να διαχωρίζει τα χαμηλά από τα υψηλά επίπεδα του ύψους της σύνταξης.

Λέξεις κλειδιά: Γραμμική Παλινδρόμηση, Παλινδρόμηση Πεμπτημορίων, SHARE, Συντάξεις

Abstract

Multiple Linear Regression is the most common method of detecting the conditional mean behavior of a variable, when is affected by a set of different variables. On the contrary, Quantile Regression studies the behavior of the response variable over the entire range of its distribution. The current study compares the models, their particularities and their differences. Furthermore, data from the 8th wave of SHARE survey were used to detecting which factors affect the annual pension amount. The statistical analysis of the data was realized using SPSS 26, using regression models. The purpose of the study was to compare the two regression models to determine which had better interpretive capacity at the height of pensions. From the analysis, Linear Regression assumptions were not met for the conclusions to be reliable, while Quantile Regression could possibly separate the low from the high levels of the pension amount.

Key words: Linear Regression, Quantile Regression, SHARE, Pensions

Περιεχόμενα

Κατάλογος Πινάκων

Κατάλογος Σχημάτων

Κατάλογος Εικόνων

Κατάλογος Συντομογραφιών

SSE	Sum of Squared Errors
SST	Total Sum of Squares
SSR	Regression Sum of Squares
BLUE	Best Linear Unbiased Estimator
SHARE	Survey of Health, Ageing and Retirement in Europe
HRS	Health Retirement Study
ELSA	English Longitudinal Survey on Ageing
CATI	Computer Assisted Telephone Interviewing
CAPI	Computer Assisted Personal Interviewing
ISCED 1997	International Standard Classification of Education
VIF	Variance Inflation Factor
MAE	Mean Absolute Error

Εισαγωγή

Αντικείμενο της παρούσας εργασίας είναι η σύγκριση δύο εναλλακτικών μεθόδων παλινδρόμησης, της Γραμμικής και της Παλινδρόμησης Πεμπτημορίων. Η Γραμμική είναι μία ευρέως χρησιμοποιούμενη τεχνική για την εξέταση των σχέσεων μεταξύ των μεταβλητών, αλλά δεν είναι κατάλληλη για όλα τα δεδομένα. Έτσι εισάγεται η Παλινδρόμηση Πεμπτημορίων ως εναλλακτική μέθοδος που αντιμετωπίζει τους περιορισμούς της πρώτης μεθόδου.

Προκειμένου να βγουν συμπεράσματα για την αξιοπιστία των δύο μεθόδων, δημιουργείται ένα μοντέλο που περιγράφει τις μεταβολές στο ετήσιο ύψος της σύνταξης που προκαλούνται από κοινωνικοοικονομικούς παράγοντες.

Όσον αφορά τη δομή της εργασίας αποτελείται από τρία κεφάλαια:

- Στο πρώτο περιγράφεται η πρώτη μέθοδος παλινδρόμησης, η Γραμμική. Γίνεται ανάλυση του μοντέλου και των υποθέσεων που πρέπει να ικανοποιούνται. Στην συνέχεια παρουσιάζεται μία μέθοδος εκτίμησης του υποδείγματος και οι ιδιότητες των συγκεκριμένων εκτιμητών. Έπειτα, δίνονται κάποιοι δείκτες, ώστε να μετρηθεί η ερμηνευτική ικανότητα του υποδείγματος, και κάποιοι έλεγχοι για την στατιστική σημαντικότητα των παραμέτρων, αλλά και ολόκληρου του μοντέλου. Παρουσιάζεται ο τρόπος εισαγωγής των κατηγορικών μεταβλητών στο υπόδειγμα και τέλος, τα διαστήματα πρόβλεψης της εξαρτημένης μεταβλητής.
- Στο δεύτερο κεφάλαιο παρουσιάζεται το θεωρητικό πλαίσιο της δεύτερης μεθόδου, της Παλινδρόμησης Πεμπτημορίων. Σε αυτό το κεφάλαιο γίνεται ανάλυση του συγκεκριμένου μοντέλου, παρουσιάζεται η μέθοδος εκτίμησης των συντελεστών και ο συντελεστής ερμηνευτικής ικανότητας. Ακόμα, γίνεται η σύγκριση των δύο μεθόδων παλινδρόμησης.
- Στο τρίτο κεφάλαιο πραγματοποιείται ανάλυση των δεδομένων της έρευνας SHARE και παρουσιάζονται τα αποτελέσματα της στατιστικής επεξεργασίας των δεδομένων. Πιο συγκεκριμένα, γίνεται μια εισαγωγή στα στοιχεία SHARE και στα δεδομένα, τα οποία θα χρησιμοποιηθούν στην ερευνητική ανάλυση. Ύστερα, εφαρμόζεται η Γραμμική Παλινδρόμηση και δημιουργείται ένα μοντέλο πρόβλεψης του ύψους των συντάξεων. Στη συνέχεια, εφαρμόζεται η Παλινδρόμηση Πεμπτημορίων στο υπόδειγμα που έχει δημιουργηθεί και γίνεται σύγκριση των δύο μεθόδων.

Τέλος, παρουσιάζονται τα τελικά συμπεράσματα που προέκυψαν από το θεωρητικό πλαίσιο των δύο πρώτων κεφαλαίων, αλλά και τα αποτελέσματα του ερευνητικού μέρους του τρίτου κεφαλαίου.

ΚΕΦΑΛΑΙΟ 1

Πολλαπλή Γραμμική Παλινδρόμηση

1.1 Εισαγωγή

Η παλινδρόμηση είναι η μέθοδος η οποία εξετάζει τη σχέση μεταξύ δύο ή περισσότερων τυχαίων μεταβλητών. Η ανάλυση της παλινδρόμησης επιδιώκει την εύρεση μιας μαθηματικής σχέσης, η οποία εκφράζει την επίδραση μιας μεταβλητής (ή πολλών μεταβλητών) πάνω σε μία άλλη μεταβλητή.

Επομένως, υπάρχει μία μεταβλητή, η οποία χρήζει μελέτης και ονομάζεται εξαρτημένη, και μερικές άλλες μεταβλητές, οι οποίες ενδεχομένως να ερμηνεύουν την πρώτη μεταβλητή και λέγονται ανεξάρτητες μεταβλητές. Στην περίπτωση όπου οι μεταβλητές συνδέονται γραμμικά τότε αναφερόμαστε στην Γραμμική Παλινδρόμηση.

Η συνήθης οικονομετρική προσέγγιση είναι η εκτίμηση των αιτιωδών σχέσεων. Δηλαδή, προηγείται ένα θεωρητικό υπόδειγμα, το οποίο αναφέρει ότι η εξαρτημένη μεταβλητή προκύπτει από μεταβολές των ανεξάρτητων μεταβλητών, και στη συνέχεια το υπόδειγμα της παλινδρόμησης προσεγγίζει -ή μπορεί να προσεγγίζει γραμμικά- την αιτιώδη θεωρητική σχέση.

Η παρουσίαση που ακολουθεί βασίζεται στους Ζαφειρόπουλος και Μυλωνάς (2018), Κατρακυλίδης, Κοντέος και Σαριαννίδης (2017), Ζαχαροπούλου (2011), Κάτος (2004), Συριόπουλος και Φίλιππας (2010) και Nimon and Oswald (2013). Επίσης, ακολουθεί τις σημειώσεις του μαθήματος «Ανάλυση Παλινδρόμησης και Ανάλυση Διακύμανσης» του μεταπτυχιακού «Εφαρμοσμένη Στατιστική» του Πανεπιστημίου Πειραιώς.

1.2 Μοντέλο με k ανεξάρτητες μεταβλητές

Το μοντέλο Πολλαπλής Γραμμικής Παλινδρόμησης εκφράζεται ως μια γραμμική συνάρτηση της εξαρτημένης μεταβλητής σε όρους των ανεξάρτητων μεταβλητών. Το υπόδειγμα για k ανεξάρτητες μεταβλητές, σε όρους μεμονωμένης παρατήρησης, γράφεται ως εξής:

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_k X_{ki} + e_i \quad , \qquad i = 1,2, \dots, n$$

όπου:

- Y_i είναι η τιμή της εξαρτημένης τυχαίας μεταβλητής Y για την i παρατήρηση. Ονομάζεται και ερμηνευόμενη μεταβλητή, εφόσον είναι αυτή της οποίας τις μεταβολές καλούμαστε να ερμηνεύσουμε αξιολογώντας τις τιμές των υπολοίπων ανεξάρτητων μεταβλητών.
- X_{1i}, X_{2i},…, X_{ki} είναι οι τιμές των ανεξάρτητων μεταβλητών X_1, X_2,…, X_k στην i παρατήρηση. Ονομάζονται και ερμηνευτικές μεταβλητές, αφού ερμηνεύουν ένα μέρος της μεταβλητότητας της εξαρτημένης μεταβλητής Y.
- b_0, b_1, b_2,…,b_k είναι οι συντελεστές παλινδρόμησης. Το b_0 ονομάζεται σταθερός όρος και εκφράζει την τιμή της εξαρτημένης μεταβλητής Y όταν οι ανεξάρτητες X_1, X_2,…, X_k ισούνται με μηδέν. Τα b_j για j=1,2,…,k εκφράζουν την μεταβολή στη μέση τιμή της εξαρτημένης μεταβλητής Y ($E[Y \mid X_{1i}, X_{2i}, \ldots, X_{ki}]$) όταν η μεταβλητή X_j μεταβάλλεται κατά 1 μονάδα, ενώ παράλληλα οι υπόλοιπες ανεξάρτητες μεταβλητές παραμένουν σταθερές. Ισχύει ότι αν $b_i>0$ τότε υπάρχει θετική γραμμική σχέση μεταξύ της Y και της ανεξάρτητης μεταβλητής X_i. Αντίστοιχα, αν $b_i<0$ τότε υπάρχει αρνητική γραμμική σχέση μεταξύ της Y και της X_i.
- ο δείκτης i δείχνει την κάθε παρατήρηση του δείγματος. Θεωρούμε ότι το δείγμα έχει n παρατηρήσεις.
- ο δείκτης k δείχνει το πλήθος των ανεξάρτητων μεταβλητών που χρησιμοποιούνται στο μοντέλο.
- e_i είναι τυχαίες ανεξάρτητες μεταβλητές που ονομάζονται τυχαία σφάλματα και ακολουθούν κανονική κατανομή με $N(0,\sigma^2)$. Τα τυχαία σφάλματα περιλαμβάνουν:
  - ερμηνευτικές μεταβλητές που επηρεάζουν την εξαρτημένη μεταβλητή Y, αλλά δεν έχουν συμπεριληφθεί στο μοντέλο
  - εισαγωγή περιττής ερμηνευτικής μεταβλητής
  - σφάλματα μέτρησης κατά τη συλλογή των δεδομένων
  - σφάλματα εξειδίκευσης του μοντέλου, όπως για παράδειγμα η σχέση της εξαρτημένης μεταβλητής Y με τις ανεξάρτητες X_1, X_2,…,X_k να είναι μη γραμμική, ενώ διερευνάται γραμμικό υπόδειγμα

1.3 Βασικές υποθέσεις του υποδείγματος

Προκειμένου να εκτελεστεί ορθά η ανάλυση της Γραμμικής Παλινδρόμησης πρέπει να ελεγχθούν ορισμένες υποθέσεις, ώστε τα αποτελέσματα και οι ερμηνείες τους να είναι αξιόπιστα. Οι υποθέσεις είναι οι εξής:

- κάθε μία από τις ανεξάρτητες μεταβλητές X_1, X_2,…,X_k συσχετίζεται γραμμικά με την εξαρτημένη μεταβλητή Y
- το τυχαίο σφάλμα e είναι τυχαία μεταβλητή με τις ακόλουθες ιδιότητες:
  - έχει μέση τιμή ίση με το μηδέν, δηλαδή $E(e_i)=0$. Αυτό σημαίνει ότι το τυχαίο σφάλμα e μπορεί να πάρει θετικές και αρνητικές τιμές, με την

προϋπόθεση όμως ότι είναι μηδέν το κατά μέσο όρο αποτέλεσμα της επίδρασης του τυχαίου σφάλματος στην εξαρτημένη μεταβλητή Y.

- έχει σταθερή διακύμανση (σ^2) σε κάθε επίπεδο των ανεξάρτητων μεταβλητών X_i, δηλαδή ισχύει ότι $Var(e_i)= \sigma^2$. Η υπόθεση αυτή ονομάζεται *υπόθεση της ομοσκεδαστικότητας*. Στην περίπτωση όπου η διακύμανση δεν διατηρείται σταθερή λέμε ότι υπάρχει ετεροσκεδαστικότητα.
- δεν υπάρχει αυτοσυσχέτιση μεταξύ των τυχαίων σφαλμάτων, δηλαδή $Cov(e_i,e_j)=0$. Πιο συγκεκριμένα, αν η σχέση δύο τυχαίων σφαλμάτων e_i και e_j χαρακτηρίζεται από ανεξαρτησία τότε η συνδιακύμανση θα ισούται με μηδέν. Η υπόθεση αυτή ονομάζεται *υπόθεση της ανεξαρτησίας των σφαλμάτων*.
- ακολουθεί κανονική κατανομή (με μέση τιμή ίση με μηδέν και σταθερή διακύμανση). Αποτέλεσμα αυτού είναι η τυχαία μεταβλητή Y να ακολουθεί κανονική κατανομή με μέση τιμή $b_0 + b_1X_{1i} + b_2X_{2i} + \dots + b_kX_{ki}$ και διακύμανση σταθερή και ίση με σ^2, δηλαδή:

$$Yi \sim N(b_0 + b_1X_{1i} + b_2X_{2i} + \dots + b_kX_{ki}, \sigma^2)$$

- καμία από τις ανεξάρτητες μεταβλητές $X_1, X_2,\dots,X_k$ δεν μπορεί να εκφραστεί σαν γραμμικός μετασχηματισμός μίας ή περισσότερων από τις ερμηνευτικές μεταβλητές. Η υπόθεση αυτή ονομάζεται *υπόθεση της μη ύπαρξης πολυσυγγραμμικότητας*. Στην περίπτωση που δύο (ή περισσότερες) ανεξάρτητες μεταβλητές είναι συγγραμμικές τότε δεν μπορούμε να απομονώσουμε τις ατομικές επιδράσεις της κάθε μεταβλητής στην εξαρτημένη Y και επομένως η ύπαρξη και των δύο (ή περισσοτέρων) ερμηνευτικών μεταβλητών δεν είναι σωστή. Συνεπώς, οι ερμηνευτικές μεταβλητές πρέπει να είναι ανεξάρτητες μεταξύ τους.
- ο αριθμός των παρατηρήσεων n είναι μεγαλύτερος από τον αριθμό των ανεξάρτητων μεταβλητών k συν 1, δηλαδή $k+1<n$. Αυτό πρέπει να συμβαίνει ώστε οι βαθμοί ελευθερίας να είναι θετικός αριθμός για να είναι εφικτή η διενέργεια στατιστικών ελέγχων στο υπόδειγμα.

Όταν ικανοποιούνται οι συγκεκριμένες υποθέσεις τότε το μοντέλο ονομάζεται κανονικό μοντέλο πολλαπλής παλινδρόμησης και μπορούμε να προχωρήσουμε στην ανάλυση μέσω της μεθόδου της Γραμμικής Παλινδρόμησης.

1.4 Μέθοδος των ελαχίστων τετραγώνων

Η εκτίμηση του γραμμικού υποδείγματος μπορεί να γίνει με διάφορες μεθόδους, ωστόσο στην διεθνή βιβλιογραφία συναντάται συνήθως η μέθοδος των ελαχίστων τετραγώνων. Πιο συγκεκριμένα, οι εκτιμητές των συντελεστών παλινδρόμησης προκύπτουν από την ελαχιστοποίηση του αθροίσματος των τετραγώνων των τυχαίων σφαλμάτων, δηλαδή την ελαχιστοποίηση του SSE. Το μέγεθος αυτό λέγεται

Άθροισμα Τετραγώνων των Σφαλμάτων (Sum of Squared Errors) και δίνεται από τη σχέση:

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$

Από τα μαθηματικά αποδεικνύεται ότι το ελάχιστο της συνάρτησης του SSE αντιστοιχεί στις τιμές των συντελεστών $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$,…,$\hat{b}_k$, οι οποίοι ονομάζονται συντελεστές ελαχίστων τετραγώνων. Η ελαχιστοποίηση του SSE αντιστοιχεί στην ελαχιστοποίηση της απόσταση των παρατηρήσεων από την ευθεία παλινδρόμησης. Το εκτιμώμενο μοντέλο που προκύπτει ονομάζεται γραμμικό μοντέλο παλινδρόμησης ελαχίστων τετραγώνων και είναι το εξής:

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \cdots + \hat{b}_k X_k$$

Το μοντέλο αυτό αντιστοιχεί στην μέση μεταβολή της εξαρτημένης μεταβλητής Y, εξαιτίας της μεταβολής κάποιας ερμηνευτικής μεταβλητής X_i κατά μία μονάδα, όταν οι υπόλοιπες ανεξάρτητες μεταβλητές διατηρούνται σταθερές. Υποθέτει ότι υπάρχει μία ενιαία γραμμική σχέση για όλα τα δεδομένα.

Οι εκτιμητές της συγκεκριμένης μεθόδου είναι αμερόληπτοι και παρουσιάζουν την ελάχιστη δυνατή διακύμανση.

1.5 Ιδιότητες των εκτιμητών ελαχίστων τετραγώνων

Οι εκτιμητές που προκύπτουν από την μέθοδο των ελαχίστων τετραγώνων έχουν ένα σύνολο από επιθυμητές ιδιότητες και κατατάσσονται στους Άριστους Γραμμικούς Αμερόληπτους Εκτιμητές (Best Linear Unbiased Estimator–BLUE). Οι ιδιότητες που τους κατατάσσουν σε αυτή την κατηγορία εκτιμητών είναι:

- οι εκτιμητές $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$,…,$\hat{b}_k$ είναι γραμμικοί, διότι αποτελούν γραμμικό συνδυασμό της εξαρτημένης μεταβλητής Y
- είναι αμερόληπτοι, διότι κατά μέσο όρο ισούνται με τις αντίστοιχες πραγματικές πληθυσμιακές παραμέτρους b_0,b_1,b_2,...,b_k, δηλαδή $E[\hat{b}_0]=b_0$, $E[\hat{b}_1]=b_1$, $E[\hat{b}_2]=b_2$, … , $E[\hat{b}_k]=b_k$
- είναι άριστοι, δηλαδή σύμφωνα με το θεώρημα των Gauss – Markov οι συγκεκριμένοι εκτιμητές έχουν τη μικρότερη διακύμανση από όλους τους γραμμικούς και αμερόληπτους εκτιμητές. Ωστόσο, για να ισχύει ο χαρακτηρισμός «άριστοι» πρέπει να είναι αποτελεσματικοί και συνεπείς. Η αποτελεσματικότητα αναφέρεται στην μικρότερη δυνατή διακύμανση των αμερόληπτων εκτιμητών, ενώ η συνέπεια σημαίνει ότι καθώς το μέγεθος του

δείγματος τείνει στο άπειρο οι τιμές των εκτιμητών τείνουν στις πραγματικές πληθυσμιακές τιμές των συντελεστών.

1.6 Συντελεστής πολλαπλού προσδιορισμού

Ένα μέτρο που υπολογίζει το πόσο καλά οι ανεξάρτητες μεταβλητές ερμηνεύουν την εξαρτημένη μεταβλητή Y είναι ο συντελεστής πολλαπλού προσδιορισμού R^2. Ο συντελεστής αυτός ορίζεται ως:

$$R^2 = \frac{SSR}{SST}$$

όπου:

- SSR είναι η ερμηνευμένη μεταβλητότητα, ή αλλιώς η μεταβλητότητα που οφείλεται στην παλινδρόμηση. Ονομάζεται *Άθροισμα Τετραγώνων της Παλινδρόμησης* (Regression Sum of Squares) και δίνεται από την σχέση:

$$SSR = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2$$

- SST είναι η συνολική μεταβλητότητα. Ονομάζεται *Συνολικό Άθροισμα Τετραγώνων* (Total Sum of Squares) και δίνεται από την σχέση:

$$SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

Η σχέση που συνδέει αυτά τα δύο μεγέθη είναι:

$$SST = SSR + SSE$$

Με απλά λόγια, η συνολική μεταβλητότητα της εξαρτημένης μεταβλητής Y οφείλεται στην μεταβλητότητα που ερμηνεύεται από την παλινδρόμηση και στην μεταβλητότητα που δεν ερμηνεύεται από αυτήν και ευθύνεται το τυχαίο σφάλμα.

Ο συντελεστής πολλαπλού προσδιορισμού αντιστοιχεί στο ποσοστό της συνολικής μεταβλητότητας της εξαρτημένης μεταβλητές Y που ερμηνεύεται από το υπόδειγμα πολλαπλής παλινδρόμησης. Ως ποσοστό μπορεί να πάρει τιμές από 0 έως 1, δηλαδή ισχύει ότι:

$$0 \leq R^2 \leq 1$$

Όσο ο συντελεστής προσεγγίζει την μονάδα τόσο καλύτερα ερμηνεύει το μοντέλο τις μεταβολές της εξαρτημένης μεταβλητής Y. Ωστόσο, αξίζει να σημειωθεί ότι αν

παραβιάζεται μία ή περισσότερες από τις συνθήκες των Gauss-Markoντότε μπορεί να προκύψουν υψηλές τιμές του R^2 οι οποίες οδηγούν σε λανθασμένα συμπεράσματα κατά την ερμηνεία.

1.6.1 Διορθωμένος συντελεστής πολλαπλού προσδιορισμού

Η τιμή του συντελεστής πολλαπλού προσδιορισμού αυξάνεται με την εισαγωγή στο μοντέλο επιπρόσθετων μεταβλητών, ακόμα κι όταν αυτές δεν συνεισφέρουν στην ερμηνεία της εξαρτημένης μεταβλητής. Για αυτόν τον λόγο συνίσταται η χρήση του διορθωμένου συντελεστή προσδιορισμού $\bar{R}^2$, ο οποίος δίνεται από την σχέση:

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1}$$

Ο συντελεστής αυτός αυξάνεται, μειώνεται ή παραμένει ίδιος με την εισαγωγή στο μοντέλο επιπρόσθετων ανεξάρτητων μεταβλητών ανάλογα με το αν αυτές συνεισφέρουν ή όχι στην ερμηνεία της εξαρτημένης μεταβλητής Y. Αν η εισαγόμενη μεταβλητή συνεισφέρει σημαντικά τότε ο συντελεστής αυξάνεται, ενώ αντίθετα μειώνεται.

Σημειώνεται ότι ο διορθωμένος συντελεστής πολλαπλού προσδιορισμού μπορεί να λάβει και αρνητικές τιμές.

1.6.2 Συντελεστής πολλαπλής συσχέτισης

Ο συντελεστής προσδιορισμού συνδέεται με τον συντελεστή συσχέτισης μέσω της σχέσης:

$$r = \pm\sqrt{R^2}$$

Ο συντελεστής αυτός ονομάζεται *συντελεστής γραμμικής συσχέτισης του Pearson* και μετρά τον βαθμό συσχετίσεως μεταξύ της εξαρτημένης μεταβλητής Y και όλων των ανεξάρτητων μεταβλητών X_1, X_2,…,X_k.

Οι τιμές που παίρνει ο συγκεκριμένος συντελεστής είναι από -1 έως 1. Αν είναι θετικός σημαίνει ότι οι μεταβλητές εμφανίζουν θετική γραμμική σχέση, ενώ αντίθετα αν είναι αρνητικός τότε υπάρχει αρνητική γραμμική σχέση μεταξύ τους. Όσο ο συντελεστής προσεγγίζει σε απόλυτη τιμή την μονάδα τόσο πιο ισχυρή γραμμική σχέση υπάρχει. Όταν λαμβάνει την τιμή μηδέν σημαίνει ότι οι μεταβλητές είναι ασυσχέτιστες μεταξύ τους.

1.7 Έλεγχος σημαντικότητας των συντελεστών

Είναι σημαντικό να ελεγχθεί αν η εξαρτημένη μεταβλητή Y εξαρτάται από τις ερμηνευτικές μεταβλητές που χρησιμοποιήθηκαν στο υπόδειγμα πολλαπλής

παλινδρόμησης. Για τον λόγο αυτόν γίνονται έλεγχοι υποθέσεων για κάθε έναν συντελεστή του μοντέλου. Οι υποθέσεις διατυπώνονται ως εξής:

H_0: $b_i=0$, ο συντελεστής δεν είναι στατιστικά σημαντικός
H_1: $b_i\neq0$, ο συντελεστής είναι στατιστικά σημαντικός

Η μηδενική υπόθεση απορρίπτεται (σε επίπεδο σημαντικότητας α) αν $|t| > t_{n-k-1;\frac{\alpha}{2}}$ με βάση την κριτική τιμή t, η οποία λαμβάνεται από τους σχετικούς πίνακες της κατανομής t με (n-k-1) βαθμούς ελευθερίας και επίπεδο σημαντικότητας α. Η τιμή της στατιστικής τυπολογίζεται από τον τύπο:

$$t = \frac{\hat{b}_i - b_i}{S_{\hat{b}_i}}$$

Απορρίπτοντας την μηδενική υπόθεση συμπεραίνουμε ότι υπάρχει στατιστικά σημαντική σχέση μεταξύ της εξαρτημένης μεταβλητής Y και της ανεξάρτητης μεταβλητής X_i.

1.7.1 Διαστήματα εμπιστοσύνης των συντελεστών

Ένας δεύτερος τρόπος να γίνουν οι έλεγχοι υποθέσεων είναι με την βοήθεια διαστημάτων εμπιστοσύνης, δηλαδή ενός φάσματος τιμών της πραγματικής παραμέτρου με βάση τις σημειακές εκτιμήσεις. Το διάστημα εμπιστοσύνης των συντελεστών υπολογίζεται ως εξής:

$$\hat{b}_i - t_{n-k-1;\frac{\alpha}{2}} \cdot S_{\hat{b}_i} \;\leq\; b_j \;\leq\; \hat{b}_i + t_{n-k-1;\frac{\alpha}{2}} \cdot S_{\hat{b}_i}$$

Αν το μηδέν βρίσκεται ανάμεσα σε αυτά τα δύο όρια, τότε δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση. Δηλαδή δεν υπάρχει στατιστικά σημαντική σχέση μεταξύ της εξαρτημένης μεταβλητής Y και της ανεξάρτητης X_i.

1.8 Έλεγχος σημαντικότητας του μοντέλου

Προκειμένου να ελεγχθεί η σημαντικότητα ολόκληρου του μοντέλου χρησιμοποιείται ο έλεγχος όλων των παραμέτρων των υπό εξέταση μεταβλητών. Ο έλεγχος αυτός αναφέρεται στην συνολική ικανότητα των ανεξάρτητων μεταβλητών να προσδιορίσουν τη συμπεριφορά της εξαρτημένης μεταβλητής Y. Οι υποθέσεις που γίνονται είναι οι εξής:

H_0: $b_1=b_2=\ldots=b_k=0$, το μοντέλο δεν είναι στατιστικά σημαντικό
H_1: τουλάχιστον ένας συντελεστής $b_j\neq0$, το μοντέλο είναι στατιστικά σημαντικό

Η μηδενική υπόθεση απορρίπτεται (σε επίπεδο σημαντικότητας α) όταν $F > F_{k;n-k-1;a}$, με βάση την κατανομή F με k βαθμούς ελευθερίας του αριθμητή, (n-k-1) βαθμούς ελευθερίας

του παρανομαστή και επίπεδο σημαντικότητας α. Η στατιστική ελέγχου F υπολογίζεται από τον τύπο:

$$F = \frac{SSR/k}{SSE/(n-k-1)}$$

Η απόρριψη της μηδενικής υπόθεσης σημαίνει ότι τουλάχιστον ένας από τους συντελεστές του υποδείγματος είναι διάφορος του μηδενός, δηλαδή είναι στατιστικά σημαντική τουλάχιστον μία από τις ανεξάρτητες μεταβλητές. Με άλλα λόγια, σε αυτήν την περίπτωση το μοντέλο μπορεί να προβλέψει τη συμπεριφορά της εξαρτημένης μεταβλητής Y. Ωστόσο, σύμφωνα με τους Draper και Smith (1981) πρέπει να ισχύει ότι $F > 4F_{k;n-k-1;a}$ για να είναι ικανοποιητική η ερμηνευτική ικανότητα του μοντέλου.

Αξίζει να σημειωθεί ότι είναι προτιμότερο να εφαρμοστεί ο έλεγχος F αντί να πραγματοποιηθούν όλοι οι επιμέρους έλεγχοι t, διότι με τον έλεγχο t δεν απαλείφεται το πρόβλημα της πολυσυγγραμμικότητας. Πιο συγκεκριμένα, μπορεί μεταξύ δύο ή περισσότερων ανεξάρτητων μεταβλητών να υπάρχει γραμμική σχέση, η οποία με τον έλεγχο t δεν φανερώνεται.

1.9 Ψευδομεταβλητές

Μέχρι στιγμής έχει γίνει αναφορά μόνο σε ποσοτικές μεταβλητές. Ωστόσο, σε πολλές περιπτώσεις μία ή περισσότερες ποιοτικές μεταβλητές ασκούν σημαντική επίδραση στην εξαρτημένη μεταβλητή. Λέγοντας ποιοτικές μεταβλητές εννοούμε τις μεταβλητές των οποίων τα χαρακτηριστικά δεν είναι μετρήσιμα.

Η εισαγωγή των ποιοτικών μεταβλητών στο μοντέλο παλινδρόμησης γίνεται με την χρήση των *ψευδομεταβλητών*. Η ψευδομεταβλητή αποδίδει αριθμητική έκφραση στα ποιοτικά χαρακτηριστικά. Συμβολίζεται, συνήθως, με D και οι τιμές της ονομάζονται επίπεδα.

Στην πράξη, για μία ποιοτική μεταβλητή, ο αριθμός των ψευδομεταβλητών που θα χρησιμοποιηθεί στο μοντέλο είναι ίσος με τον αριθμό των επιπέδων μείον ένα. Πιο συγκεκριμένα, αν μία ποιοτική μεταβλητή έχει δύο επίπεδα τότε θα χρησιμοποιηθεί μία ψευδομεταβλητή, η οποία θα λαμβάνει την τιμή 0 στο πρώτο επίπεδο και την τιμή 1 στο δεύτερο επίπεδο. Το επίπεδο που αντιστοιχεί στην τιμή 0 ονομάζεται επίπεδο αναφοράς ή ομάδα ελέγχου. Η μεταβλητή D μπορεί να εκφραστεί ως εξής:

$$D = \begin{cases} 0 & \textit{για απουσία συγκεκριμένου χαρακτηριστικού} \\ 1 & \textit{για παρουσία συγκεκριμένου χαρακτηριστικού} \end{cases}$$

Ο τρόπος που η ψευδομεταβλητή D θα εισαχθεί στο μοντέλο εξαρτάται από το μοντέλο και από το είδος των μεταβολών που χρειάζεται να μελετηθούν. Παρακάτω παρατίθενται όλες οι δυνατές περιπτώσεις.

Προς διευκόλυνση της ανάλυσης θα χρησιμοποιηθεί μία ποιοτική μεταβλητή με τρία επίπεδα (χαρακτηριστικά). Δημιουργούνται δύο ψευδομεταβλητές D_1 και D_2 που ορίζονται ως εξής:

$$D_1 = \begin{cases} 0 & για\ απουσία\ χαρακτηριστικού\ A \\ 1 & για\ παρουσία\ χαρακτηριστικού\ A \end{cases}$$

$$D_2 = \begin{cases} 0 & για\ απουσία\ χαρακτηριστικού\ B \\ 1 & για\ παρουσία\ χαρακτηριστικού\ B \end{cases}$$

1.9.1 Μεταβολή του σταθερού όρου

Στην πρώτη περίπτωση, εξετάζουμε την επίδραση της ποιοτικής μεταβλητής μέσα από τα επίπεδά της, ανεξαρτήτως των ποσοτικών χαρακτηριστικών και επομένως εισάγουμε τις ψευδομεταβλητές D_1 και D_2 προσθετικά στο υπόδειγμα.

Το μοντέλο παλινδρόμησης εξειδικεύεται ως εξής:

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + \cdots + b_kX_{ki} + \boldsymbol{\gamma_1 D_{1i}} + \boldsymbol{\gamma_2 D_{2i}} + e_i \quad , \qquad i = 1,2, \dots, n$$

το οποίο είναι ισοδύναμο με τρία μοντέλα, ένα για κάθε χαρακτηριστικό, ως εξής:

$$Y_i = \begin{cases} (b_0 + \boldsymbol{\gamma_1}) + b_1X_{1i} + b_2X_{2i} + \cdots + b_kX_{ki} + e_i & για\ παρουσία\ χαρακτηριστικού\ A \\ (b_0 + \boldsymbol{\gamma_2}) + b_1X_{1i} + b_2X_{2i} + \cdots + b_kX_{ki} + e_i & για\ παρουσία\ χαρακτηριστικού\ B \\ b_0 + b_1X_{1i} + b_2X_{2i} + \cdots + b_kX_{ki} + e_i & για\ παρουσία\ χαρακτηριστικού\ Γ \end{cases}$$

Η μεταβολή του σταθερού όρου από b_0 σε $(b_0+\gamma_1)$ ή $(b_0+\gamma_2)$ υποδεικνύει την διαφορά στην τιμή της εξαρτημένης μεταβλητής Y από την ύπαρξη του χαρακτηριστικού A ή B, αντίστοιχα.

Αξίζει να σημειωθεί ότι η διαφοροποίηση του σταθερού όρου συμβαίνει τότε μόνον, όταν οι συντελεστές παλινδρόμησης γ_1 και γ_2 είναι στατιστικά σημαντικοί. Ο έλεγχος γίνεται, όπως και στους υπόλοιπους συντελεστές, με την χρήση της στατιστικής $t = \frac{\hat{\gamma}_i - \gamma_i}{s_{\hat{\gamma}_i}}$.

Η ανάλυση επεκτείνεται και στην περίπτωση που η ποιοτική μεταβλητή έχει περισσότερα επίπεδα.

1.9.2 Μεταβολή της κλίσης

Στην δεύτερη περίπτωση, εξετάζουμε την μεταβολή της κλίσης της συνάρτησης, και όχι του σταθερού όρου. Αυτό πρακτικά σημαίνει ότι ελέγχουμε εάν η ποιοτική μεταβλητή μεταβάλει την εξαρτημένη μεταβλητή Y αναφορικά με μία ή περισσότερες ανεξάρτητες μεταβλητές X_i. Επομένως, εισάγουμε τις ψευδομεταβλητές D_1 και D_2 πολλαπλασιαστικά με τις ανεξάρτητες μεταβλητές του υποδείγματος.

Προς απλοποίηση των τύπων θα θεωρηθεί ότι υπάρχει μία ποσοτική ερμηνευτική μεταβλητή X και μία ποιοτική με τρία επίπεδα.

Το μοντέλο παλινδρόμησης εξειδικεύεται ως εξής:

$$Y_i = b_0 + b_1 X_i + \boldsymbol{\delta_1 D_{1i} X_i} + \boldsymbol{\delta_2 D_{2i} X_i} + e_i \quad , \qquad i = 1,2, \dots, n$$

το οποίο είναι ισοδύναμο με τρία μοντέλα, ένα για κάθε χαρακτηριστικό, ως εξής:

$$Y_i = \begin{cases} b_0 + (\boldsymbol{b_1 + \delta_1}) X_i + e_i & \textit{για παρουσία χαρακτηριστικού } A \\ b_0 + (\boldsymbol{b_1 + \delta_2}) X_i + e_i & \textit{για παρουσία χαρακτηριστικού } B \\ b_0 + b_1 X_i + e_i & \textit{για παρουσία χαρακτηριστικού } \Gamma \end{cases}$$

Η κλίση της συνάρτησης παλινδρόμησης μεταβάλλεται από b_1σε (b_1+δ_1) ή (b_1+δ_2), ενώ η σταθερά b_0 παραμένει αμετάβλητη. Αυτό σημαίνει ότι η εξαρτημένη μεταβλητή Υ διαφοροποιείται από την αλληλεπίδραση του ποσοτικού χαρακτηριστικού και ενός επιπέδου της ποιοτικής μεταβλητής. Ωστόσο, η μεταβολή της κλίσης πραγματοποιείται μόνον όταν οι συντελεστές παλινδρόμησης δ_1 και δ_2 είναι στατιστικά διάφοροι του μηδενός.

Η ανάλυση μπορεί να επεκταθεί αντίστοιχα και στην περίπτωση που η ποιοτική μεταβλητή έχει περισσότερα επίπεδα και το υπόδειγμα περιλαμβάνει περισσότερες από μία ποσοτικές ερμηνευτικές μεταβλητές.

1.9.3 Μεταβολή της κλίσης και του σταθερού όρου

Στην τρίτη περίπτωση, εξετάζουμε τον συνδυασμό των δύο παραπάνω περιπτώσεων, δηλαδή την ταυτόχρονη μεταβολή του σταθερού όρου και της κλίσης της συνάρτησης παλινδρόμησης. Η ψευδομεταβλητές D_1 και D_2 εισάγονται προσθετικά, αλλά και πολλαπλασιαστικά στις ανεξάρτητες μεταβλητές.

Όπως στην προηγούμενη περίπτωση, θα θεωρηθεί ότι υπάρχει μία ποσοτική ερμηνευτική μεταβλητή Χ και μία ποιοτική με τρία επίπεδα, προς απλοποίηση των τύπων.

Το μοντέλο παλινδρόμησης εξειδικεύεται ως εξής:

$$Y_i = b_0 + b_1 X_i + \boldsymbol{\gamma_1 D_{1i}} + \boldsymbol{\gamma_2 D_{2i}} + \boldsymbol{\delta_1 D_{1i} X_i} + \boldsymbol{\delta_2 D_{2i} X_i} + e_i \quad , \qquad i = 1,2, \dots, n$$

το οποίο είναι ισοδύναμο με τρία μοντέλα, ένα για κάθε χαρακτηριστικό, ως εξής:

$$Y_i = \begin{cases} (\boldsymbol{b_0 + \gamma_1}) + (\boldsymbol{b_1 + \delta_1}) \mathrm{X_i} + e_i & \textit{για παρουσία χαρακτηριστικού } A \\ (\boldsymbol{b_0 + \gamma_2}) + (\boldsymbol{b_1 + \delta_2}) \mathrm{X_i} + e_i & \textit{για παρουσία χαρακτηριστικού } B \\ b_0 + b_1 X_i + e_i & \textit{για παρουσία χαρακτηριστικού } \Gamma \end{cases}$$

Από τις παραπάνω εξισώσεις προκύπτει ότι τόσο η σταθερά όσο και η κλίση της συνάρτησης μεταβάλλονται ανάλογα με το επίπεδο της ποιοτικής μεταβλητής. Ωστόσο, η ταυτόχρονη μεταβολή γίνεται μόνον όταν οι συντελεστές παλινδρόμησης γ_1, γ_2, δ_1και δ_2 είναι στατιστικά σημαντικοί.

Η ανάλυση μπορεί να επεκταθεί αντίστοιχα και στην περίπτωση που η ποιοτική μεταβλητή έχει περισσότερα επίπεδα και το υπόδειγμα περιλαμβάνει περισσότερες από μία ποσοτικές ερμηνευτικές μεταβλητές.

1.10 Προβλέψεις

Απώτερος σκοπός της ανάλυσης παλινδρόμησης είναι η πρόβλεψη των τιμών της εξαρτημένης μεταβλητής Y, οι οποίες αντιστοιχούν σε δοσμένες τιμές των ερμηνευτικών μεταβλητών X_1, X_2,…,X_k. Για οποιεσδήποτε τιμές x_1,x_2,…,x_k των μεταβλητώνX_1, X_2,…,X_k, αντίστοιχα, και έχοντας εκτιμήσει τους συντελεστές b_0, b_1, b_2,…,b_k είναι εφικτή η σημειακή πρόβλεψη της τιμής της Y από το μοντέλο:

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \cdots + \hat{b}_k x_k$$

1.10.1 Μέση πρόβλεψη της μεταβλητής Y

Το διάστημα πρόβλεψης του μέσου της εξαρτημένης μεταβλητής Y, για δοσμένες τιμές των ερμηνευτικών μεταβλητών και για συντελεστή εμπιστοσύνης 1-α, είναι το εξής:

$$\hat{Y} - t_{n-k-1,\alpha/2} \cdot S_{\hat{Y}} \leq E(Y) \leq \hat{Y} + t_{n-k-1,\alpha/2} \cdot S_{\hat{Y}}$$

όπου $S_{\hat{Y}}$ είναι η τυπική απόκλιση που υπολογίζεται από τον αμερόληπτο εκτιμητή της διακύμανσης της πρόβλεψης.

Πιο συγκεκριμένα, αν πάρουμε ένα σύνολο από παρατηρήσεις τότε η μέση τιμή της Y θα βρίσκεται σε αυτό το διάστημα, με συντελεστή εμπιστοσύνης 1-α.

1.10.2 Ατομική πρόβλεψη της μεταβλητής Y

Το διάστημα πρόβλεψης της ατομικής τιμής της εξαρτημένης μεταβλητής Y, για δοσμένες τιμές των ερμηνευτικών μεταβλητών και για συντελεστή εμπιστοσύνης 1-α, είναι το εξής:

$$\hat{Y} - t_{n-k-1,\alpha/2} \cdot S_f \leq Y \leq \hat{Y} + t_{n-k-1,\alpha/2} \cdot S_f$$

όπου S_f είναι η τυπική απόκλιση που υπολογίζεται από τον αμερόληπτο εκτιμητή της διακύμανσης του σφάλματος της πρόβλεψης. Ως σφάλμα πρόβλεψης ορίζεται η διαφορά της εκτιμημένης τιμής $\hat{Y}$ από την πραγματική τιμή Y.

Πιο συγκεκριμένα, αν πάρουμε μία νέα παρατήρηση τότε η τιμή της Y θα βρίσκεται σε αυτό το διάστημα, με συντελεστή εμπιστοσύνης 1-α.

Είναι σημαντικό οι νέες τιμές των μεταβλητών X_1, X_2,…,X_k να βρίσκονται κοντά στα όρια του δείγματος βάσει του οποίου κατασκευάστηκε το διάστημα της πρόβλεψης. Για παρατηρήσεις που βρίσκονται μακριά από το δείγμα η πρόβλεψη είναι παρακινδυνευμένη, διότι όσο απομακρυνόμαστε από τα όρια του δείγματος αυξάνεται η πιθανότητα μεταβολής του προτύπου συμπεριφοράς.

1.11 Συμπεράσματα

Η Πολλαπλή Γραμμική Παλινδρόμηση είναι μία μέθοδος η οποία προσπαθεί να ερμηνεύσει τη μέση συμπεριφορά της μεταβλητής απόκρισης. Πιο συγκεκριμένα, θεωρεί ότι η σχέση της εξαρτημένης από τις ανεξάρτητες μεταβλητές είναι γραμμική και σταθερή σε όλο το εύρος της κατανομής της και αποσκοπεί στη μοντελοποίηση της σχέσης αυτής, με σκοπό την εύρεση των βέλτιστων εκτιμητών. Οι εκτιμητές αυτοί είναι σταθεροί σε ολόκληρη την κατανομή.

Επιπρόσθετα, η Γραμμική Παλινδρόμηση θέτει ορισμένες υποθέσεις που πρέπει να ικανοποιούνται, ώστε να είναι αξιόπιστα τα αποτελέσματα της. Οι υποθέσεις αυτές αφορούν την κανονική κατανομή, την σταθερή διακύμανση και την ανεξαρτησία των σφαλμάτων. Αν παραβιάζονται οι συγκεκριμένες υποθέσεις τότε οι ερμηνείες που προκύπτουν από το εκτιμημένο μοντέλο δεν είναι αξιόπιστες.

ΚΕΦΑΛΑΙΟ 2

Παλινδρόμηση Πεμπτημορίων

2.1 Εισαγωγή

Στις περισσότερες εφαρμογές της οικονομετρίας αναφερόμαστε σε μέσους όρους και προκύπτουν συμπεράσματα για τη μέση συμπεριφορά της εξαρτημένης μεταβλητής. Ωστόσο, σε πολλές περιπτώσεις είναι ενδιαφέρουσα η μελέτη ολόκληρης της κατανομής της εξαρτημένης μεταβλητής. Η μέθοδος η οποία δείχνει τι συμβαίνει σε όλο το εύρος της κατανομής ονομάζεται Παλινδρόμηση Πεμπτημορίων.

Η συγκεκριμένη μέθοδος είναι μία επέκταση της Γραμμικής Παλινδρόμησης, η οποία δεν αναφέρεται στην μέση τιμή της εξαρτημένης μεταβλητής, αλλά σε κάποιο ποσοστιαίο σημείο της. Χρησιμοποιείται σε περιπτώσεις όπου απαιτείται η μελέτη των ακραίων τιμών της ερμηνευόμενης μεταβλητής, προκειμένου να προκύψουν συμπεράσματα για τα ανώτερα ή κατώτερα στρώματα της υπό εξέτασης μεταβλητής.

Τα κύρια πεδία εφαρμογής της είναι στην οικονομική επιστήμη για την μελέτη της κατανομής του εισοδήματος για υψηλούς και χαμηλούς μισθούς, στις επιστήμες υγείας για τη διερεύνηση της σχέσης μεταξύ της κατάστασης υγείας ενός ατόμου και των παραγόντων σε διαφορετικά ποσοστιαία σημεία, αλλά και στην περιβαλλοντική επιστήμη για την μελέτη της κατανομής των ρύπων στο περιβάλλον. Επίσης, μπορεί να χρησιμοποιεί στον κλάδο της ασφάλισης, της διαχείρισης κινδύνων, της μηχανικής, της εκπαίδευσης και γενικότερα σε όποιον τομέα απαιτείται η διερεύνηση ολόκληρης της κατανομής της εξαρτημένης μεταβλητής. Ακόμα, αποτελεί ένα χρήσιμο εργαλείο όταν οι παραδοχές της Γραμμικής Παλινδρόμησης δεν πληρούνται.

Η παρουσίαση ακολουθεί τους Koenker και Hallock (2001), τους Rodriguez και Yao (2017), τους Maroto και Rios-Avila (2017), τον Koenker (2004), τους Johar και Katayama (2012), ton Waldmann (2018), τους Le Cook και Manning (2013), τους Ando και Tsay (2011), τους Petscher και Logan (2014) , τον Mingxiang (2015) και τον Γκανέτος (2007).

2.2 Μοντέλο με k ανεξάρτητες μεταβλητές

Το μοντέλο της Παλινδρόμησης Πεμπτημορίων στο τ ποσοστιαίο σημείο μπορεί να αναπαρασταθεί ως εξής:

$$Q_\tau(y_l) = b_0(\tau) + b_1(\tau)X_{i1} + b_2(\tau)X_{i2} + \ldots + b_k(\tau)X_{ik} \quad , \qquad i = 1,2,\ldots,n$$

όπου:

- $Q_\tau(y_i)$ είναι η τιμή της εξαρτημένης μεταβλητής Y για την i παρατήρηση στο τ ποσοστιαίο σημείο.
- X_{1i}, X_{2i},…, X_{ki} είναι οι τιμές των ανεξάρτητων μεταβλητών X_1, X_2,…, X_k στην i παρατήρηση.
- τ είναι το ποσοστιαίο σημείο στο οποίο γίνεται η ανάλυση. Σύμφωνα με τον ορισμό του ποσοστιαίου σημείου μία παρατήρηση η οποία βρίσκεται στο τ-οστό ποσοστημόριο είναι μεγαλύτερη από το τ% των παρατηρήσεων και μικρότερη από το (1-τ)% των παρατηρήσεων. Δηλαδή για τ=0,25 η πραγματική τιμή της Y θα είναι με 25% πιθανότητα μικρότερη από την τιμή που δίνει η συνάρτηση Qτ(yi), ενώ κατά 75% θα είναι μεγαλύτερη από αυτή. Το τ=0,50 αντιπροσωπεύει τη διάμεσο.
- $b_0(\tau)$, $b_1(\tau)$, $b_2(\tau)$,…,$b_k(\tau)$ είναι οι συντελεστές παλινδρόμησης στο τ ποσοστιαίο σημείο. Στη συγκεκριμένη μέθοδο αντί να είναι σταθεροί, οι συντελεστές παλινδρόμησης είναι συναρτήσεις και εξαρτώνται από το ποσοστιαίο σημείο. Δηλαδή τα b_j για j=1,2,…,k εκφράζουν την μεταβολή - *σε ένα συγκεκριμένο ποσοστημόριο τ* - της εξαρτημένης μεταβλητής Y όταν η μεταβλητή X_j μεταβάλλεται κατά 1 μονάδα, ενώ παράλληλα οι υπόλοιπες ανεξάρτητες μεταβλητές παραμένουν σταθερές. Σε αντίθεση με τη Γραμμική Παλινδρόμηση, εδώ για τις ερμηνείες των αποτελεσμάτων χρειάζεται να προσδιοριστεί σε ποιο ποσοστημόριο της Y αναφερόμαστε.
- ο δείκτης i δείχνει την κάθε παρατήρηση του δείγματος. Θεωρούμε ότι το δείγμα έχει n παρατηρήσεις.
- ο δείκτης k δείχνει το πλήθος των ανεξάρτητων μεταβλητών που χρησιμοποιούνται στο μοντέλο.
- e_i είναι τυχαίες ανεξάρτητες μεταβλητές που ονομάζονται τυχαία σφάλματα. Σε αντίθεση με την Γραμμική Παλινδρόμηση, δεν γίνονται υποθέσεις για την κατανομή, τη μέση τιμή και τη διακύμανση των τυχαίων σφαλμάτων στην συγκεκριμένη μέθοδο.

2.3 Μέθοδος εκτίμησης παραμέτρων

Σκοπός της Παλινδρόμησης Πεμπτημορίων είναι η εκτίμηση των τιμών των $b_j(\tau)$ για κάθε ποσοστιμόριο τ. Αυτό συμβαίνει με την ελαχιστοποίηση του μεγέθους:

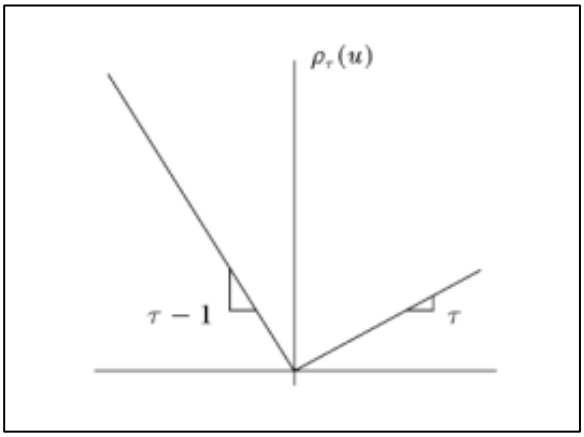
$$\sum_{i=1}^{n} \rho_\tau\big(y_i - (b_0(\tau) + b_1(\tau)X_{i1} + b_2(\tau)X_{i2} + \ldots + b_k(\tau)X_{ik})\big)$$

Ως ρ_τ θεωρείται η συνάρτηση απώλειας η οποία ορίζεται από τον τύπο:

$$\rho_\tau = \tau\, max(e, 0) \;+\; (1-\tau)\, max\,(-e, 0)$$

Στο παρακάτω σχήμα φαίνεται η γραμμική απεικόνιση της εν λόγω συνάρτησης:

Εικόνα 2.1 *Συνάρτηση ελέγχου ρ_τ*



i Πηγή: Koenker & Hallock (2001) p.146

Η συνάρτηση ρ_τ είναι η συνάρτηση ελέγχου η οποία δίνει ασύμμετρα βάρη στα τυχαία σφάλματα, ανάλογα με το τ ποσοστιαίο σημείο. Για τ=0,50 η συνάρτηση απώλειας αποτελεί τις ελάχιστες απόλυτες αποκλίσεις, αφού $\rho_\tau = \frac{1}{2}|e|$. Οπότε, η συνάρτηση ελαχιστοποίησης είναι η διάμεσος, αφού η διάμεσος ελαχιστοποιεί τις απόλυτες αποκλίσεις. Στα υπόλοιπα ποσοστιαία σημεία το τυπικό σφάλμα πολλαπλασιάζεται με τ αν το τυπικό σφάλμα είναι θετικό, ενώ αν είναι αρνητικό το τυπικό σφάλμα πολλαπλασιάζεται με (1-τ).

Με αυτόν τον τρόπο, σε κάθε ποσοστιαίο σημείο τ η ελαχιστοποίηση του ανωτέρω μεγέθους αποδίδει ένα ξεχωριστό σύνολο από εκτιμητές b_j. Συνεπώς, οι εκτιμητές δεν είναι σταθεροί όπως συμβαίνει στην Γραμμική Παλινδρόμηση, αλλά αποτελούν συναρτήσεις.

2.4 Συντελεστής προσδιορισμού

Στην Παλινδρόμηση Πεμπτημορίων χρησιμοποιείται ένας αντίστοιχος δείκτης με τον R^2 της Γραμμικής Παλινδρόμησης προκειμένου να υπολογιστεί το πόσο καλά τα δεδομένα ερμηνεύονται από το μοντέλο. Ο δείκτης αυτός είναι ο ψευδο-R^2 και υπολογίζεται ως εξής:

$$R^2_{pseudo}(\tau) = 1 - \frac{\rho_\tau}{SST}$$

Το ρ_τ αναπαριστά τη μη ερμηνευόμενη μεταβλητότητα, ενώ το SST αφορά τη συνολική μεταβλητότητα του υποδείγματος.

Η διαφορά του από τον συντελεστή προσδιορισμού της Γραμμικής Παλινδρόμησης είναι ότι το ψευδο-R^2 μετράει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητές Y που ερμηνεύεται από το υπόδειγμα σε ένα συγκεκριμένο ποσοστιμόριο και όχι στο σύνολο των δεδομένων.

Ως ποσοστό μπορεί να πάρει τιμές από 0 έως 1, δηλαδή ισχύει ότι:

$$0 \leq R^2 \leq 1$$

Εξακολουθεί να ισχύει ότι όσο ο συντελεστής προσεγγίζει την μονάδα τόσο καλύτερα ερμηνεύει το μοντέλο τις μεταβολές της εξαρτημένης μεταβλητής Y.

2.5 Έλεγχοι σημαντικότητας των συντελεστών

Προκειμένου να ελεγχθεί εάν οι ερμηνευτικές μεταβλητές του μοντέλου είναι στατιστικά σημαντικές, δηλαδή ασκούν σημαντική επίδραση στην μεταβλητή απόκρισης, γίνονται έλεγχοι υποθέσεων για κάθε έναν συντελεστή σε κάθε ποσοστιμόριο. Οι υποθέσεις αυτές είναι οι ίδιες με τις αντίστοιχες της Γραμμικής Παλινδρόμησης, με τη διαφορά ότι σε αυτή τη μέθοδο γίνεται ένας έλεγχος για κάθε συντελεστή και για κάθε ποσοστιαίο σημείο.

Οι υποθέσεις διατυπώνονται ως εξής:

H_0: $b_i(\tau)=0$, ο συντελεστής δεν είναι στατιστικά σημαντικός στο τ ποσοστιαίο σημείο

H_1: $b_i(\tau)\neq 0$, ο συντελεστής είναι στατιστικά σημαντικός στο τ ποσοστιαίο σημείο

Η τιμή της στατιστικής t και η κριτική τιμή t υπολογίζονται από τους ίδιους τύπους με την Γραμμική Παλινδρόμηση.

Η απόρριψη της μηδενικής υπόθεσης σημαίνει ότι υπάρχει στατιστικά σημαντική σχέση μεταξύ της εξαρτημένης μεταβλητής Y και της ερμηνευτικής μεταβλητής X_i στο τ ποσοστιμόριο.

2.6 Σύγκριση με την Γραμμική Παλινδρόμηση

Οι δύο μέθοδοι αποτελούν στατιστικές τεχνικές που χρησιμοποιούνται για την μοντελοποίηση των σχέσεων μεταξύ μίας εξαρτημένης μεταβλητής και μίας ή περισσοτέρων ερμηνευτικών μεταβλητών. Πιο συγκεκριμένα, εξετάζονται οι επιδράσεις που κάποιες μεταβλητές ασκούν σε κάποια άλλη. Ωστόσο, οι διαφορές είναι αρκετές.

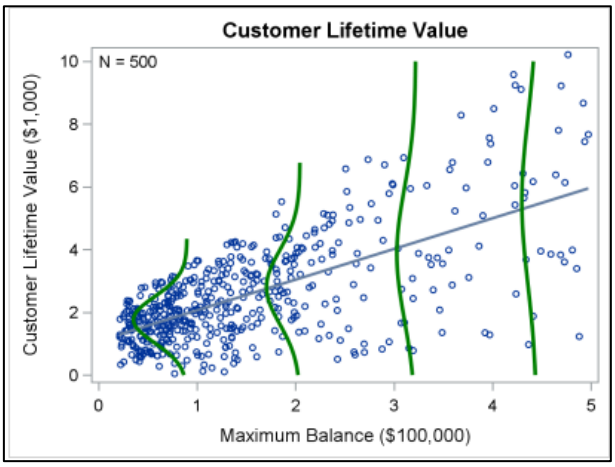
Κατ' αρχάς, η κρίσιμη διαφορά μεταξύ των δύο μεθόδων είναι ότι η Γραμμική Παλινδρόμηση επικεντρώνεται στη μέση τιμή της μεταβλητής απόκρισης, ενώ η Παλινδρόμηση Πεμπτημορίων παρέχει πληροφορίες για διαφορετικά ποσοστιαία σημεία της κατανομής. Η δεύτερη παρέχει μια πιο ολοκληρωμένη εικόνα της σχέσης των ερμηνευτικών μεταβλητών με την εξαρτημένη, αφού περιγράφει τη συμπεριφορά της μεταβλητής απόκρισης σε όλο το εύρος της.

Η Γραμμική Παλινδρόμηση χρησιμοποιείται όταν η έρευνα επικεντρώνεται στην εκτίμηση του μέσου αποτελέσματος των παραγόντων ή όταν ικανοποιούνται οι υποθέσεις της. Σε αντίθεση με αυτό, η Παλινδρόμηση Πεμπτημορίων χρησιμοποιείται όταν διερευνάτε πώς οι

παράγοντες επηρεάζουν διαφορετικά σημεία της κατανομής της μεταβλητής απόκρισης ή σε περιπτώσεις δεδομένων μη κανονικά κατανεμημένων ή λοξών.

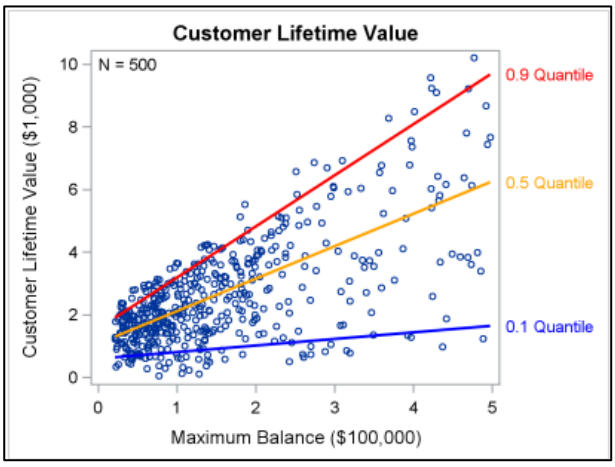
Στις παρακάτω εικόνες παρουσιάζεται ένα παράδειγμα παλινδρόμησης με μία ανεξάρτητη μεταβλητή. Η μεταβλητή απόκρισης είναι η αξία της διάρκειας ζωής του πελάτη και η ερμηνευτική μεταβλητή είναι το μέγιστο υπόλοιπο του τραπεζικού λογαριασμού του. Στην εικόνα 2.1 εφαρμόζεται Γραμμική Παλινδρόμηση (μπλε γραμμή), ενώ στην εικόνα 2.2 Παλινδρόμηση Πεμπτημορίων στα ποσοστιαία σημεία τ=0,10, τ=0,50 και τ=0,90 (μπλε, κίτρινη και κόκκινη γραμμή, αντίστοιχα).

Εικόνα 2.3 *Εφαρμογή Γραμμικής Παλινδρόμησης σε δεδομένα 500 πελατών τραπεζών*



iii Πηγή: Rodrigues & Yao (2017), p.2

Εικόνα 2.2 *Εφαρμογή Παλινδρόμησης Πεμπτημορίων σε δεδομένα 500 πελατών τραπεζών*



ii Πηγή: Rodrigues & Yao (2017), p.2

Στην εικόνα 2.2 η ευθεία παλινδρόμησης είναι η ευθεία που έχει προκύψει από την μέθοδο των ελαχίστων τετραγώνων. Ωστόσο, περιγράφει τη μέση κατάσταση του δείγματος και πολλά σημεία είναι αρκετά μακριά από αυτήν, ειδικά όσο αυξάνεται η ερμηνευτική μεταβλητή «υπόλοιπο λογαριασμού». Στην εικόνα 2.3 τα περισσότερα σημεία βρίσκονται κοντά σε κάποια από τις τρεις ευθείες και άρα δεν έχουν μεγάλη απόκλιση από αυτές. Η Παλινδρόμηση Πεμπτημορίων φαίνεται να περιγράφει καλύτερα όλο το σύνολο των σημείων.

Οι εκτιμητές της Γραμμικής Παλινδρόμησης είναι εκείνοι που ελαχιστοποιούν το άθροισμα των τετραγώνων των σφαλμάτων, ενώ στην Παλινδρόμηση Πεμπτημορίων οι εκτιμητές υπολογίζονται σε κάθε ποσοστιμόριο από την ελαχιστοποίηση του αθροίσματος των απόλυτων τιμών των σφαλμάτων. Οπότε, στη δεύτερη μέθοδο δημιουργείται ένα σύνολο εκτιμητών για κάθε ποσοστιαίο σημείο.

Επίσης, η Παλινδρόμηση Πεμπτημορίων δεν υποθέτει την κανονική κατανομή των σφαλμάτων ή τη σταθερή τους διακύμανση. Στις εικόνες 2.2 και 2.3 είναι ευδιάκριτο ότι η διακύμανση αυξάνεται, όπως και η μέση τιμή της μεταβλητής απόκρισης, καθώς αυξάνεται το «μέγιστο υπόλοιπο του τραπεζικού λογαριασμού». Αυτό σημαίνει ότι η συγκεκριμένη μέθοδος λειτουργεί και με ετεροσκεδαστικά δεδομένα, που δεν ακολουθούν κανονική κατανομή και η μέση τιμή τους μεταβάλλεται.

Η Γραμμική Παλινδρόμηση είναι αρκετά ευαίσθητη στις ακραίες παρατηρήσεις, γιατί ελαχιστοποιεί το άθροισμα των τετραγώνων των σφαλμάτων. Δηλαδή, οι ακραίες τιμές που έχουν μεγάλα τετραγωνικά σφάλματα επηρεάζουν δυσανάλογα την εκτίμηση των συντελεστών. Σε αντίθεση με αυτό, η Παλινδρόμηση Πεμπτημορίων ελαχιστοποιεί μια συνάρτηση ελέγχου των απόλυτων σφαλμάτων, οπότε είναι πιο ανθεκτική σε ακραίες τιμές.

Οι δύο μέθοδοι έχουν εφαρμογή σε πολλούς τομείς, ωστόσο μελετάνε διαφορετικά θέματα με διαφορετικό τρόπο. Πιο συγκεκριμένα, στις ανωτέρω εικόνες οι παλινδρομήσεις εφαρμόζονται για τη σχέση μεταξύ της αξίας της διάρκειας ζωής του πελάτη και του μέγιστου υπολοίπου του τραπεζικού λογαριασμού του. Αλλά στην Γραμμική Παλινδρόμηση ερμηνεύεται η μέση αξία της διάρκειας ζωής, ενώ στην Παλινδρόμηση Πεμπτημορίων εξετάζονται τα υψηλά στρώματα αξίας διάρκειας ζωής σε σχέση με τα πιο χαμηλά.

ΚΕΦΑΛΑΙΟ 3

Εμπειρική Ανάλυση

Στο παρόν κεφάλαιο γίνεται η εμπειρική σύγκριση των δύο μεθόδων παλινδρόμησης που έχουν αναλυθεί παραπάνω. Δημιουργείται ένα υπόδειγμα Γραμμικής Παλινδρόμησης και ένα Παλινδρόμησης Πεμπτημορίων με σκοπό τη σύγκριση τους ώστε να διαπιστωθεί ποιο από τα δύο έχει καλύτερη ερμηνευτική ικανότητα στην εξαρτημένη μεταβλητή.

Τα δεδομένα που θα χρησιμοποιηθούν είναι στοιχεία από την έρευνα SHARE. Τα στοιχεία αυτά είναι διεπιστημονικά και βασίζονται σε ατομικές παρατηρήσεις. Επιλέχτηκαν επειδή εμπεριέχουν πολλά θέματα για τα όποια θα είχε ενδιαφέρον να γίνει μία προσέγγιση στα ποσοστιμόρια. Τα κύρια θέματα που εμπεριέχονται αφορούν τους κλάδους της υγείας, της γήρανσης και της συνταξιοδότησης.

Αναλυτικότερα τα δεδομένα της ανάλυσης, τα μοντέλα παλινδρόμησης και οι μεταβλητές παρουσιάζονται παρακάτω.

3.1 Η έρευνα SHARE

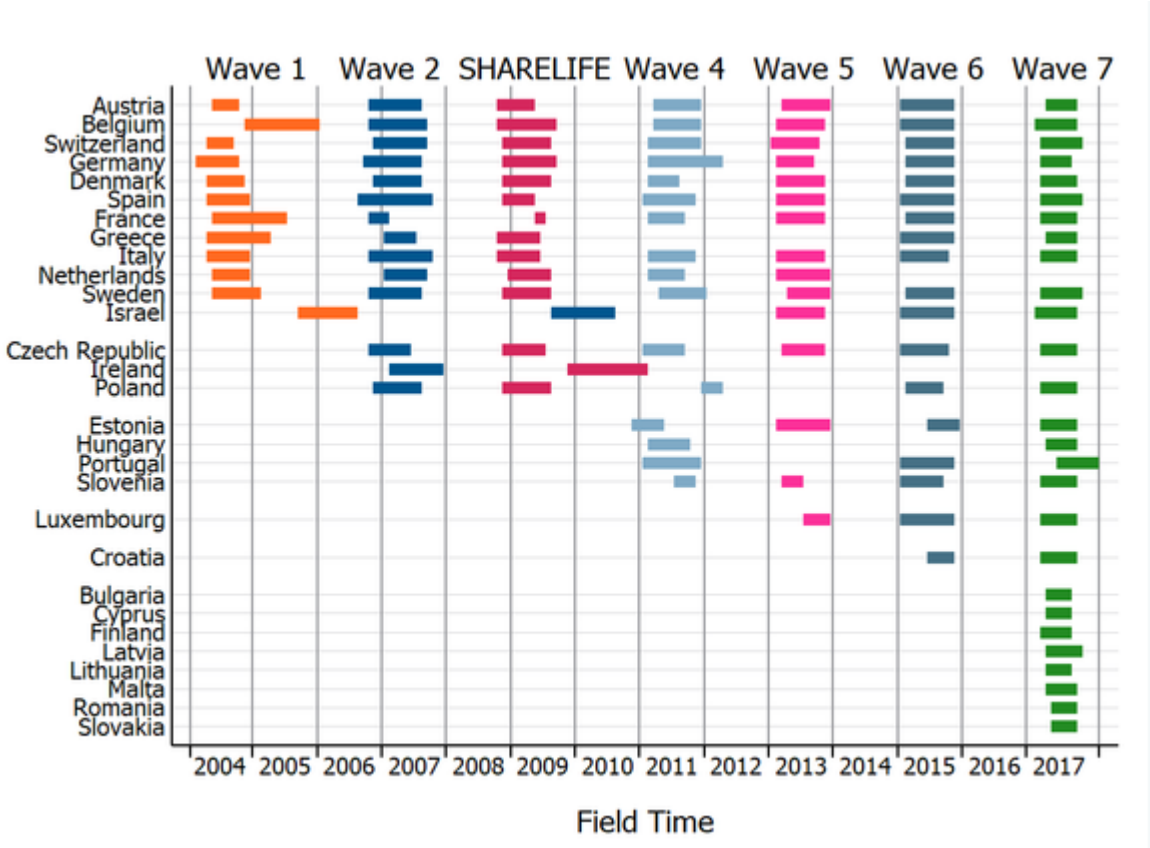
Η Έρευνα για την Υγεία, τη Γήρανση και τη Συνταξιοδότηση στην Ευρώπη (SHARE) είναι μια βάση δεδομένων, η οποία εμπεριέχει στοιχεία από διαφορετικούς επιστημονικούς κλάδους για τις χώρες της Ευρωπαϊκής Ένωσης και το Ισραήλ. Σκοπός της έρευνας είναι η μελέτη της κοινωνικοοικονομικής κατάστασης και της υγείας των ατόμων ηλικίας μεγαλύτερης των 50 ετών. Αποτελεί μια έρευνα πάνελ, δηλαδή εξετάζονται διαχρονικά τα ίδια άτομα και παρακολουθείτε η αλλαγή στις ζωές τους.

Το SHARE είναι πρωτοβουλία της Ευρωπαϊκής Επιτροπής, η οποία μάλιστα το χρηματοδοτεί. Αντίστοιχες έρευνες πραγματοποιούνται σε ολόκληρο τον κόσμο. Παράδειγμα αποτελεί το Health Retirement Study (HRS) στις ΗΠΑ και το English Longitudinal Survey on Ageing (ELSA) στην Αγγλία. Επίσης, συγκρίσιμες έρευνες διεξάγονται και στην Ιαπωνία, Κίνα, Βραζιλία και Μεξικό.

Η διεξαγωγή της έρευνας ξεκίνησε το 2004 με τη συμμετοχή στο 1^{ο} κύμα έντεκα χωρών, οι οποίες ήταν οι εξής: Αυστρία, Γερμανία, Σουηδία, Ολλανδία, Ισπανία, Ιταλία, Γαλλία, Δανία, Ελλάδα, Ελβετία και Βέλγιο. Την επόμενη χρονιά προστέθηκαν στοιχεία από το Ισραήλ. Το σύνολο των συμμετεχόντων στο 1^{ο} κύμα ήταν 30.419 άτομα. Το 2^{ο} κύμα πραγματοποιήθηκε το 2006-2007 σε τρεις επιπλέον χώρες, την Ιρλανδία, την Πολωνία και την Τσεχία, ενώ το 3^{ο}κύμα έγινε το 2008-2009. Κατά το 4^{ο} κύμα (2011) προστέθηκε η Ουγγαρία, η Πορτογαλία, η Σλοβενία και η Εσθονία. Το 5^{ο} κύμα πραγματοποιήθηκε το 2013 με την προσθήκη του Λουξεμβούργου. Η Ελλάδα δεν συμμετείχε κατά την διεξαγωγή του 4^{ου} και 5^{ου} κύματος, αλλά «επέστρεψε» στο 6^{ο} κύμα που έγινε το 2015 και στο οποίο προστέθηκε

η Κροατία. Κατά το 7ο κύμα (2017) προστέθηκε στο δείγμα η Λιθουανία, η Βουλγαρία, η Κύπρος, η Φιλανδία, η Λετονία, η Μάλτα, η Ρουμανία και η Σλοβακία. Το 7ο κύμα είναι το πρώτο στο οποίο συμμετείχαν όλες οι χώρες της ΕΕ.

Στην παρακάτω εικόνα παρουσιάζεται η συμμετοχή των χωρών στα πρώτα 7 κύματα της έρευνας SHARE.



Εικόνα 3.1 *Συμμετοχή χωρών στα κύματα 1-7. Διαθέσιμο στο :* https://share.cerge-ei.cz/data_overview_EN.htm *(Τελευταία πρόσβαση: 15/11/2023)*

Τέλος, το 8ο κύμα πραγματοποιήθηκε στο διάστημα 2019 έως 2020. Το σύνολο των συμμετεχόντων είναι 46.733 άτομα. Το συγκεκριμένο κύμα διεκόπη εξαιτίας της πανδημίας, έχοντας ολοκληρωθεί μόνο το 70% της δειγματοληψίας. Ως αποτέλεσμα των νέων υγειονομικών μέτρων πραγματοποιήθηκαν δύο έρευνες, η Corona Survey1 (2020) και η Corona Survey2(2021), οι οποίες αφορούν την επίδραση του Covid-19 στις ζωές των ανθρώπων και τις αλλαγές κατά τη διάρκεια της πανδημίας. Οι δυο τελευταίες έρευνες πραγματοποιήθηκαν με την μέθοδο CATI (Computer Assisted Telephone Interviewing), δηλαδή μέσω τηλεφωνικής συνέντευξης, σε αντίθεση με τα πρώτα 8 κύματα, τα οποία πραγματοποιήθηκαν με την μέθοδο CAPI (Computer Assisted Personal Interviewing), δηλαδή μέσω προσωπικής συνέντευξης.

Στην παρούσα διπλωματική εργασία θα χρησιμοποιηθούν δεδομένα από το 8ο κύμα της έρευνας SHARE.

3.2 Τα δεδομένα

Η παρούσα ανάλυση επικεντρώνεται στους συνταξιούχους και διερευνά τους παράγοντες που επιδρούν στο ύψος της σύνταξής τους. Από το σύνολο των συμμετεχόντων στο 8ο κύμα της έρευνας SHARE έχουν αφαιρεθεί άτομα μικρότερα των 50 ετών και άτομα που δεν λαμβάνουν σύνταξη. Με τον όρο σύνταξη εννοούμε την κύρια δημόσια σύνταξη γήρατος, την επικουρική, την πρόωρη, την κύρια ή/και επικουρική σύνταξη λόγω θανάτου του συζύγου και την πολεμική σύνταξη. Το ύψος της σύνταξης υπολογίζεται σε ετήσια βάση.

Οι μεταβλητές οι οποίες θα ερευνηθεί αν επηρεάζουν το ύψος της σύνταξης ενός ατόμου είναι το φύλο, η χώρα, τα χρόνια εκπαίδευσης, η οικογενειακή κατάσταση, η ηλικία συνταξιοδότησης, εάν το άτομο εργάζεται παράλληλα με την συνταξιοδότηση και κάποιες μεταβλητές που αφορούν της τελευταία εργασία πριν την συνταξιοδότηση. Όσον αφορά την τελευταία εργασία, οι μεταβλητές που εξετάζονται είναι εάν το άτομο αποτελούσε εργαζόμενο ή ελεύθερο επαγγελματία και εάν η εργασία απαιτούσε γνώσεις χρήσης ηλεκτρονικού υπολογιστή. Οι μεταβλητές αυτές χρησιμοποιούνται ευρύτατα από ερευνητές, για παράδειγμα στην διερεύνηση του χάσματος του φύλου στις συντάξεις (Betti, Bettio και Tinios, 2015).

Οι χώρες οι οποίες συμμετέχουν στο 8ο κύμα έχουν χωριστεί σε τρεις κατηγορίες: Κεντροευρωπαϊκές, Σκανδιναβικές και Μεσογειακές, σύμφωνα με το κοινωνικό κράτος του Esping-Andersen (Penda,2017). Στον παρακάτω πίνακα παρουσιάζεται η ομαδοποίηση τους στις τρεις κατηγορίες.

Πίνακας 3.1 *Κατηγοριοποίηση χωρών*

Κεντροευρωπαϊκές		**Σκανδιναβικές**	**Μεσογειακές**
Λουξεμβούργο	Αυστρία	Σουηδία	Ισπανία
Γαλλία	Σλοβενία	Δανία	Ιταλία
Γερμανία	Τσεχία	Φινλανδία	Ελλάδα
Βέλγιο	Πολωνία		Κύπρος
Ολλανδία	Εσθονία		Βουλγαρία
Αυστρία	Ουγγαρία		Μάλτα
Σλοβενία	Λιθουανία		Ρουμανία
Γαλλία	Λετονία		Κροατία
Γερμανία	Σλοβακία		
Βέλγιο	Ελβετία		
Ολλανδία			

Τα διάφορα επίπεδα εκπαίδευσης έχουν κωδικοποιηθεί βάση του ISCED 1997 (International Standard Classification of Education) (UNESCO, 2006). Η συγκεκριμένη ταξινόμηση καθιστά τα επίπεδα εκπαίδευσης συγκρίσιμα ανεξάρτητα από τη δομή των εθνικών εκπαιδευτικών συστημάτων. Το επίπεδο 0 αναφέρεται σε εκπαίδευση προσχολικού σταδίου, ενώ το επίπεδο 1 στην παροχή των βασικών γνώσεων γραφής, ανάγνωσης, μαθηματικών και άλλων μαθημάτων που παρέχονται στα 6 χρόνια πρωτοβάθμιας

εκπαίδευσης. Στην συνέχεια, κατά το επίπεδο 2 ολοκληρώνεται η βασική εκπαίδευση, συνήθως και υποχρεωτική, η οποία είναι η κατώτερη δευτεροβάθμια εκπαίδευση και κατά το επίπεδο 3 ολοκληρώνεται η ανώτερη δευτεροβάθμια εκπαίδευση. Στο επίπεδο 4 συμπεριλαμβάνονται προγράμματα τα οποία δεν μπορούν να θεωρηθούν τριτοβάθμια εκπαίδευση, αλλά ωστόσο το περιεχόμενο τους είναι ανώτερο της δευτεροβάθμιας. Το επίπεδο 5 αναφέρεται σε προγράμματα τριτοβάθμιας εκπαίδευσης, τα οποία δεν οδηγούν σε προχωρημένη ερευνητική πιστοποίηση. Τέλος, στο επίπεδο 6 είναι η τριτοβάθμια εκπαίδευση, με την έννοια ότι οδηγεί στην παροχή πτυχίου, μεταπτυχιακού ή διδακτορικού.

Όσον αφορά την οικογενειακή κατάσταση, το δείγμα χωρίζεται στους έγγαμους, στους άγαμους, στους διαζευγμένους και στους χήρους. Στην κατηγορία «έγγαμος» ανήκουν όσοι είναι παντρεμένοι, είτε συζούν με τον/την σύζυγο είτε διαμένουν σε διαφορετικές κατοικίες. Στην κατηγορία «άγαμος» ανήκουν τα άτομα που δεν έχουν παντρευτεί ποτέ.

Η ηλικία συνταξιοδότησης αναφέρεται στην ηλικία που είχε ο συνεντευξιαζόμενος όταν έλαβε σύνταξη για πρώτη φορά. Η συγκεκριμένη μεταβλητή έχει ομαδοποιηθεί σε κλάσεις, προκειμένου να ερευνηθεί εάν το ύψος της σύνταξης μεταβάλλεται ανάλογα με ένα εύρος ηλικίας. Εν προκειμένω, το δείγμα έχει χωριστεί σε όσους πήραν σύνταξη έως 56 χρονών, σε αυτούς που συνταξιοδοτήθηκαν μεταξύ 57 και 62 χρόνων, και τέλος, σε αυτούς που πήραν σύνταξη μετά τα 63.

3.3 Περιγραφική Ανάλυση

Καταρχάς, πρέπει να περιγραφεί το προφίλ των συνταξιούχων της Ευρώπης. Τα κύρια χαρακτηριστικά που θα μας απασχολήσουν θα διερευνηθούν μέσω δημογραφικών και κοινωνικοοικονομικών διαστάσεων. Πιο συγκεκριμένα, τα χαρακτηριστικά που θα περιγραφούν είναι το φύλο, η ηλικία, η χώρα, η οικογενειακή κατάσταση, το επίπεδο εκπαίδευσης, η ηλικία συνταξιοδότησης, η ύπαρξη παράλληλης εργασίας και κάποια στοιχεία για την τελευταία εργασία πριν την συνταξιοδότηση.

3.3.1 Μεταβλητή «Φύλο»

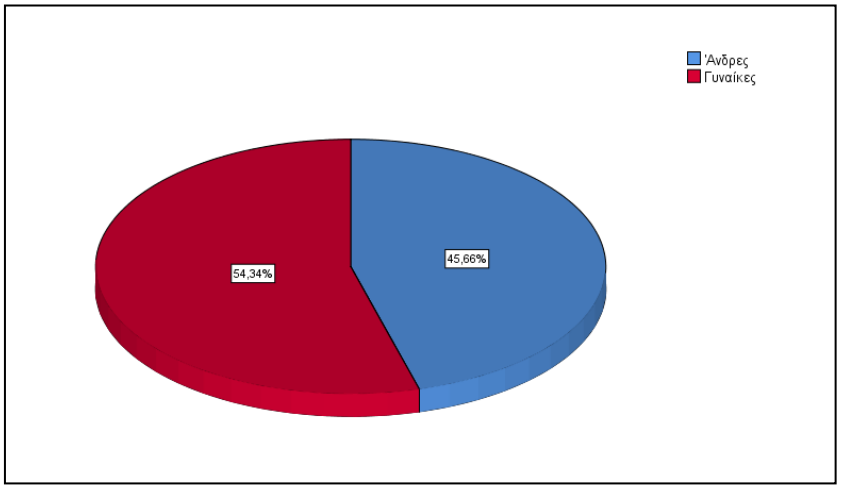
Το δείγμα αποτελείται από 28.195 άτομα, από τα οποία τα 12.875 είναι άνδρες και τα 15.320 είναι γυναίκες. Το ποσοστό των ανδρών ανέρχεται στο 45,7% και των γυναικών στο 54,3%.

Πίνακας 3.2 *Πίνακας συχνοτήτων για τη μεταβλητή «Φύλο»*

Φύλο	N	%
Άνδρες	12.875	45,7
Γυναίκες	15.320	54,3

Στο διάγραμμα πίτας φαίνεται οπτικά ότι το πλήθος των γυναικών, το οποίο εμφανίζεται με κόκκινο χρώμα, ξεπερνάει το πλήθος των ανδρών συμμετεχόντων, το οποίο είναι με μπλε.

Διάγραμμα 3.1 *Διάγραμμα πίτας για τη μεταβλητή «Φύλο»*



3.3.2 Μεταβλητή «Ηλικία»

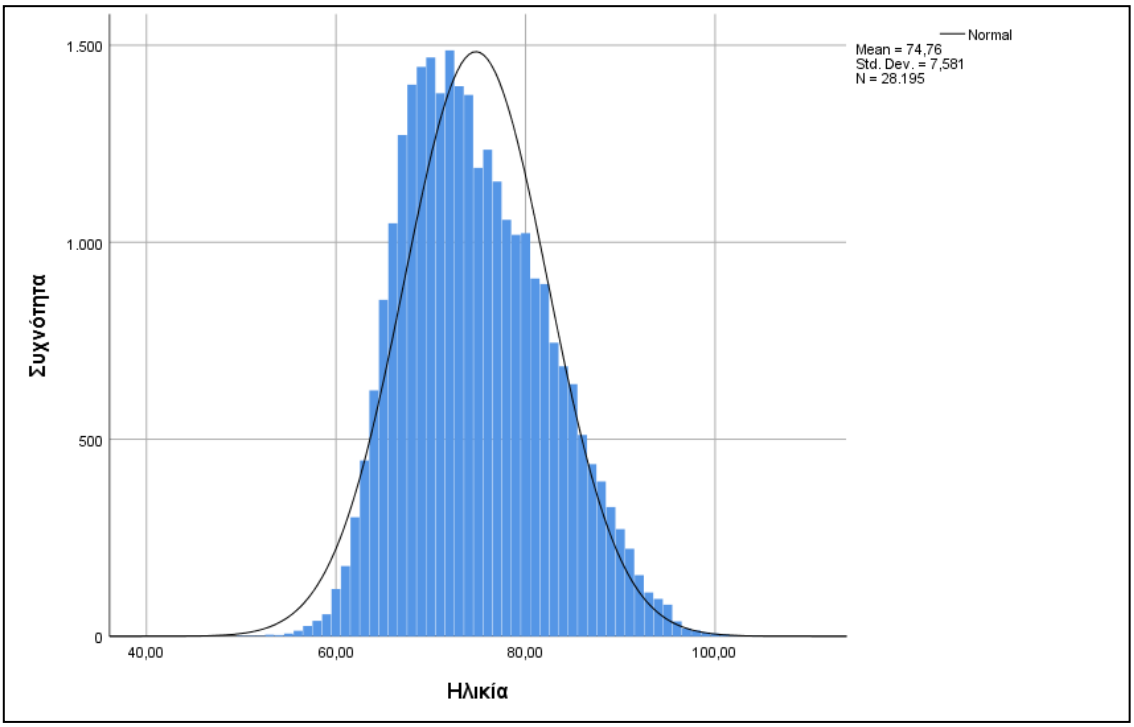
Όσον αφορά την ηλικία των συμμετεχόντων, η μέση τιμή είναι τα 75 έτη περίπου. Η τυπική απόκλιση, δηλαδή η διασπορά των παρατηρήσεων γύρω από το μέσο όρο, είναι 7,58. Τέλος, ο πιο μικρός σε ηλικία συμμετέχων είναι 50 χρόνων (η έρευνα SHARE αφορά άτομα άνω των 50 ετών), ενώ ο μεγαλύτερος είναι 104 χρόνων.

Πίνακας 3.3 *Περιγραφικά στατιστικά στοιχεία για τη μεταβλητή «Ηλικία»*

	Ηλικία
Ελάχιστη Τιμή	50
Μέγιστη Τιμή	104
Μέση Τιμή	74,76
Τυπική Απόκλιση	7,58

Παρακάτω παρουσιάζεται το ιστόγραμμα συχνοτήτων, όπου επιβεβαιώνονται σχηματικά τα ανωτέρω σχόλια. Φαίνεται ότι οι παρατηρήσεις παρουσιάζουν θετική ασυμμετρία. Αυτό σημαίνει ότι οι περισσότερες παρατηρήσεις συγκεντρώνονται αριστερά της κορυφής της καμπύλης.

Διάγραμμα 3.2 *Ιστόγραμμα συχνοτήτων για τη μεταβλητή «Ηλικία»*



Στη συνέχεια, διαχωρίστηκε το δείγμα σε τρεις κλάσεις ανάλογα με την ηλικία. Η ομάδα με το μεγαλύτερο πλήθος ατόμων είναι οι ηλικίες 70 έως 79. Η συγκεκριμένη ομάδα αποτελεί το 45,2% του συνολικού δείγματος. Οι άλλες δύο ομάδες έχουν περίπου τον ίδιο αριθμό συμμετεχόντων, αφού η κλάση «50-69» αποτελεί το 27,8% του δείγματος, ενώ η κλάση «80+» το 26,9% αυτού.

Πίνακας 3.4 *Πίνακας συχνοτήτων για τις ηλικιακές ομάδες*

Ηλικιακή Κλάση	N	%
50-69	7.841	27,8
70-79	12.758	45,2
80+	7.596	26,9

3.3.3 Μεταβλητή «Χώρα»

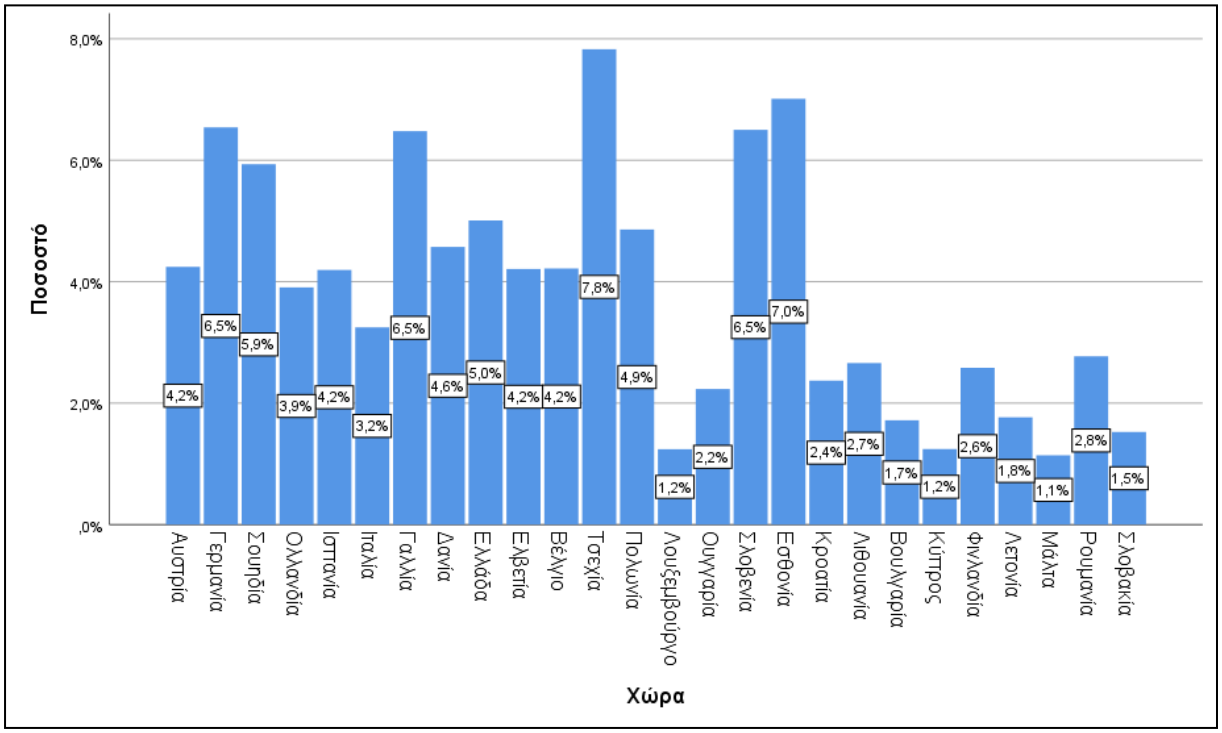
Οι χώρες, οι οποίες θα αναλυθούν, είναι 26 και παρουσιάζονται στον παρακάτω πίνακα συχνοτήτων (Πίνακας 3.4). Το υψηλότερο ποσοστό συμμετοχής εμφανίζει η Τσεχία με ποσοστό 7,8% του συνόλου των συμμετεχόντων και 2.207 άτομα από το σύνολο των 28.195. Η χώρα με τους λιγότερους συμμετέχοντες είναι η Μάλτα, η οποία συμμετέχει με 322 συνταξιούχους (ποσοστό 1,1%).

Πίνακας 3.5 *Πίνακας συχνοτήτων για τη μεταβλητή «Χώρα»*

Χώρα	N	%	Χώρα	N	%
Αυστρία	1.197	4,2	**Λουξεμβούργο**	350	1,2
Γερμανία	1.844	6,5	**Ουγγαρία**	629	2,2
Σουηδία	1.673	5,9	**Σλοβενία**	1.833	6,5
Ολλανδία	1.101	3,9	**Εσθονία**	1.977	7
Ισπανία	1.182	4,2	**Κροατία**	668	2,4
Ιταλία	916	3,2	**Λιθουανία**	750	2,7
Γαλλία	1.827	6,5	**Βουλγαρία**	484	1,7
Δανία	1.289	4,6	**Κύπρος**	351	1,2
Ελλάδα	1.412	5	**Φινλανδία**	728	2,6
Ελβετία	1.187	4,2	**Λετονία**	498	1,8
Βέλγιο	1.189	4,2	**Μάλτα**	322	1,1
Τσεχία	2.207	7,8	**Ρουμανία**	781	2,8
Πολωνία	1.370	4,9	**Σλοβακία**	430	1,5

Από το παρακάτω διάγραμμα επιβεβαιώνεται ότι η Τσεχία κατέχει το μεγαλύτερο ποσοστό του δείγματος. Αμέσως μετά, φαίνεται να ξεχωρίζει η Εσθονία, η Γερμανία, η Σλοβενία, η Γαλλία και η Σουηδία. Τα χαμηλότερα ποσοστά φαίνεται σχηματικά ότι κατέχει η Μάλτα, η Κύπρος και το Λουξεμβούργο.

Διάγραμμα 3.3 *Ραβδόγραμμα σχετικών συχνοτήτων για τη μεταβλητή «Χώρα»*



Όπως έχει αναφερθεί και παραπάνω, στην εργασία οι χώρες έχουν χωριστεί σε τρεις κατηγορίες: τις Κεντροευρωπαϊκές, τις Σκανδιναβικές και τις Μεσογειακές. Ακολουθεί πίνακας συχνοτήτων, στον οποίο είναι εμφανές ότι οι Κεντροευρωπαϊκές χώρες καταλαμβάνουν το μεγαλύτερο ποσοστό (65,2%), ακολουθούν οι Μεσογειακές με ποσοστό 21,7% και τέλος οι Σκανδιναβικές, οι οποίες αποτελούν το 13,1 % του δείγματος.

Πίνακας 3.6 *Πίνακας συχνοτήτων για τις ομάδες χωρών*

Ομάδα Χωρών	N	%
Κεντροευρωπαϊκές Χώρες	18389	65,2
Σκανδιναβικές Χώρες	3690	13,1
Μεσογειακές Χώρες	6116	21,7

3.3.4 Μεταβλητή «Οικογενειακή Κατάσταση»

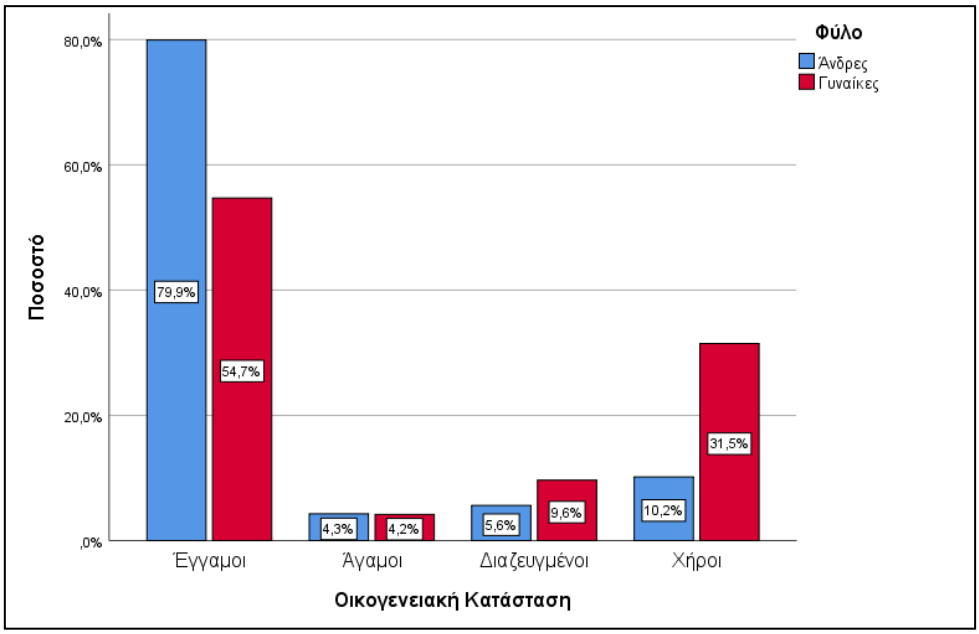
Στο σύνολο των ερωτηθέντων το μεγαλύτερο ποσοστό παρατηρείται ότι είναι έγγαμοι (ποσοστό 66,2%), ακολουθούν οι χήροι με ποσοστό 21,7%, οι διαζευγμένοι με ποσοστό 7,8% και τέλος οι άγαμοι αποτελούν μόλις το 4,2% του δείγματος. Και στα δύο φύλα, το μεγαλύτερο ποσοστό κατέχουν οι έγγαμοι, ενώ οι κατηγορίες «άγαμοι» και «διαζευγμένοι» αποτελούν πολύ μικρά ποσοστά του συνόλου των συνταξιούχων. Τέλος, οι γυναίκες χήρες αποτελούν το 17,1% του δείγματος, σε αντίθεση με το 4,6% των ανδρών, αντίστοιχα.

Πίνακας 3.7 *Πίνακας συχνοτήτων για τη μεταβλητή «Οικογενειακή Κατάσταση»*

Οικογενειακή Κατάσταση	Σύνολο		Άνδρες		Γυναίκες	
	N	%	N	%	N	%
Έγγαμοι	18.671	66,2	10.292	36,5	8.379	29,7
Άγαμοι	1.194	4,2	552	2	642	2,3
Διαζευγμένοι	2.200	7,8	723	2,6	1477	5,2
Χήροι	6.130	21,7	1308	4,6	4.822	17,1

Από το ραβδόγραμμα, ενδιαφέρον παρουσιάζει η μεγάλη διαφορά στο ποσοστό των γυναικών χήρων σε σχέση με τους άνδρες χήρους. Οι χήρες γυναίκες αποτελούν το 31,5% του συνολικού γυναικείου δείγματος, ενώ οι χήροι άνδρες μόλις το 10,2% του αντίστοιχου ανδρικού.

Διάγραμμα 3.3 *Ραβδόγραμμα σχετικών συχνοτήτων για τη μεταβλητή «Οικογενειακή Κατάσταση» ανά «Φύλο»*



iv Τα ποσοστά αναφέρονται στο σύνολο του εκάστοτε φύλου

3.3.5 Μεταβλητή «Εκπαίδευση»

Όπως έχει αναφερθεί και παραπάνω, η εκπαίδευση κατηγοριοποιείται, βάσει του διεθνούς προτύπου ISCED 1997, σε 6 επίπεδα. Στο σύνολο των ατόμων που συμμετείχαν στην έρευνα, το μεγαλύτερο ποσοστό εμφανίζεται στο επίπεδο 3 (ποσοστό 36,7%), το οποίο αντιστοιχεί στην ολοκλήρωση της ανώτερης δευτεροβάθμιας εκπαίδευσης. Πολύ μικρό είναι το ποσοστό των αναλφάβητων ατόμων (ποσοστό 4,3%), αλλά και των αποφοίτων τριτοβάθμιας εκπαίδευσης (ποσοστό 0,7%). Στις Κεντροευρωπαϊκές χώρες παρατηρείται παρόμοια κατάσταση με το σύνολο του δείγματος, δηλαδή το μεγαλύτερο ποσοστό (42,5%) αντιστοιχεί στο επίπεδο 3 του ISCED 1997. Όσον αφορά τις Σκανδιναβικές χώρες, το επίπεδο 5 αποτελεί το μεγαλύτερο ποσοστό (35%). Το επίπεδο αυτό αντιστοιχεί σε προγράμματα τριτοβάθμιας εκπαίδευσης χωρίς πιστοποίηση. Τέλος, στις Μεσογειακές χώρες δεν φαίνεται να υπάρχει μεγάλη απόκλιση στα υψηλότερα ποσοστά, αφού η διαφορά ανάμεσα στα επίπεδα 1 (ποσοστό 24,2%), 2 (ποσοστό 20,5%) και 3 (ποσοστό 23,7%) είναι μικρή.

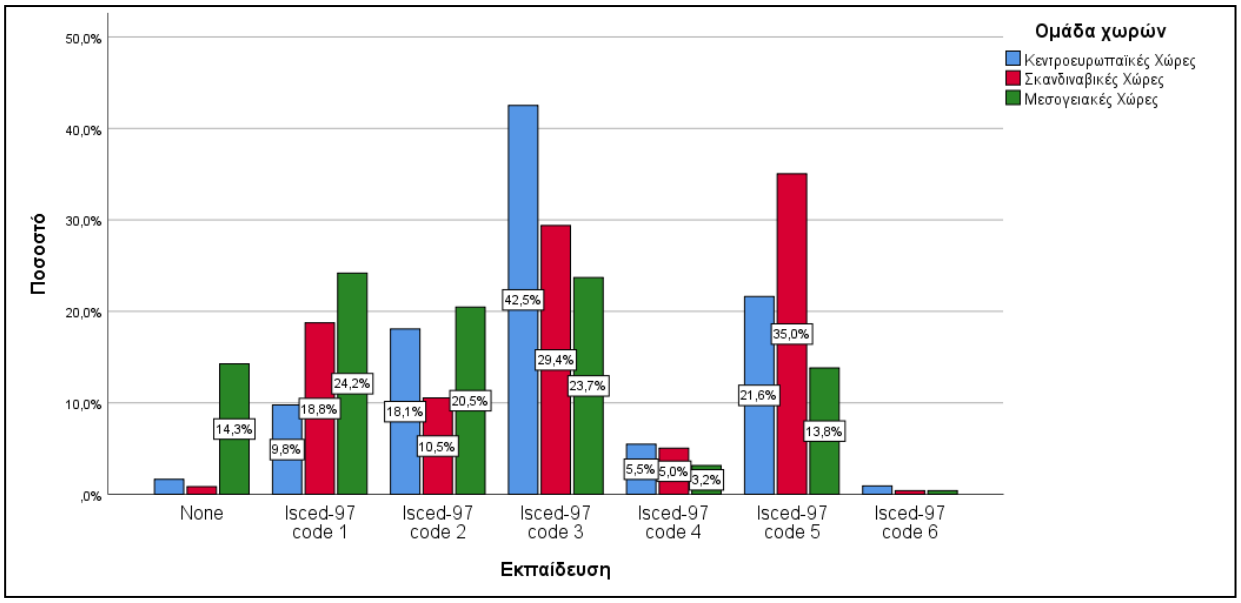
Πίνακας 3.8 *Πίνακας συχνοτήτων για τη μεταβλητή «Εκπαίδευση»*

Εκπαίδευση	Σύνολο		Κεντροευρωπαϊκές Χώρες		Σκανδιναβικές Χώρες		Μεσογειακές Χώρες	
	N	%	N	%	N	%	N	%
Καμία	1205	4,3	301	1,6	31	0,8	873	14,3
Πρωτοβάθμια	3968	14,1	1797	9,8	692	18,8	1479	24,2
Κατώτερη δευτεροβάθμια	4964	17,6	3323	18,1	389	10,5	1252	20,5

Ανώτερη δευτεροβάθμια	10349	36,7	7816	42,5	1084	29,4	1449	23,7
Μετά-δευτεροβάθμια	1386	4,9	1007	5,5	186	5	193	3,2
Πρώτο στάδιο τριτοβάθμιας	6115	21,7	3977	21,6	1293	35	845	13,8
Τριτοβάθμια	208	0,7	168	0,9	15	0,4	25	0,4

Διαγραμματικά φαίνεται ότι το μεγαλύτερο ποσοστό των Κεντροευρωπαϊκών χωρών αντιστοιχεί στο επίπεδο 3, των Σκανδιναβικών χωρών στο επίπεδο 5 και των Μεσογειακών χωρών στο επίπεδο 1 του ISCED 1997.

Διάγραμμα 3.4 *Ραβδόγραμμα σχετικών συχνοτήτων για τη μεταβλητή «Εκπαίδευση» ανά ομάδα χωρών*



ν Τα ποσοστά αναφέρονται στο σύνολο της εκάστοτε ομάδας χωρών

Όσον αφορά τα χρόνια εκπαίδευσης, φαίνεται από τον παρακάτω πίνακα ότι η ελάχιστη τιμή είναι 0, δηλαδή υπάρχουν άτομα στο δείγμα με μηδενική εκπαίδευση, ενώ η μέγιστη τιμή είναι 25, δηλαδή 25 χρόνια εκπαίδευσης. Η μέση τιμή των χρόνων εκπαίδευσης των συνταξιούχων του δείγματος είναι 11 χρόνια περίπου. Τέλος, η τυπική απόκλιση είναι 4,065 μονάδες.

Πίνακας 3.9 *Περιγραφικά στατιστικά στοιχεία για τη μεταβλητή «Χρόνια εκπαίδευσης»*

	Χρόνια εκπαίδευσης
Ελάχιστη Τιμή	0
Μέγιστη Τιμή	25
Μέση Τιμή	11,02
Τυπική Απόκλιση	4,065

3.3.6 Μεταβλητή «Ηλικία συνταξιοδότησης»

Αρχικά, το πρώτο πράγμα που παρατηρείται είναι ότι στην μεταβλητή «Ηλικία συνταξιοδότησης» υπάρχουν ελλιπείς τιμές, το σύνολο των οποίων αποτελεί το 36,8% του δείγματος. Αυτές οι τιμές είναι πιθανό να έχουν δημιουργηθεί επειδή ο ερωτώμενος έχει απαντήσει σε αυτήν την ερώτηση σε κάποιο από τα προηγούμενα κύματα.

Πίνακας 3.10 *Πίνακας συχνοτήτων για τις ελλιπείς τιμές της μεταβλητής «Ηλικία συνταξιοδότησης»*

	N	**%**
Έγκυρες τιμές	17.830	63,2
Ελλειπείς τιμές	10.365	36,8

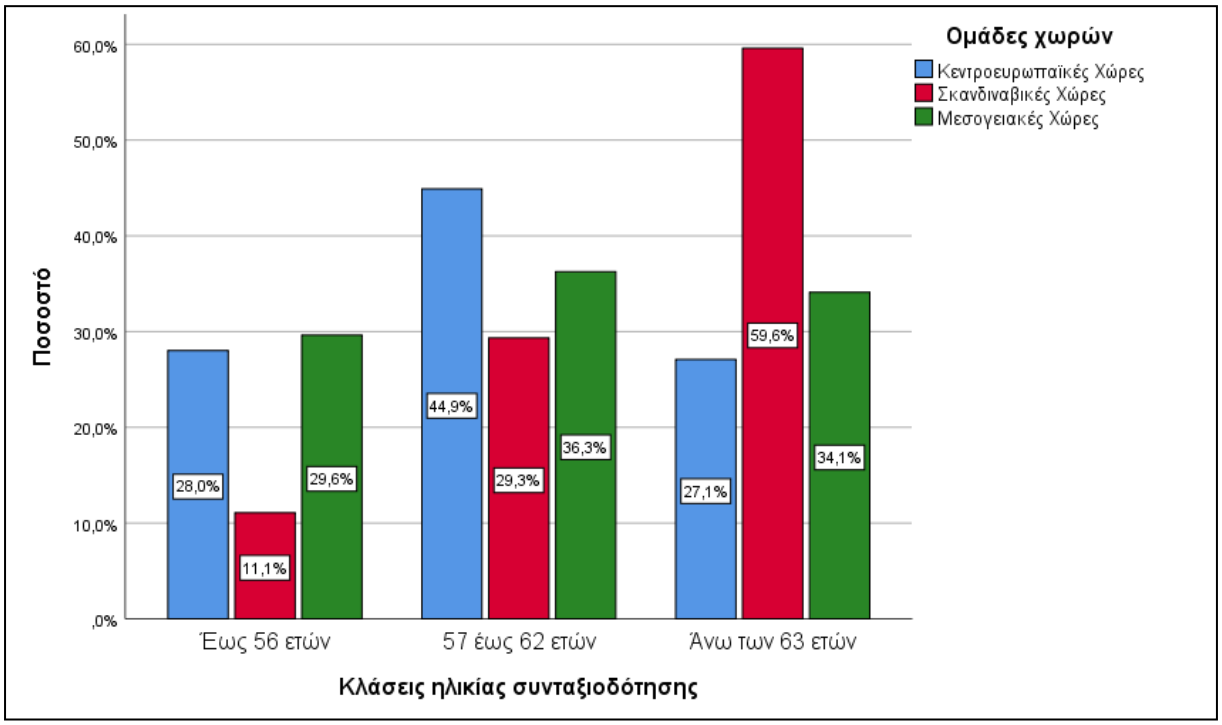
Στη συνέχεια, το σύνολο των 17.830 παρατηρήσεων έχει χωριστεί σε τρεις ηλικιακές κλάσεις, οι οποίες είναι: τα άτομα που συνταξιοδοτήθηκαν πριν τα 56 έτη, εκείνοι που πήραν πρώτη φορά σύνταξη μεταξύ 57 και 62 χρόνων και τέλος, αυτοί που πήραν σύνταξη μετά τα 63 έτη. Η κατηγορία «57 έως 62 ετών» κατέχει το υψηλότερο ποσοστό (41,1%), ακολουθεί η κατηγορία «άνω των 63 ετών» (με ποσοστό 32,5%) και το μικρότερο ποσοστό (26,4%) ανήκει στην κατηγορία «έως 56 ετών».

Πίνακας 3.11 *Πίνακας συχνοτήτων για τη μεταβλητή «Ηλικία συνταξιοδότησης»*

Κλάσεις Ηλικίας Συνταξιοδότησης	**Σύνολο**		**Κεντροευρωπαϊκές Χώρες**		**Σκανδιναβικές Χώρες**		**Μεσογειακές Χώρες**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Έως 56 ετών	4.711	26,4	3.281	28	229	11,1	1.201	29,6
57 έως 62 ετών	7.333	41,1	5.257	44,9	607	29,3	1.469	36,3
Άνω των 63 ετών	5.786	32,5	3.171	27,1	1.233	59,6	1.382	34,1

Από το ραβδόγραμμα παρατηρείται ότι, όσον αφορά τις Σκανδιναβικές χώρες, το υψηλότερο ποσοστό συμμετοχής (59,6%) είναι στην κλάση με τις υψηλότερες ηλικίες συνταξιοδότησης, το δεύτερο ποσοστό (29,3%) ανήκει στην μεσαία κλάση και τέλος, το μικρότερο ποσοστό (11,1%) αφορά την κλάση με τις ηλικίες συνταξιοδότησης έως 56 ετών. Ωστόσο, στις Κεντροευρωπαϊκές χώρες το υψηλότερο ποσοστό (44,9%) βρίσκεται στην κατηγορία «57 έως 62 ετών», με τις άλλες δύο κατηγορίες να έχουν πολύ μικρή απόκλιση στα ποσοστά. Τέλος, στις Μεσογειακές χώρες τα τρία ποσοστά συμμετοχής στο δείγμα κυμαίνονται σε πολύ κοντινά επίπεδα.

Διάγραμμα 3.5 *Ραβδόγραμμα σχετικών συχνοτήτων για τη μεταβλητή «Ηλικία συνταξιοδότησης» ανά ομάδα χωρών*



vi Τα ποσοστά αναφέρονται στο σύνολο της εκάστοτε ομάδας χωρών

3.3.7 Μεταβλητή «Παράλληλη εργασία με τη συνταξιοδότηση»

Οι ερωτηθέντες έχουν απαντήσει σχετικά με το εάν έχουν εργαστεί τις τελευταίες 4 εβδομάδες. Μόνο το 0,3% δεν έχει δώσει απάντηση.

Πίνακας 3.12 *Πίνακας συχνοτήτων για τις ελλιπείς τιμές της μεταβλητής «Παράλληλη εργασία με τη συνταξιοδότηση»*

	N	%
Έγκυρες τιμές	28.099	99,7
Ελλειπείς τιμές	96	0,3

Το μεγαλύτερο πλήθος των συνταξιούχων δεν συνεχίζουν να εργάζονται μετά τη συνταξιοδότηση. Ωστόσο, ένα 11,4% έχει δηλώσει ότι έχει εργαστεί τις τελευταίες 4 εβδομάδες, ανεξάρτητα της σύνταξης την οποία λαμβάνει.

Πίνακας 3.13 *Πίνακας συχνοτήτων για τη μεταβλητή «Παράλληλη εργασία με τη συνταξιοδότηση»*

Παράλληλη εργασία με τη συνταξιοδότηση	N	%
Ναι	3.211	11,4
Όχι	24.888	88,6

3.3.8 Μεταβλητές που αφορούν πληροφορίες της τελευταίας εργασίας πριν τη συνταξιοδότηση

Στη συνέχεια, έχουν ερευνηθεί κάποιες μεταβλητές οι οποίες αφορούν την τελευταία εργασία πριν συνταξιοδοτηθεί το άτομο. Πιο συγκεκριμένα, παρουσιάζονται περιγραφικά χαρακτηριστικά σχετικά με το εάν ο συνταξιούχος ήταν υπάλληλος ή ελεύθερος επαγγελματίας και σχετικά με το εάν η τελευταία εργασία του απαιτούσε τη γνώση χρήσης ηλεκτρονικού υπολογιστή.

Στις συγκεκριμένες μεταβλητές το μεγαλύτερο μέρος του δείγματος δεν έχει απαντήσει. Λιγότερο από το 20% έχει δώσει απάντηση στις ερωτήσεις που αφορούν την τελευταία εργασία του. Τα στοιχεία που παρουσιάζονται παρακάτω αφορούν το πλήθος των έγκυρων τιμών.

Πίνακας 3.14*Πίνακας συχνοτήτων για τις ελλιπείς τιμές των μεταβλητών για την τελευταία εργασία*

	N	%
Έγκυρες τιμές	5.422	19,2
Ελλειπείς τιμές	22.773	80,8

3.3.8.i Υπάλληλος ή ελεύθερος επαγγελματίας

Από το σύνολο των ατόμων που απάντησαν, το 53,7% δήλωσε ότι κατά την τελευταία του εργασία ήταν υπάλληλος του δημοσίου τομέα, το 38,3% δήλωσε ότι εργαζόταν στον ιδιωτικό τομέα, ενώ μόνο το 8% ήταν ελεύθεροι επαγγελματίες.

Πίνακας 3.15 *Πίνακας συχνοτήτων για τη μεταβλητή «Υπάλληλος ή ελεύθερος επαγγελματίας»*

Υπάλληλος ή ελεύθερος επαγγελματίας	N	%
Υπάλληλος ιδιωτικού τομέα	2.079	38,3
Υπάλληλος δημοσίου τομέα	2.910	53,7
Ελεύθερος επαγγελματίας	433	8

3.3.8.ii Απαίτηση χρήσης ηλεκτρονικού υπολογιστή

Όσον αφορά την απαίτηση γνώσης χρήσης ηλεκτρονικού υπολογιστή κατά την τελευταία εργασία πριν την συνταξιοδότηση, περίπου το 75% έχει απαντήσει αρνητικά.

Πίνακας 3.16 *Πίνακας συχνοτήτων για τη μεταβλητή «Απαίτηση χρήσης ηλεκτρονικού υπολογιστή»*

Απαίτηση χρήσης ηλεκτρονικού υπολογιστή	N	%
Ναι	1.407	25,7
Όχι	4.072	74,3

3.3.9 Εξαρτημένη μεταβλητή «Ετήσιο ύψος σύνταξης»

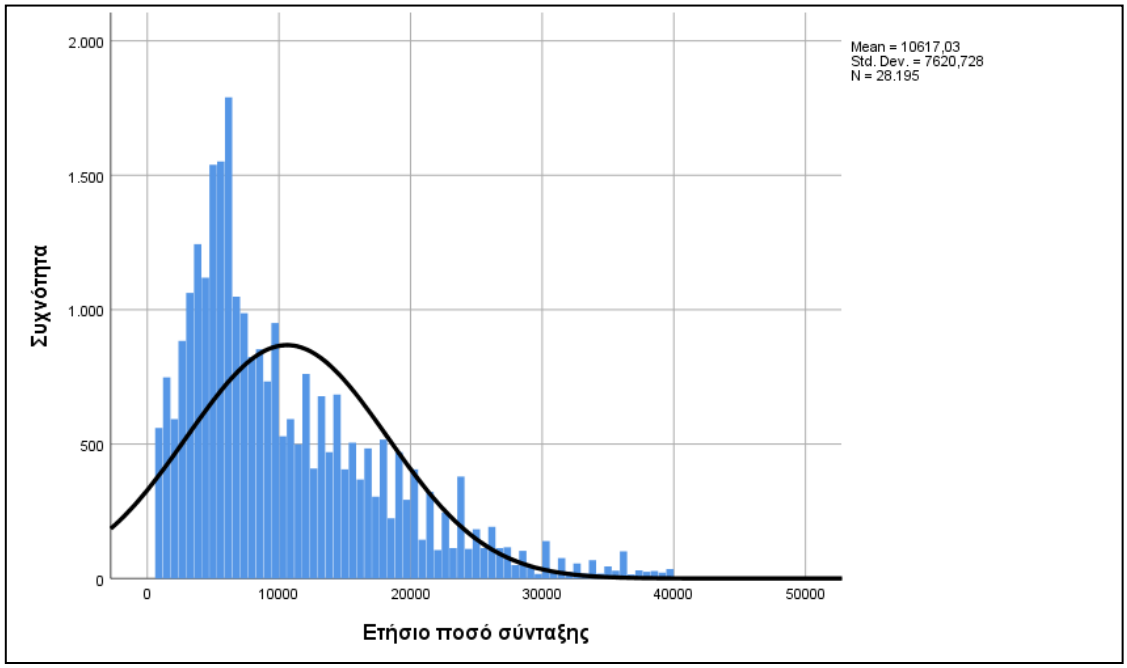
Όσον αφορά το ετήσιο ποσό της σύνταξης, η μέση τιμή είναι περίπου 10.600€ και η τυπική απόκλιση είναι 7.620. Η ελάχιστη τιμή του δείγματος είναι 600€ και η μέγιστη 40.000€, διότι έχουν διαγραφεί ακραίες παρατηρήσεις που πιθανόν ήταν λάθος.

Πίνακας 3.17 *Περιγραφικά στατιστικά στοιχεία για τη μεταβλητή «Ετήσιο ποσό σύνταξης»*

	Ετήσιο ύψος σύνταξης
Ελάχιστη Τιμή	600
Μέγιστη Τιμή	40.000
Μέση Τιμή	10.617,03
Τυπική Απόκλιση	7.620,73

Παρακάτω παρουσιάζεται το ιστόγραμμα συχνοτήτων. Φαίνεται ότι οι περισσότερες παρατηρήσεις συγκεντρώνονται στα χαμηλότερα ποσά σύνταξης και ελάχιστες βρίσκονται σε ποσά μεγαλύτερα των 30.000€ το χρόνο. Τέλος, η κατανομή απομακρύνεται πολύ από την κανονική.

Διάγραμμα 3.6 *Ιστόγραμμα συχνοτήτων για τη μεταβλητή «Ετήσιο ύψος σύνταξης»*



3.4 Πολλαπλή Γραμμική Παλινδρόμηση

Αφότου παρουσιάστηκαν συνοπτικά τα δεδομένα της έρευνας, ακολουθεί η διερεύνηση των παραγόντων που επιδρούν στο ύψος της σύνταξης των Ευρωπαίων συνταξιούχων. Η πρώτη μέθοδος, η οποία θα χρησιμοποιηθεί, είναι η Πολλαπλή Γραμμική Παλινδρόμηση.

Όπως έχει αναφερθεί και παραπάνω, οι ανεξάρτητες μεταβλητές οι οποίες θα ελεγχθεί εάν μεταβάλλουν το ύψος της σύνταξης είναι οι εξής:

- το φύλο του ατόμου,
- η χώρα (ή η ομάδα χωρών, που έχει προκύψει από την κατηγοριοποίηση των χωρών σε τρεις κατηγορίες),
- τα χρόνια εκπαίδευσης,
- η οικογενειακή κατάσταση,
- η ηλικία συνταξιοδότησης,
- εάν εργάζεται παράλληλα με την συνταξιοδότηση,
- εάν στην τελευταία εργασία του ήταν υπάλληλος ή ελεύθερος επαγγελματίας και
- εάν υπήρχε η απαίτηση γνώσης χρήσης ηλεκτρονικού υπολογιστή στην τελευταία εργασία του.

Οι περισσότερες από τις ανεξάρτητες μεταβλητές είναι κατηγορικές. Αυτό σημαίνει ότι θα εισαχθούν στο υπόδειγμα με χρήση ψευδομεταβλητών. Πιο συγκεκριμένα:

- Όσον αφορά τη μεταβλητή «Φύλο» (D_1):

$$D_1 = \begin{cases} 0 & (\textit{επίπεδο αναφοράς})\ \textit{Άνδρες} \\ 1 & \textit{Γυναίκες} \end{cases}$$

- Όσον αφορά τη μεταβλητή «ομάδα χωρών» (D_2 και D_3):

$$D_2 = \begin{cases} 0 & \textit{όχι Κεντροευρωπαϊκές χώρες} \\ 1 & \textit{Κεντροευρωπαϊκές χώρες} \end{cases}$$

$$D_3 = \begin{cases} 0 & \textit{όχι Σκανδιναβικές χώρες} \\ 1 & \textit{Σκανδιναβικές χώρες} \end{cases}$$

Επίπεδο αναφοράς: Μεσογειακές χώρες

- Όσον αφορά τη μεταβλητή «Οικογενειακή κατάσταση» (D_4, D_5 και D_6):

$$D_4 = \begin{cases} 0 & \textit{όχι έγγαμοι} \\ 1 & \textit{έγγαμοι} \end{cases}$$

$$D_5 = \begin{cases} 0 & \textit{όχι άγαμοι} \\ 1 & \textit{άγαμοι} \end{cases}$$

$$D_6 = \begin{cases} 0 & \textit{όχι διαζευγμένοι} \\ 1 & \textit{διαζευγμένοι} \end{cases}$$

Επίπεδο αναφοράς: Χήροι

- Όσον αφορά τη μεταβλητή «Ηλικία συνταξιοδότησης» (D_7 και D_8)

$$D_7 = \begin{cases} 0 & \text{όχι έως 56 ετών} \\ 1 & \text{έως 56 ετών} \end{cases}$$

$$D_8 = \begin{cases} 0 & \text{όχι 57 έως 62 ετών} \\ 1 & \text{57 έως 62 ετών} \end{cases}$$

Επίπεδο αναφοράς: άνω των 63 ετών

- Όσον αφορά τη μεταβλητή «Παράλληλη εργασία με συνταξιοδότηση» (D_9):

$$D_9 = \begin{cases} 0 & (\text{επίπεδο αναφοράς}) \text{ όχι} \\ 1 & \text{ναι} \end{cases}$$

- Όσον αφορά τη μεταβλητή «Υπάλληλος ή ελ. Επαγγελματίας» (D_{10} και D_{11}):

$$D_{10} = \begin{cases} 0 & \text{όχι υπάλληλος ιδιωτικού τομέα} \\ 1 & \text{υπάλληλος ιδιωτικού τομέα} \end{cases}$$

$$D_{11} = \begin{cases} 0 & \text{όχι υπάλληλος δημοσίου τομέα} \\ 1 & \text{υπάλληλος δημοσίου τομέα} \end{cases}$$

Επίπεδο αναφοράς: ελεύθερος επαγγελματίας

- Όσον αφορά τη μεταβλητή «Απαίτηση χρήσης ηλεκτρονικού υπολογιστή» (D_{12}):

$$D_{12} = \begin{cases} 0 & (\text{επίπεδο αναφοράς}) \text{ όχι} \\ 1 & \text{ναι} \end{cases}$$

Η μεταβλητή «Χρόνια εκπαίδευσης» είναι ποσοτική και θα συμβολίζεται με X_1.

Η γενική μορφή του μοντέλου είναι η εξής:

$$\text{Ύψος σύνταξης } (Y) = b_0 + b_1 X_1 + \gamma_1 D_1 + \gamma_2 D_2 + +\gamma_3 D_3 + \gamma_4 D_4 + \gamma_5 D_5 +$$

$$+\gamma_6 D_6 + \gamma_7 D_7 + \gamma_8 D_8 + \gamma_9 D_9 + \gamma_{10} D_{10} + \gamma_{11} D_{11} + \gamma_{12} D_{12}$$

3.4.1 Εύρεση του βέλτιστου μοντέλου

Στα οικονομετρικά υποδείγματα οι ερμηνευτικές μεταβλητές που χρησιμοποιούνται είναι εκείνες από τις οποίες προκύπτουν οι συντάξεις σύμφωνα με την θεωρητική ανασκόπηση. Ωστόσο, εδώ η ανάλυση είναι επικεντρωμένη στο στατιστικό κομμάτι , οπότε οι μεταβλητές που θα συμπεριληφθούν στο μοντέλο παλινδρόμησης είναι εκείνες οι οποίες ασκούν σημαντική στατιστική επίδραση στην εξαρτημένη μεταβλητή.

Η εύρεση του βέλτιστου μοντέλου θα γίνει με τη μέθοδο της προοδευτικής προσθήκης μεταβλητών (Forward Selection) (Lindsey και Sheather, 2010). Η συγκεκριμένη μέθοδος

ξεκινάει με το κενό μοντέλο, δηλαδή μόνο με τον σταθερό όρο, και προσθέτει διαδοχικά μία μεταβλητή στο μοντέλο. Η πρώτη μεταβλητή που εισέρχεται είναι εκείνη με την μεγαλύτερη θετική ή αρνητική συσχέτιση με την εξαρτημένη μεταβλητή. Στο επόμενο βήμα, προστίθεται η μεταβλητή με την αμέσως επόμενη στατιστική σημαντικότητα και η διαδικασία σταματάει όταν συμπεριληφθούν όλες οι μεταβλητές ή όταν δεν υπάρχει κάποια μεταβλητή η οποία δεν έχει ενταχθεί στο μοντέλο και παράλληλα δεν πληροί το κριτήριο ένταξης.

Το κριτήριο ένταξης, που χρησιμοποιείται στην παρούσα ανάλυση, είναι το επίπεδο σημαντικότητας α=0,05. Πιο συγκεκριμένα, εισάγονται στο μοντέλο οι μεταβλητές με επίπεδο σημαντικότητας μικρότερο του 0,05. Το κριτήριο επιλογής του βέλτιστου μοντέλου είναι η μεγαλύτερη τιμή (μεταξύ των μοντέλων) του διορθωμένου συντελεστή προσδιορισμού $\bar{R}^2$.

Στον πίνακα που ακολουθεί παρουσιάζονται οι συντελεστές των μοντέλων που προέκυψαν με τη μέθοδο της προοδευτικής προσθήκης μεταβλητών, καθώς και οι τιμές του κριτηρίου επιλογής του βέλτιστου μοντέλου.

Πίνακας 3.18 *Μοντέλα Γραμμικής Παλινδρόμησης με τη μέθοδο Forward Selection*

	Μεταβλητές	**Μοντέλο 1**	**Μοντέλο 2**	**Μοντέλο 3**	**Μοντέλο 4**	**Μοντέλο 5**	**Μοντέλο 6**	**Μοντέλο 7**
D_3	**Σκανδιναβικές χώρες**	12.013,5 (190,1)	10.937,1 (194,6)	10.829,6 (189,0)	10.899,6 (188,8)	10.956,5 (189,1)	10.905,5 (189,2)	10.559,7 (205,8)
D_{12}	**Απαίτηση χρήσης ηλ.υπολογιστή**	-	2.723,0 (153,8)	2.750,1 (149,3)	2.369,3 (162,8)	2.404,6 (162,8)	2.332,8 (163,5)	2.347,6 (163,2)
D_1	**Φύλο**	-	-	-2.288,4 (124,8)	-2.241,1 (124,7)	-2.250,9 (124,5)	-2.169,3 (125,9)	-2.117,8 (126,3)
X_1	**Χρόνια εκπαίδευσης**	-	-	-	106,7 (18,4)	111,6 (18,4)	111,2 (18,4)	123,9 (18,6)
D_9	**Παράλληλη εργασία**	-	-	-	-	-1.117,0 (275,6)	-1.145,6 (275,3)	-1.161,4 (274,9)
D_7	**Έως 56 ετών**	-	-	-	-	-	-572,5 (139,4)	-601,6 (139,4)
D_2	**Κεντροευρωπαικές χώρες**	-	-	-	-	-	-	-580,5 (136,8)
b_0	**Σταθερός όρος**	5.163,8 (70,7)	4.612,9 (75,5)	5.890,7 (101,1)	4.825,9 (209,7)	4.823,7 (209,4)	4.975,8 (212,4)	5.149,9 (216,0)

$\bar{R}^2$	0,422	0,453	0,485	0,488	0,489	0,491	0,492

vii Παρουσιάζονται οι εκτιμήσεις των συντελεστών και σε παρένθεση τα τυπικά σφάλματα των μοντέλων

Η πρώτη μεταβλητή που εισέρχεται στο μοντέλο είναι η D_3, η οποία αναφέρεται στο εάν το άτομο προέρχεται από Σκανδιναβική χώρα ή όχι. Η τιμή του διορθωμένου συντελεστή προσδιορισμού είναι 0,422. Στο δεύτερο μοντέλο προστίθεται η μεταβλητή «Απαίτηση χρήσης ηλεκτρονικού υπολογιστή» και ο συντελεστής αυξάνεται στο 0,453. Κατά το τρίτο βήμα της μεθόδου, η μεταβλητή «Φύλο» που εισάγεται μεταβάλει τον συντελεστή στο 0,485. Στη συνέχεια, προστίθεται η ποσοτική μεταβλητή «Χρόνια εκπαίδευσης» και ο συντελεστής μεταβάλλεται στο 0,488. Η πέμπτη μεταβλητή που εισάγεται στο μοντέλο είναι η «Παράλληλη εργασία με τη συνταξιοδότηση» και η τιμή του συντελεστή είναι 0,489. Κατά το έκτο βήμα, προστίθεται η ψευδομεταβλητή D_7, η οποία αναφέρεται στο εάν το άτομο πήρε σύνταξη έως τα 56 έτη και ο συντελεστή μεταβάλλεται στο 0,491.

Το μοντέλο που προκρίνεται είναι αυτό που προκύπτει από το έβδομο βήμα της μεθόδου προοδευτικής προσθήκης μεταβλητών με την εισαγωγή της ψευδομεταβλητής D_2, η οποία αφορά εάν το άτομο είναι από Κεντροευρωπαϊκή χώρα ή όχι. Το συγκεκριμένο υπόδειγμα περιλαμβάνει τις μεταβλητές: Χρόνια εκπαίδευσης (X_1), Φύλο (D_1), Κεντροευρωπαϊκές χώρες (D_2), Σκανδιναβικές χώρες (D_3), Συνταξιοδότηση έως 56 ετών (D_7), Παράλληλη εργασία με τη συνταξιοδότηση (D_9) και Απαίτηση χρήσης ηλεκτρονικού υπολογιστή (D_{12}), οι οποίες είναι στατιστικά σημαντικές στο επίπεδο σημαντικότητας α=0,05 που έχει οριστεί σαν κριτήριο ένταξης.

Ωστόσο, η ανάλυση θα συνεχίσει με το μοντέλο 4, διότι η προσθήκη των τελευταίων τριών μεταβλητών προσφέρει ελάχιστα στην ερμηνεία της εξαρτημένης μεταβλητής από το υπόδειγμα. Η διαφορά του $\bar{R}^2$ του $7^{ου}$ μοντέλου από του $4^{ου}$ είναι μόλις 0,004 μονάδες.

Στον πίνακα 3.18 παρουσιάζονται όλα τα στοιχεία του υποδείγματος.

Πίνακας 3.19 *Βέλτιστο μοντέλο Γραμμικής Παλινδρόμησης*

	Μεταβλητές	**Μοντέλο 4**	**Standardized beta**	**p-value**
D_3	**Σκανδιναβικές χώρες**	10.899,6* (188,8)	0,589	0,000
D_{12}	**Απαίτηση χρήσης ηλ.υπολογιστή**	2.369,3* (162,8)	0,162	0,000
D_1	**Φύλο**	-2.241,1* (124,7)	-0,174	0,000
X_1	**Χρόνια εκπαίδευσης**	106,7* (18,4)	0,062	0,000

b_0	**Σταθερός όρος**	4.825,9* (209,7)
	$\bar{R}^2$	0,488

viii Παρουσιάζονται οι εκτιμήσεις των συντελεστών και σε παρένθεση τα τυπικά σφάλματα των μοντέλων. Οι στατιστικά σημαντικές εκτιμήσεις συνοδεύονται με «» σε επίπεδο σημαντικότητας α=0,05.*

Το βέλτιστο μοντέλο που προέκυψε είναι το:

$$Ύψος\ σύνταξης = 4.825{,}9 + 106{,}7\ X_1 - 2.241{,}1\ D_1 + 10.899{,}6\ D_3\ + 2.369{,}3\ D_{12}$$

Παρατηρείται θετική σχέση των χρόνων εκπαίδευσης με το ύψος της σύνταξης, δηλαδή όσο αυξάνονται τα χρόνια που έχει εκπαιδευτεί κάποιος τόσο αυξάνεται και το ποσό της σύνταξης που λαμβάνει. Επίσης, οι γυναίκες λαμβάνουν μικρότερη σύνταξη σε σχέση με τους άντρες, αφού υπάρχει αρνητική σχέση της ψευδομεταβλητής «Φύλο» (D_1, όπου D_1=1 αναφέρεται στις γυναίκες) με το ετήσιο ποσό σύνταξης. Τα άτομα που προέρχονται από Σκανδιναβική χώρα έχουν μεγαλύτερη σύνταξη σε σχέση με εκείνους με διαφορετική καταγωγή, αφού υπάρχει θετική σχέση της ψευδομεταβλητής «Σκανδιναβικές Χώρες» (D_3, όπου D_3=1 αναφέρεται σε άτομα με καταγωγή από Σκανδιναβία) με το ύψος σύνταξης. Τέλος, η απαίτηση χρήσης ηλεκτρονικού υπολογιστή στην τελευταία εργασία πριν την συνταξιοδότηση αυξάνει το ύψος της σύνταξης του ατόμου, αφού υπάρχει θετική σχέση της μεταβλητής D_{12} (όπου D_{12}=1 αναφέρεται στην ύπαρξη απαίτησης) σε σχέση με το ύψος της σύνταξης.

Στη συνέχεια ακολουθεί η ερμηνεία των συντελεστών της πολλαπλής παλινδρόμησης;

- Όσον αφορά τον σταθερό όρο: το ύψος της ετήσιας σύνταξης διαμορφώνεται στα 4.825,9€, εφόσον το άτομο είναι άνδρας με μηδενική εκπαίδευση, προερχόμενος από χώρα που δεν υπάγεται στη Σκανδιναβία και η τελευταία του εργασία πριν τη συνταξιοδότηση δεν απαιτούσε γνώσεις ηλεκτρονικού υπολογιστή.
- Όσον αφορά τον συντελεστή β_1 της μεταβλητής «Χρόνια εκπαίδευσης»: για κάθε έναν χρόνο εκπαίδευσης, το ετήσιο ποσό σύνταξης αυξάνεται κατά 106,7€ (όταν όλες οι άλλες μεταβλητές διατηρούνται σταθερές).
- Όσον αφορά τον συντελεστή β_2 της μεταβλητής «Φύλο»: το ύψος της ετήσιας σύνταξης είναι κατά 2.241,1€ μικρότερο εάν πρόκειται για γυναίκα σε σχέση με την περίπτωση του άνδρα (όταν όλες οι άλλες μεταβλητές διατηρούνται σταθερές).
- Όσον αφορά τον συντελεστή β_3 της μεταβλητής «Σκανδιναβικές χώρες»: το ύψος της ετήσιας σύνταξης είναι κατά 10.899,6€ μεγαλύτερο εάν το άτομο προέρχεται από χώρα της Σκανδιναβίας, σε σχέση με το να προέρχεται από άλλη περιοχή (όταν όλες οι άλλες μεταβλητές διατηρούνται σταθερές).
- Όσον αφορά τον συντελεστή β_4 της μεταβλητής «Απαίτηση γνώσης χρήσης ηλεκτρονικού υπολογιστή: το ύψος της ετήσιας σύνταξης αυξάνεται κατά 2.369,3€ εάν η τελευταία εργασία απαιτούσε γνώση ηλεκτρονικού υπολογιστή σε σχέση με

την περίπτωση να μην υπήρχε αυτή η απαίτηση (όταν όλες οι άλλες μεταβλητές διατηρούνται σταθερές).

Όπως έχει ήδη αναφερθεί το κριτήριο ένταξης των μεταβλητών στο υπόδειγμα είναι το επίπεδο σημαντικότητας α=0,05.Οπότε είναι σαφές, εκ των προτέρων, ότι όλες οι μεταβλητές θα είναι στατιστικά σημαντικές σε αυτό το επίπεδο σημαντικότητας. Ωστόσο, παρατηρείται από τον πίνακα 3.18 ότι όλες οι μεταβλητές είναι στατιστικά σημαντικές σε *κάθε* επίπεδο σημαντικότητας (αφού p-value=0,00).

Η συνολική σημαντικότητα του μοντέλου προκύπτει από τον πίνακα ANOVA. Οι υποθέσεις που γίνονται είναι οι εξής:

H_0: το μοντέλο δεν είναι στατιστικά σημαντικό

H_1: το μοντέλο είναι στατιστικά σημαντικό

Πίνακας 3.20 *Πίνακας ANOVA για συνολική μεταβλητότητα*

ANOVA	
F	1.303,58
p-value	0,00

Σε οποιοδήποτε επίπεδο σημαντικότητας, απορρίπτεται η μηδενική υπόθεση του ελέγχου. Αυτό σημαίνει ότι το μοντέλο είναι στατιστικά σημαντικό.

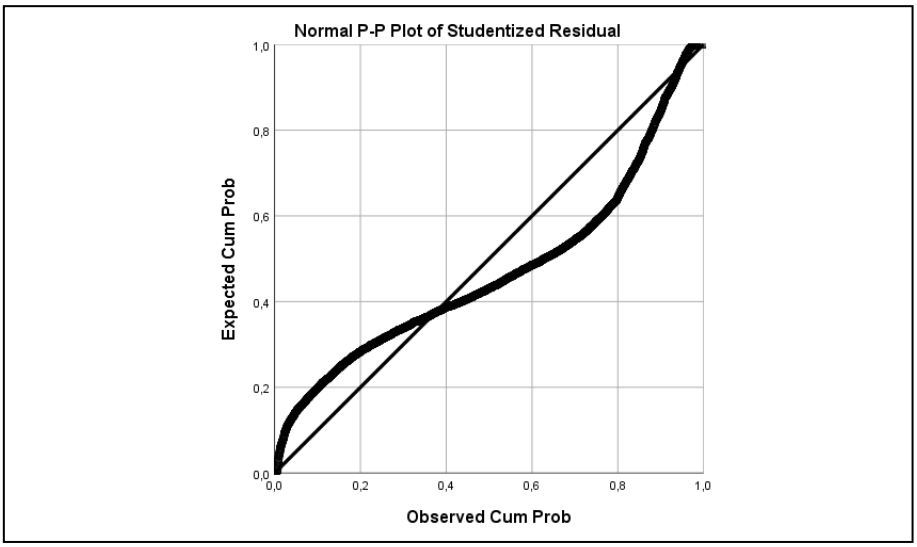
Επιπρόσθετα, από τον πίνακα 3.18 έχουμε ότι η τιμή του διορθωμένου συντελεστή προσδιορισμού είναι 0,488. Η τιμή αυτή σημαίνει ότι το 48,8% της διακύμανσης των συντάξεων ερμηνεύεται από τις ανεξάρτητες μεταβλητές του υποδείγματος.

Ωστόσο, προκειμένου να χρησιμοποιηθεί ορθά η Πολλαπλή Γραμμική Παλινδρόμηση πρέπει να ικανοποιούνται ορισμένες υποθέσεις. Στην συνέχεια, γίνεται έλεγχος αυτών των υποθέσεων, ώστε να διαπιστωθεί εάν τα αποτελέσματα της συγκεκριμένης μεθόδου είναι αξιόπιστα.

3.4.2 Έλεγχος για την υπόθεση της κανονικότητας

Η πρώτη υπόθεση που ελέγχεται είναι εκείνη της κανονικότητας των σφαλμάτων. Από το παρακάτω διάγραμμα μπορεί να διαπιστωθεί ότι τα σημεία δεν τείνουν στη διαγώνιο που σχηματίζεται από τις παρατηρούμενες αθροιστικές πιθανότητες με τις αναμενόμενες. Αυτό σημαίνει ότι τα σφάλματα δεν ακολουθούν κανονική κατανομή.

Διάγραμμα 3.7 *P-P Plot για έλεγχο κανονικότητας*



Προκειμένου να επιβεβαιωθεί το συμπέρασμα που προέκυψε από τον γραφικό έλεγχο θα γίνει κατάλληλος στατιστικός έλεγχος. Θα εκτελεστεί το κριτήριο Kolmogorov-Smirnov, το οποίο είναι ένας μη παραμετρικός έλεγχος που ελέγχει εάν τα δεδομένα έχουν καλή προσαρμογή στο μοντέλο. Οι υποθέσεις που γίνονται είναι οι εξής:

H_0: τα τυπικά σφάλματα ακολουθούν κανονική κατανομή

H_1: τα τυπικά σφάλματα δεν ακολουθούν κανονική κατανομή

Πίνακας **3. 21** *Έλεγχος Kolmogorov-Smirnov*

Kolmogorov - Smirnov test	
p-value	0,00

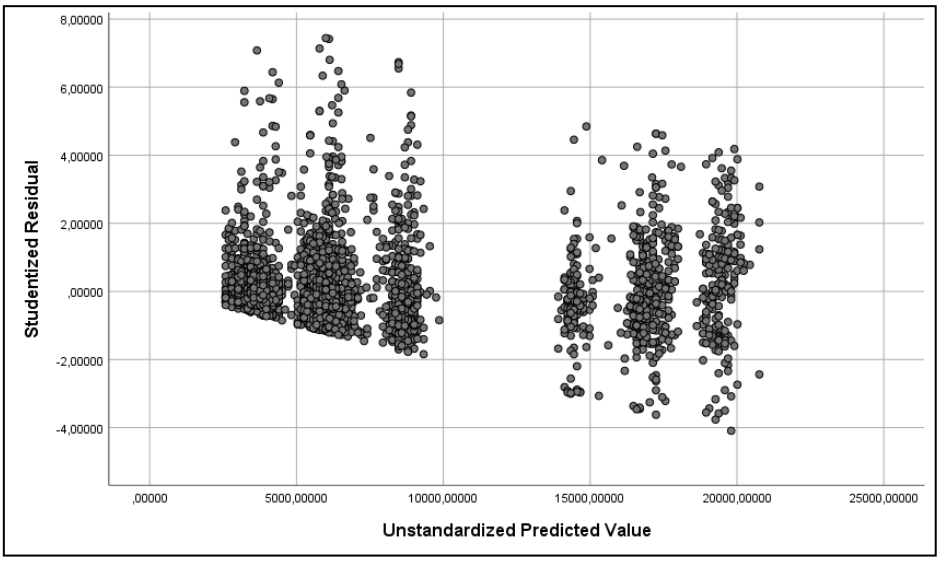
Σε οποιοδήποτε επίπεδο σημαντικότητας, απορρίπτεται η μηδενική υπόθεση του ελέγχου. Αυτό σημαίνει ότι τα τυπικά σφάλματα δεν ακολουθούν κανονική κατανομή. Οπότε η υπόθεση της κανονικότητας παραβιάζεται.

3.4.3 Έλεγχος για την υπόθεση της ομοσκεδαστικότητας

Στη συνέχεια γίνεται έλεγχος για να διαπιστωθεί εάν παραβιάζεται η υπόθεση της ομοσκεδαστικότητας των σφαλμάτων. Ένας οπτικός τρόπος είναι το διάγραμμα διασποράς του ζεύγους των s-τυποποιημένων καταλοίπων και των εκτιμώμενων τιμών της εξαρτημένης μεταβλητής, το οποίο διάγραμμα παρουσιάζεται παρακάτω.

Από το διάγραμμα φαίνεται ότι το νέφος σημείων παρουσιάζει συγκεκριμένο γεωγραφικό μοτίβο, γεγονός που μας προειδοποιεί ότι δεν υπάρχει ομοσκεδαστικότητα.

Διάγραμμα 3.8 *Διάγραμμα διασποράς καταλοίπων για έλεγχο ομοσκεδαστικότητας*



Για επιβεβαίωση των ανωτέρω συμπερασμάτων πραγματοποιείται ο έλεγχος Mann - Whitney U, ο οποίος είναι ένας μη παραμετρικός έλεγχος που χρησιμοποιείται όταν το δείγμα δεν κατανέμεται κανονικά. Το συγκεκριμένο κριτήριο αξιολογεί εάν οι διαφορές στις διακυμάνσεις των σφαλμάτων είναι στατιστικά σημαντικές. Οι υποθέσεις που γίνονται είναι οι εξής:

H_0: υπάρχει ομοσκεδαστικότητα

H_1: υπάρχει ετεροσκεδαστικότητα

Πίνακας 3.22 *Έλεγχος Mann-Whitney U*

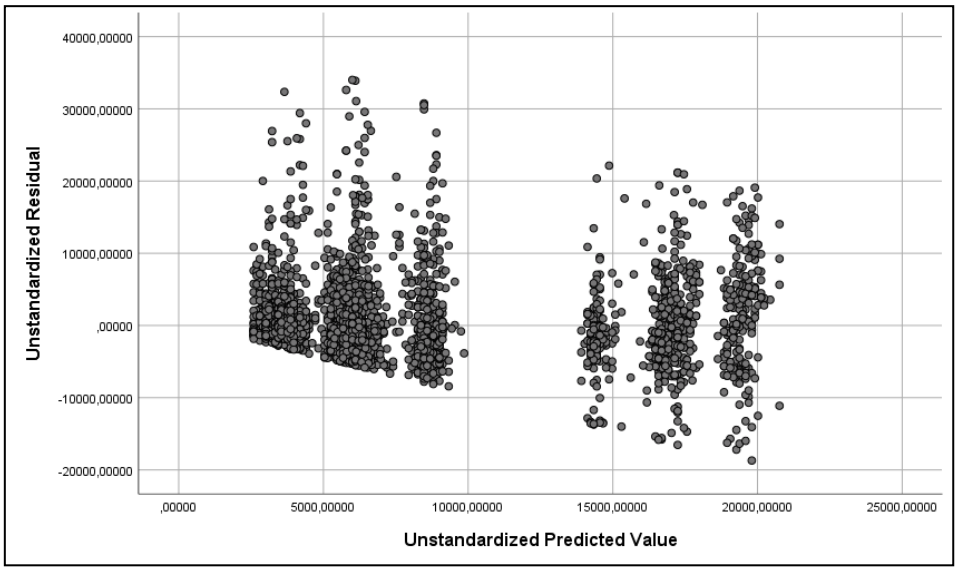
Mann-Whitney U test	
p-value	0,00

Σε οποιοδήποτε επίπεδο σημαντικότητας, απορρίπτεται η μηδενική υπόθεση του ελέγχου. Πιο συγκεκριμένα, η διακύμανση των τυπικών σφαλμάτων δεν διατηρείται σταθερή. Οπότε, η υπόθεση της ομοσκεδαστικότητας παραβιάζεται.

3.4.4 Έλεγχος για την υπόθεση της ανεξαρτησίας

Ο επόμενος έλεγχος που πραγματοποιείται αφορά την υπόθεση της ανεξαρτησίας των σφαλμάτων. Από το διάγραμμα φαίνεται ότι τα ζεύγη των μη τυποποιημένων καταλοίπων έναντι των εκτιμώμενων τιμών της εξαρτημένης μεταβλητής δημιουργούν ένα συστηματικό μοτίβο από σημεία. Φαίνεται ότι η υπόθεση της ανεξαρτησίας των σφαλμάτων δεν ικανοποιείται.

Διάγραμμα 3.9 *Διάγραμμα διασποράς καταλοίπων για έλεγχο ανεξαρτησίας*



Ο έλεγχος που χρησιμοποιείται, ώστε να επιβεβαιωθεί το ανωτέρω συμπέρασμα, είναι ο Wald-Wolfowitz Runs. Το συγκεκριμένο κριτήριο αποτελεί έναν μη παραμετρικό έλεγχο, ο οποίος ελέγχει την τυχαιότητα σε μια σειρά αριθμών, στην προκειμένη των καταλοίπων. Οι υποθέσεις που γίνονται είναι οι εξής:

H_0: τα τυπικά σφάλματα είναι ανεξάρτητα

H_1: τα τυπικά σφάλματα δεν είναι ανεξάρτητα

Πίνακας 3.23 *Έλεγχος Wald-Wolfowitz Runs*

Runs test	
p-value	0,00

Σε οποιοδήποτε επίπεδο σημαντικότητας, απορρίπτεται η μηδενική υπόθεση του ελέγχου. Από τον έλεγχο προκύπτει ότι τα σφάλματα δεν είναι ανεξάρτητα. Άρα, η υπόθεση της ανεξαρτησίας παραβιάζεται.

3.4.5 Έλεγχος για την υπόθεση της πολυσυγγραμμικότητας

Η υπόθεση για την ύπαρξη πολυσυγγραμμικότητας ελέγχει τον βαθμό συσχέτισης των ερμηνευτικών μεταβλητών. Θα χρησιμοποιηθεί ο δείκτης VIF (Variance Inflation Factor), ο οποίος εκφράζει πόσο πιο μεγάλη είναι η διακύμανση του εκτιμώμενου συντελεστή της μεταβλητής, αν συγκριθεί με την περίπτωση απουσίας πολυσυγγραμμικότητας (VIF=1). Μεγάλες τιμές του VIF αναδεικνύουν έντονο πρόβλημα πολυσυγγραμμικότητας, ενώ ικανοποιητικές θεωρούνται οι τιμές μικρότερες του 2.

Πίνακας 3.24 *Δείκτης VIF*

Μεταβλητές	VIF
Χρόνια εκπαίδευσης	1,205
Φύλο	1,005
Σκανδιναβικές χώρες	1,114
Απαίτηση χρήσης ηλ. υπολογιστή	1,325

Οι τιμές του δείκτη VIF για όλες τις ερμηνευτικές μεταβλητές είναι αρκετά μικρές (VIF<2), οπότε δεν υπάρχουν υψηλές συσχετίσεις. Αυτό σημαίνει ότι η υπόθεση της μη ύπαρξης πολυσυγγραμμικότητας δεν παραβιάζεται.

3.4.6 Αξιολόγηση Πολλαπλού Γραμμικού Υποδείγματος

Στην ανάλυση που προηγήθηκε, είδαμε ότι ο διορθωμένος συντελεστής προσδιορισμού, ο οποίος υπολογίζει το πόσο καλά οι ανεξάρτητες μεταβλητές ερμηνεύουν την εξαρτημένη, είναι αρκετά χαμηλός ($\bar{R}^2 = 0{,}488$). Αυτό σημαίνει ότι η ποιότητα προσαρμογής της Γραμμικής Παλινδρόμησης στα δεδομένα είναι μέτρια.

Επιπρόσθετα, παρουσιάστηκαν οι έλεγχοι για τις υποθέσεις που απαιτούνται για να είναι αξιόπιστα τα αποτελέσματα της Πολλαπλής Γραμμικής Παλινδρόμησης. Η υπόθεση της μη ύπαρξης πολυσυγγραμμικότητας ικανοποιείται, ωστόσο οι υποθέσεις της κανονικότητας, της ομοσκεδαστικότητας και της ανεξαρτησίας των σφαλμάτων παραβιάζονται. Γεγονός που καθιστά την Γραμμική Παλινδρόμηση μη κατάλληλη μέθοδο για τα συγκεκριμένα δεδομένα.

3.5 Παλινδρόμηση Πεμπτημορίων

Στη συνέχεια, η ανάλυση επικεντρώνεται σε μια δεύτερη μέθοδο, την Παλινδρόμηση Πεμπτημορίων. Όπως έχει αναφερθεί, η συγκεκριμένη μέθοδος δεν απαιτεί συγκεκριμένες υποθέσεις προκειμένου να χρησιμοποιηθεί ορθά και να είναι αξιόπιστες οι ερμηνείες.

Το υπόδειγμα που θα χρησιμοποιηθεί περιλαμβάνει τις ερμηνευτικές μεταβλητές που κρίθηκαν στατιστικά σημαντικές από τη μέθοδο Forward Selection που πραγματοποιήθηκε στο πλαίσιο της Πολλαπλής Γραμμικής Παλινδρόμησης. Πιο συγκεκριμένα, οι μεταβλητές αυτές είναι οι εξής: «Χρόνια εκπαίδευσης», «Φύλο», «Σκανδιναβικές χώρες», «Απαίτηση γνώσης χρήσης ηλεκτρονικού υπολογιστή». Η παρούσα εργασία είναι μία διερευνητική ανάλυση για τις ιδιότητες των μεθόδων και όχι μία απόπειρα εξέτασης των συντάξεων.

3.5.1 Μοντέλο Παλινδρόμησης Πεμπτημορίων

Η γενική μορφή του μοντέλου είναι η εξής:

$$Ύψος\ σύνταξης\ (Y) = b_0(\tau) + b_1(\tau)\ X_1 + \gamma_1(\tau)\ D_1 + +\gamma_3(\tau)\ D_3 + \gamma_{12}(\tau)\ D_{12}$$

Η ανάλυση θα πραγματοποιηθεί για τα ποσοστιαία σημεία τ=0,10 , τ=0,25 , τ=0,50 , τ=0,75 και τ=0,90. Το ποσοστιαίο σημείο τ=0,50 αναφέρεται στη διάμεσο των δεδομένων, ενώ τα υπόλοιπα σε ανώτερο ή κατώτερο επίπεδο της εξαρτημένης μεταβλητής.

Ακολουθεί ο πίνακας με το σύνολο των μοντέλων που προκύπτουν από την εφαρμογή της Παλινδρόμησης Πεμπτημορίων για τα ανωτέρω ποσοστιαία σημεία.

Πίνακας 3.25 *Μοντέλα Παλινδρόμησης Πεμπτημορίων*

	Μεταβλητές	**Γραμμική Παλινδρόμηση**	**Q 0,10**	**Q 0,25**	**Q 0,50**	**Q 0,75**	**Q 0,90**
D_3	**Σκανδιναβικές χώρες**	10.899,6* (188,8)	8.009,5* (92,7)	9.279,4* (97,1)	11.225,4* (112,9)	13.395,0* (186,3)	13.675,0* (396,6)
D_{12}	**Απαίτηση χρήσης ηλ.υπολογιστή**	2.369,3* (162,8)	639,5* (79,9)	1.004,5* (83,7)	1.439,7* (97,3)	3.030,0* (160,6)	5.640,0* (341,8)
D_1	**Φύλο**	-2.241,1* (124,7)	-611,0* (61,2)	-930,1* (64,1)	-1.470,1* (74,5)	-2.880,0* (123,0)	-4.800,0* (18,3)
X_1	**Χρόνια εκπαίδευσης**	106,7* (18,4)	15,0 (9,0)	86,8* (9,5)	96,0* (11,0)	37,5* (18,2)	0,0 (38,7)
b_0	**Σταθερός όρος**	4.825,9* (209,7)	10.089,5* (145,0)	11.710.0* (151,9)	14.945,3* (176,5)	20.295,0* (291,4)	25.315,0* (620,3)
	Pseudo $\bar{R}^2$	(0,488)	0,128	0,225	0,312	0,402	0,399
	MAE	2.873,5	3.890,1	3.124,6	2.732,3	3.378,3	5.617,0

ix Παρουσιάζονται οι εκτιμήσεις των συντελεστών και σε παρένθεση τα τυπικά σφάλματα των μοντέλων. Στην γραμμική παλινδρόμηση το Pseudo R^2 αναφέρεται στο $\bar{R}^2$ που υπολογίστηκε στο προηγούμενο κεφάλαιο. Οι στατιστικά σημαντικές εκτιμήσεις συνοδεύονται με «» σε επίπεδο σημαντικότητας α=0,05.*

Αρχικά, από τον πίνακα παρατηρείται ότι οι θετικές και αρνητικές σχέσεις, που είχαν προκύψει από την Γραμμική Παλινδρόμηση, μεταξύ των ερμηνευτικών μεταβλητών και της ανεξάρτητης παραμένουν ίδιες. Πιο συγκεκριμένα, υπάρχει θετική σχέση των χρόνων εκπαίδευσης, της καταγωγής από Σκανδιναβική χώρα και της απαίτησης γνώσης χρήσης ηλεκτρονικού υπολογιστή σε σχέση με το ύψος της σύνταξης. Ενώ, αρνητική σχέση φαίνεται να υπάρχει μεταξύ γυναικών και σύνταξης.

Στη συνέχεια, παρατηρείται ότι οι περισσότερες εκτιμήσεις είναι στατιστικά σημαντικές (σε επίπεδο σημαντικότητας α=0,05). Η μεταβλητή «Χρόνια εκπαίδευσης» δεν είναι στατιστικά σημαντική στα ποσοστιαία σημεία τ=0,10 και τ=0,90. Αυτό σημαίνει ότι τα πολύ χαμηλά και πολύ υψηλά επίπεδα σύνταξης δεν επηρεάζονται από το πόσα χρόνια έχει εκπαιδευτεί ένα άτομο.

Ενδεικτικά θα δοθούν οι ερμηνείες των συντελεστών στα ποσοστιαία σημεία τ=0,25 και τ=0,75:

- Ποσοστιαίο σημείο τ=0,25:
  - Όσον αφορά τον σταθερό όρο: το ύψος της ετήσιας σύνταξης διαμορφώνεται στα 11.710,0€ στο 0,25 ποσοστιαίο σημείο, εφόσον το άτομο είναι άνδρας με μηδενική εκπαίδευση, προερχόμενος από χώρα που δεν υπάγεται στη Σκανδιναβία και η τελευταία του εργασία πριν τη συνταξιοδότηση δεν απαιτούσε γνώσεις ηλεκτρονικού υπολογιστή.
  - Όσον αφορά τον συντελεστή β_1 της μεταβλητής «Χρόνια εκπαίδευσης»: για κάθε έναν χρόνο εκπαίδευσης, το ετήσιο ποσό σύνταξης αυξάνεται κατά 86,8€ στο 0,25 ποσοστιαίο σημείο (όταν όλες οι άλλες μεταβλητές διατηρούνται σταθερές).
  - Όσον αφορά τον συντελεστή β_2 της μεταβλητής «Φύλο»: το ύψος της ετήσιας σύνταξης είναι κατά 930,1€ μικρότερο εάν πρόκειται για γυναίκα σε σχέση με την περίπτωση του άνδρα, στο 0,25 ποσοστιαίο σημείο (όταν όλες οι άλλες μεταβλητές διατηρούνται σταθερές).
  - Όσον αφορά τον συντελεστή β_3 της μεταβλητής «Σκανδιναβικές χώρες»: στο 0,25 ποσοστιαίο σημείο, το ύψος της ετήσιας σύνταξης είναι κατά 9.279,4€ μεγαλύτερο εάν το άτομο προέρχεται από χώρα της Σκανδιναβίας, σε σχέση με το να προέρχεται από άλλη περιοχή (όταν όλες οι άλλες μεταβλητές διατηρούνται σταθερές).
  - Όσον αφορά τον συντελεστή β_4 της μεταβλητής «Απαίτηση γνώσης χρήσης ηλεκτρονικού υπολογιστή»: το ύψος της ετήσιας σύνταξης, στο 0,25 ποσοστιαίο σημείο, αυξάνεται κατά 1.004,5€ εάν η τελευταία εργασία απαιτούσε γνώση ηλεκτρονικού υπολογιστή σε σχέση με την περίπτωση να μην υπήρχε αυτή η απαίτηση (όταν όλες οι άλλες μεταβλητές διατηρούνται σταθερές).
- Ποσοστιαίο σημείο τ=0,75:
  - Όσον αφορά τον σταθερό όρο: το ύψος της ετήσιας σύνταξης διαμορφώνεται στα 20.295,0€ στο 0,75 ποσοστιαίο σημείο, εφόσον το άτομο είναι άνδρας με μηδενική εκπαίδευση, προερχόμενος από χώρα που δεν υπάγεται στη Σκανδιναβία και η τελευταία του εργασία πριν τη συνταξιοδότηση δεν απαιτούσε γνώσεις ηλεκτρονικού υπολογιστή.
  - Όσον αφορά τον συντελεστή β_1 της μεταβλητής «Χρόνια εκπαίδευσης»: για κάθε έναν χρόνο εκπαίδευσης, το ετήσιο ποσό σύνταξης αυξάνεται κατά 37,5€ στο 0,75 ποσοστιαίο σημείο (όταν όλες οι άλλες μεταβλητές διατηρούνται σταθερές).
  - Όσον αφορά τον συντελεστή β_2 της μεταβλητής «Φύλο»: το ύψος της ετήσιας σύνταξης είναι κατά 2.880,0€ μικρότερο εάν πρόκειται για γυναίκα σε σχέση με την περίπτωση του άνδρα, στο 0,75 ποσοστιαίο σημείο (όταν όλες οι άλλες μεταβλητές διατηρούνται σταθερές).

- Όσον αφορά τον συντελεστή β_3 της μεταβλητής «Σκανδιναβικές χώρες»: στο 0,75 ποσοστιαίο σημείο, το ύψος της ετήσιας σύνταξης είναι κατά 13.395,0€ μεγαλύτερο εάν το άτομο προέρχεται από χώρα της Σκανδιναβίας, σε σχέση με το να προέρχεται από άλλη περιοχή (όταν όλες οι άλλες μεταβλητές διατηρούνται σταθερές).
- Όσον αφορά τον συντελεστή β_4 της μεταβλητής «Απαίτηση γνώσης χρήσης ηλεκτρονικού υπολογιστή»: το ύψος της ετήσιας σύνταξης, στο 0,75 ποσοστιαίο σημείο, αυξάνεται κατά 3.030,0€ εάν η τελευταία εργασία απαιτούσε γνώση ηλεκτρονικού υπολογιστή σε σχέση με την περίπτωση να μην υπήρχε αυτή η απαίτηση (όταν όλες οι άλλες μεταβλητές διατηρούνται σταθερές).

Στον πίνακα 3.24, εκτός από τις εκτιμήσεις των συντελεστών, παρουσιάζονται και τα ψευδο-R^2 του μοντέλου σε κάθε ποσοστιαίο σημείο. Το μεγαλύτερο ψευδο-R^2 διακρίνεται στο τ=0,75, όπου ψευδο-R^2=0,402. Αυτό σημαίνει ότι το συγκεκριμένο ποσοστιαίο σημείο έχει την καλύτερη προσαρμογή των δεδομένων, δηλαδή ερμηνεύει καλύτερα την διακύμανση της εξαρτημένης μεταβλητής «Ετήσιο ποσό σύνταξης». Αντίστοιχα υψηλό είναι και το ψευδο-R^2 στο τ=0,90 (ψευδο-R^2=0,399) σε σχέση με τα μικρότερα ποσοστιαία σημεία. Αυτό σημαίνει ότι το μοντέλο φαίνεται να αποδίδει καλύτερα στα υψηλότερα ποσοστιαία σημεία.

Ο διορθωμένος συντελεστής προσδιορισμού που υπολογίστηκε από την Πολλαπλή Γραμμική Παλινδρόμηση είναι 0,488 - μεγαλύτερος από τα ψευδο-R^2 σε όλα τα ποσοστιαία σημεία της Παλινδρόμησης Πεμπτημορίων. Ωστόσο, δεν μπορεί να γίνει άμεση σύγκριση μεταξύ τους, διότι το ψευδο-R^2 αποτυπώνει διαφορετικές πτυχές της απόδοσης του μοντέλου. Πιο συγκεκριμένα, παρέχει πληροφορίες για την ερμηνευτική ικανότητα του μοντέλου σε συγκεκριμένα ποσοστιμόρια.

Ένα μέτρο που μπορεί να συγκρίνει τις δύο μεθόδους είναι το Μέσο Απόλυτο Σφάλμα (MAE – Mean Absolute Error), το οποίο υπολογίζει τη μέση τιμή της απόλυτης διαφοράς μεταξύ των παρατηρούμενων και των προβλεπόμενων τιμών. Μια μικρή τιμή του MAE υποδηλώνει καλύτερη προγνωστική ακρίβεια. Η μικρότερη τιμή στα υπό εξέταση μοντέλα φαίνεται ότι υπάρχει στο τ=0,50 της Παλινδρόμησης Πεμπτημορίων (όπου MAE=2.732,2), ενώ και το αντίστοιχο της Γραμμικής Παλινδρόμησης δεν είναι πολύ μεγαλύτερο (MAE=2.873,5).

3.5.2 Μοντέλα Παλινδρόμησης με Λογαριθμικό Μετασχηματισμό

Προκειμένου να ενισχυθεί η ερμηνεία των αποτελεσμάτων των δύο μεθόδων παρουσιάζονται παρακάτω τα υποδείγματα με λογαριθμικούς μετασχηματισμούς. Η εξαρτημένη μεταβλητή «Ετήσιο ποσό σύνταξης» μετασχηματίζεται χρησιμοποιώντας τον φυσικό λογάριθμο (ln). Οπότε, οι εκτιμήσεις των συντελεστών αντιπροσωπεύουν την ποσοστιαία μεταβολή της εξαρτημένης μεταβλητής αν μεταβληθεί κατά μία μονάδα η ερμηνευτική μεταβλητή. Με αυτόν τον τρόπο είναι εφικτή η σύγκριση των ποσοστιαίων μεταβολών μεταξύ διαφορετικών επιπέδων της μεταβλητής απόκρισης.

Η γενική μορφή του μοντέλου Γραμμικής Παλινδρόμησης είναι η εξής:

$$\ln(Y) = b_0 + b_1\, X_1 + \gamma_1\, D_1 + +\gamma_3\, D_3 + \gamma_{12}\, D_{12}$$

Ενώ η αντίστοιχη μορφή της Παλινδρόμησης Πεμπτημορίων:

$$\ln(Y) = b_0(\tau) + b_1(\tau)\, X_1 + \gamma_1(\tau)\, D_1 + +\gamma_3(\tau)\, D_3 + \gamma_{12}(\tau)\, D_{12}$$

Παρακάτω παρουσιάζεται ο πίνακας με τα αποτελέσματα των δύο μεθόδων.

Πίνακας 3.26 *Μοντέλα Παλινδρόμησης με Φυσικό Λογάριθμο*

	Μεταβλητές	**Γραμμική Παλινδρόμηση**	**Q 0,10**	**Q 0,25**	**Q 0,50**	**Q 0,75**	**Q 0,90**
D_3	**Σκανδιναβικές χώρες**	1,167*	1,522*	1,436*	1,261*	1,114*	0,907*
		(0,025)	(0,043)	(0,032)	(0,023)	(0,030)	(0,045)
D_{12}	**Απαίτηση χρήσης ηλ.υπολογιστή**	0,310*	0,247*	0,216*	0,242*	0,353*	0,454*
		(0,022)	(0,037)	(0,28)	(0,020)	(0,026)	(0,039)
D_1	**Φύλο**	-0,342*	-0,252*	-0,285*	-0,296*	-0,411*	-0,467*
		(0,016)	(0,029)	(0,021)	(0,015)	(0,020)	(0,030)
X_1	**Χρόνια εκπαίδευσης**	0,013*	0,010*	0,032*	0,022*	0,006*	0,001
		(0,002)	(0,004)	(0,003)	(0,002)	(0,003)	(0,004)
b_0	**Σταθερός όρος**	8,311*	9,085*	9,102*	9,418*	9,799*	10,103*
		(0,028)	(0,068)	(0,050)	(0,020)	(0,046)	(0,070)
	Pseudo $\bar{R}^2$	(0,414)	0,144	0,215	0,283	0,337	0,290
	MAE	0,4452	0,7732	0,5361	0,4425	0,5244	0,7754

x Παρουσιάζονται οι εκτιμήσεις των συντελεστών και σε παρένθεση τα τυπικά σφάλματα των μοντέλων. Στην γραμμική παλινδρόμηση το Pseudo R^2 αναφέρεται στο $\bar{R}^2$ που υπολογίστηκε στο προηγούμενο κεφάλαιο. Οι στατιστικά σημαντικές εκτιμήσεις συνοδεύονται με «» σε επίπεδο σημαντικότητας $\alpha=0,05$. Οι τιμές αναφέρονται σε μεταβολές του λογαρίθμου της εξαρτημένης μεταβλητής*

Ο λογαριθμικός μετασχηματισμός φαίνεται ότι μείωσε κατά λίγο τον διορθωμένο συντελεστή προσδιορισμού του μοντέλου της Γραμμικής Παλινδρόμησης (από 0,488 σε 0,414). Ωστόσο, στην Παλινδρόμηση Πεμπτημορίων παρατηρείται ότι στα χαμηλά επίπεδα συντάξεων το ψευδο-R^2 αυξήθηκε κατά λίγο (στο τ=0,10 από 0,128 σε 0,144 και στο τ=0,25 από 0,225 σε 0,215), ενώ στα υψηλότερα μειώθηκε (στο τ=0,75 από 0,402 σε 0,337 και στο τ=0,90 από 0,399 σε 0,290).

Το Μέσο Απόλυτο Σφάλμα της Γραμμικής Παλινδρόμησης είναι 0,445, ενώ της Παλινδρόμησης Πεμπτημορίων κυμαίνεται από 0,443 έως 0,775, ανάλογα το ποσοστιαίο

σημείο. Η ερμηνευτική ικανότητα των δύο μοντέλων φαίνεται ότι είναι στα ίδια επίπεδα ικανοποίησης.

Οι εκτιμήσεις των συντελεστών που προέκυψαν αφορούν την μεταβολή του φυσικού λογαρίθμου της εξαρτημένης μεταβλητής. Για παράδειγμα, αν αυξηθεί κατά μία μονάδα η μεταβλητή «Χρόνια εκπαίδευσης» τότε θα αυξηθεί κατά 0,013 μονάδες το ln(Y), δεδομένου ότι διατηρούνται οι υπόλοιπες ερμηνευτικές μεταβλητές σταθερές.

Παρακάτω παρουσιάζονται τα αποτελέσματα των υποδειγμάτων ύστερα από μετατροπή των τιμών, έτσι ώστε να αναφέρονται στις μεταβολές της εξαρτημένης μεταβλητής και όχι του λογαρίθμου της.

Πίνακας 3.27 *Μοντέλα Παλινδρόμησης με ποσοστιαίες μεταβολές της Y*

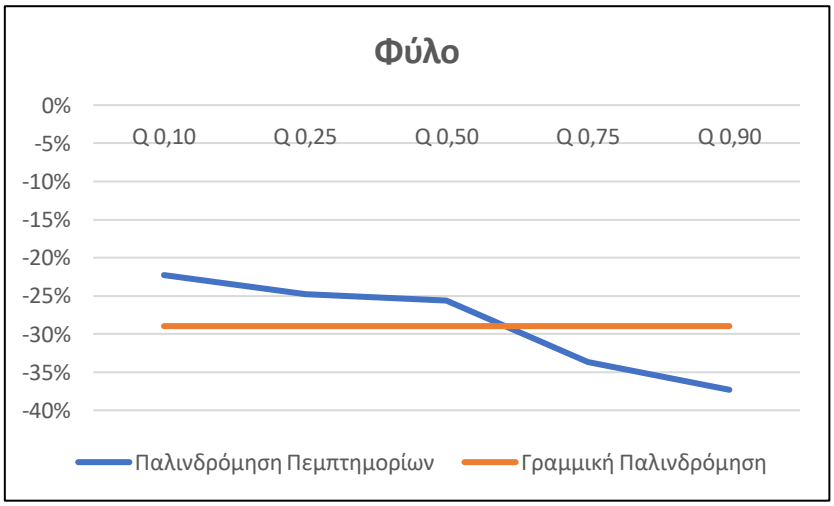
	Μεταβλητές	Γραμμική Παλινδρόμηση	Q 0,10	Q 0,25	Q 0,50	Q 0,75	Q 0,90
D_3	**Σκανδιναβικές χώρες**	2,212	3,581	3,204	2,529	2,047	1,477
D_{12}	**Απαίτηση χρήσης ηλ.υπολογιστή**	0,363	0,280	0,241	0,274	0,423	0,575
D_1	**Φύλο**	-0,290	-0,223	-0,248	-0,256	-0,337	-0,373
X_1	**Χρόνια εκπαίδευσης**	0,013	0,010	0,033	0,022	0,006	0,001
b_0	**Σταθερός όρος**	4067,4	8821,0	8972,2	12306,9	18014,7	24415,1

xi Παρουσιάζονται οι εκτιμήσεις των συντελεστών. Οι τιμές αναφέρονται σε ποσοστιαίες μεταβολές της εξαρτημένης μεταβλητής, ύστερα από τον λογαριθμικό μετασχηματισμό.

Στην συνέχεια παρατίθενται οι ερμηνείες των συντελεστών σε όλο το εύρος της κατανομής της εξαρτημένης μεταβλητής «Ετήσιο ποσό σύνταξης».

Όσον αφορά τη μεταβλητή «Φύλο»:

Διάγραμμα 3.10 *Εκτιμήσεις μεταβλητής «Φύλο» σε όλο το εύρος της κατανομής*

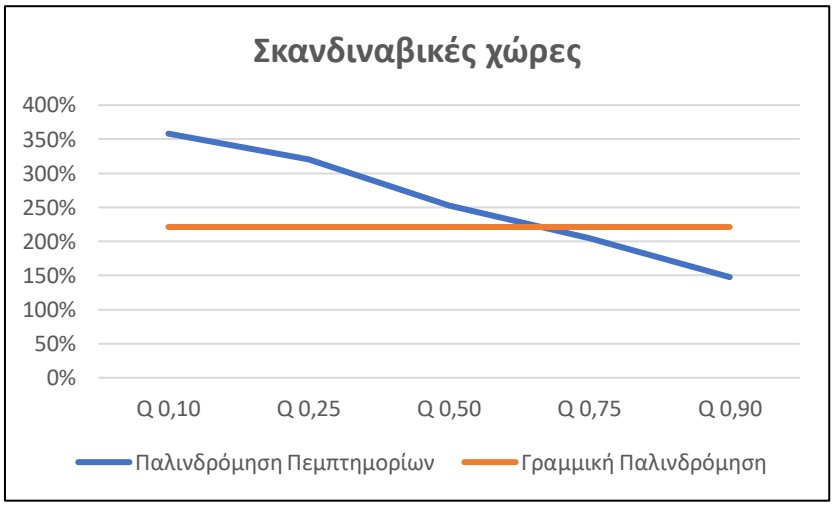


Από την Γραμμική Παλινδρόμηση προέκυψε ότι το ύψος της ετήσιας σύνταξης είναι κατά 29% μικρότερο εάν πρόκειται για γυναίκα σε σχέση με την περίπτωση του άνδρα (όταν όλες οι άλλες μεταβλητές διατηρούνται σταθερές). Η

Παλινδρόμηση Πεμπτημορίων δείχνει ότι στα χαμηλότερα ποσά σύνταξης η διαφορά μεταξύ ανδρών και γυναικών είναι μικρότερη (στο τ=0,10 η διαφορά είναι 22%), ενώ στα υψηλότερα ποσά αυξάνεται (στο τ=0,90 η διαφορά ξεπερνάει το 35%) – δεδομένου ότι όλες οι άλλες μεταβλητές διατηρούνται σταθερές. Πιο συγκεκριμένα, από το υπόδειγμα προέκυψε ότι οι γυναίκες λαμβάνουν κατά 22% μικρότερη σύνταξη εάν αναφερόμαστε σε χαμηλές συντάξεις, ενώ εάν αναφερόμαστε σε υψηλά επίπεδα σύνταξης λαμβάνουν έως και 35% μικρότερη σύνταξη σε σχέση με τους άνδρες.

Όσον αφορά τη μεταβλητή «Σκανδιναβικές χώρες»:

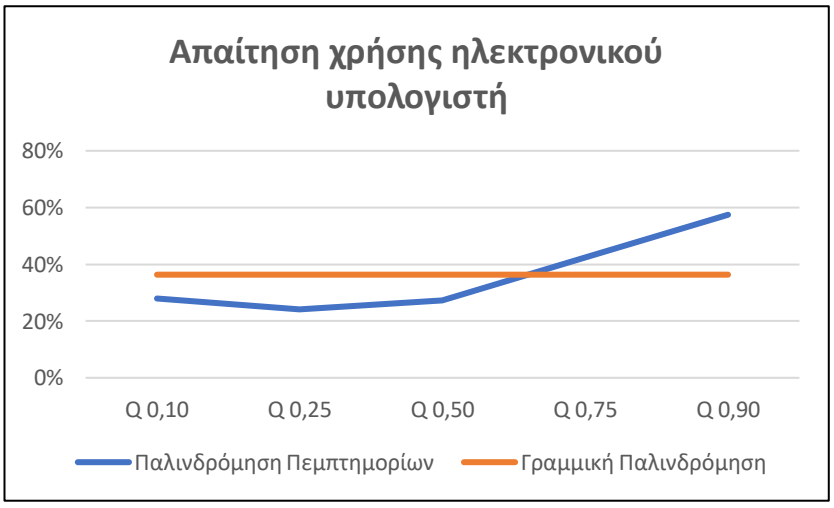
Διάγραμμα 3.11 *Εκτιμήσεις μεταβλητής «Σκανδιναβικές χώρες» σε όλο το εύρος της κατανομής*



Η Γραμμική Παλινδρόμηση υπολογίζει ότι σε όλο το εύρος των συντάξεων, η καταγωγή του ατόμου από Σκανδιναβική χώρα αυξάνει τη σύνταξη κατά 220% σε σχέση με τις άλλες χώρες. Στην Παλινδρόμηση Πεμπτημορίων φαίνεται ότι στα χαμηλά ποσά η ποσοστιαία μεταβολή πλησιάζει το 360% (στο τ=0,10 η διαφορά είναι 358%), ενώ στα υψηλότερα ποσά σύνταξης η μεταβολή φαίνεται να είναι γύρω στο 150% (στο τ=0,90 η διαφορά είναι 148%) – δεδομένου ότι όλες οι άλλες μεταβλητές διατηρούνται σταθερές.

Όσον αφορά τη μεταβλητή «Απαίτηση γνώσης χρήσης ηλεκτρονικού υπολογιστή»:

Διάγραμμα 3.12 *Εκτιμήσεις μεταβλητής «Απαίτηση χρήσης ηλ. υπολογιστή» σε όλο το εύρος της κατανομής*



Κατά τη Γραμμική Παλινδρόμηση, το ύψος της ετήσιας σύνταξης αυξάνεται κατά 36% εάν η τελευταία εργασία πριν την συνταξιοδότηση απαιτούσε χρήση ηλεκτρονικού υπολογιστή. Από την Παλινδρόμηση Πεμπτημορίων φαίνεται ότι στα χαμηλά ποσά συντάξεων δεν υπάρχει

τόσο μεγάλη μεταβολή της σύνταξης (στο τ=0,10 η διαφορά είναι 28% και στο τ=0,25 η διαφορά μειώνεται στο 24%), ενώ μετά το ποσοστιαίο σημείο τ=0,50 η μεταβολή αυξάνεται με αύξοντα ρυθμό και φτάνει έως και το 60% – δεδομένου ότι όλες οι άλλες μεταβλητές διατηρούνται σταθερές. Η διαφορά αυτή μπορεί να οφείλεται στο γεγονός ότι οι εργασίες που απαιτούν γνώση χρήσης ηλεκτρονικού υπολογιστή συνήθως αμείβονται καλύτερα.

3.5.3 Αξιολόγηση Υποδείγματος Παλινδρόμησης Πεμπτημορίων

Με την Παλινδρόμηση Πεμπτημορίων είναι εφικτή η διερεύνηση των παραγόντων που επηρεάζουν σε όλο το εύρος της εξαρτημένης μεταβλητής «Ετήσιο ύψος συντάξεων».

Οι ερμηνευτικές μεταβλητές που χρησιμοποιήθηκαν σε αυτή τη μέθοδο είναι εκείνες οι οποίες προέκυψαν στατιστικά σημαντικές από την Γραμμική Παλινδρόμηση. Αυτό είχε σαν αποτέλεσμα να μην είναι στατιστικά σημαντικές σε όλα τα ποσοστιαία σημεία που ελέγχθηκαν κατά την εφαρμογή της Παλινδρόμησης Πεμπτημορίων και το ψευδο-R^2 να μην έχει μεγάλη ερμηνευτική ικανότητα σε όλο το εύρος της κατανομής.

Με τον μετασχηματισμό της εξαρτημένης μεταβλητής σε φυσικό λογάριθμο είναι ευδιάκριτες οι ποσοστιαίες μεταβολές του ύψους των συντάξεων στο εύρος της κατανομής ανάλογα με την τιμή της εκάστοτε ερμηνευτικής μεταβλητής. Έτσι παρουσιάζονται καλύτερα οι διαφορές μεταξύ των επιπέδων των μεταβλητών στις χαμηλές από τις υψηλές συντάξεις.

Συμπεράσματα

Η Γραμμική και Παλινδρόμηση Πεμπτημορίων αποτελούν μεθόδους παλινδρόμησης που χρησιμοποιούνται για τη μοντελοποίηση της σχέσης μεταξύ μίας ή περισσότερων μεταβλητών πρόβλεψης και μίας μεταβλητής απόκρισης. Ωστόσο, υπάρχουν κρίσιμες διαφορές μεταξύ τους.

Καταρχάς, η βασική διαφοροποίηση τους έγκειται στο γεγονός ότι η Γραμμική εκτιμά τη μέση συμπεριφορά της εξαρτημένης μεταβλητής δεδομένων των τιμών των ερμηνευτικών μεταβλητών. Σε αντίθεση με αυτό, η Παλινδρόμηση Πεμπτημορίων περιγράφει ολόκληρη την κατανομή της μεταβλητής απόκρισης. Συγκρίνοντας τις εκτιμήσεις των παραμέτρων μεταξύ διαφορετικών ποσοστιαίων σημείων, η δεύτερη μέθοδος επιτρέπει την εξέταση του τρόπου με τον οποίο οι ερμηνευτικές μεταβλητές επηρεάζουν διαφορετικά μέρη της κατανομής.

Επιπρόσθετα, η Γραμμική Παλινδρόμηση απαιτεί ορισμένες προϋποθέσεις προκειμένου να εκτελείται ορθά, ενώ η Παλινδρόμηση Πεμπτημορίων δέχεται τις μη κανονικές κατανομές και επιτρέπει την ετεροσκεδαστικότητα, εφόσον εκτιμά διαφορετικά ποσοστιμόρια. Μία ακόμα διαφορά είναι ότι οι εκτιμητές της δεύτερης μεθόδου είναι πιο ανθεκτικοί ως προς τις ακραίες παρατηρήσεις της εξαρτημένης μεταβλητής.

Από τα ανωτέρω προκύπτει ότι η Παλινδρόμηση Πεμπτημορίων αποτελεί την πλέον χρήσιμη μέθοδο όταν αναλύονται δεδομένα με ακραίες τιμές, λοξές κατανομές ή όταν ο ερευνητής ενδιαφέρεται να κατανοήσει πώς οι προγνωστικοί παράγοντες επηρεάζουν διαφορετικά σημεία στην κατανομή της μεταβλητής απόκρισης αν η βασική σχέση δεν είναι γραμμική.

Το μοντέλο που ερευνήθηκε ενδεικτικά στην παρούσα εργασία κατέληξε ότι οι παράγοντες που επηρεάζουν το ετήσιο ύψος της σύνταξης είναι το φύλο, το πλήθος των χρόνων που έχει εκπαιδευτεί ένα άτομο, η καταγωγή από Σκανδιναβική χώρα και η απαίτηση γνώσης χρήσης ηλεκτρονικού υπολογιστή στην τελευταία εργασία πριν την συνταξιοδότηση. Η μεταβολή στο ύψος της σύνταξης είναι αρνητική εάν το άτομο είναι γυναίκα, ενώ εάν το άτομο κατάγεται από Σκανδιναβική χώρα ή στην τελευταία εργασία του απαιτούνταν γνώσεις ηλεκτρονικού υπολογιστή τότε το ύψος της σύνταξης μεταβάλλεται θετικά. Επίσης, όσο περισσότερα χρόνια εκπαίδευσης έχει το άτομο τόσο μεγαλύτερη σύνταξη λαμβάνει.

Κατά την ανάλυση της Γραμμικής Παλινδρόμησης διενεργήθηκαν οι έλεγχοι των υποθέσεων και προέκυψε ότι παραβιάζεται η κανονικότητα, η ομοσκεδαστικότητα και η ανεξαρτησία των σφαλμάτων. Το γεγονός αυτό καθιστά την συγκεκριμένη μέθοδο μη κατάλληλη για τα συγκεκριμένα δεδομένα και τα συμπεράσματά της δεν είναι αξιόπιστα.

Τα αποτελέσματα της Παλινδρόμησης Πεμπτημορίων είναι πιο ικανοποιητικά, εφόσον δεν απαιτείται η ύπαρξη προϋποθέσεων για την χρήση της. Ο δείκτης που χρησιμοποιείται για να ελεγχθεί πόσο καλά οι ανεξάρτητες μεταβλητές ερμηνεύουν την εξαρτημένη μεταβλητή έδειξε ότι υπάρχει καλύτερη ερμηνευτική ικανότητα του μοντέλου στα υψηλά ποσά συντάξεων.

Από τα υποδείγματα με λογαριθμικό μετασχηματισμό προέκυψε ότι η ποσοστιαία μεταβολή του ύψους της σύνταξης στην περίπτωση της Γραμμικής Παλινδρόμησης είναι μείωση της σύνταξης κατά 29% για τις γυναίκες, αύξηση κατά 220% σε άτομα από Σκανδιναβική χώρα και αύξηση κατά 36% σε εκείνους των οποίων η τελευταία εργασία απαιτούσε χρήση ηλεκτρονικού υπολογιστή. Επίσης, για κάθε επιπλέον χρόνο εκπαίδευσης το ετήσιο ποσό σύνταξης αυξάνεται κατά 1,3%.

Από την ανάλυση της Παλινδρόμησης Πεμπτημορίων φαίνεται ότι τα χρόνια εκπαίδευσης επηρεάζουν ελάχιστα τα υψηλά και τα πολύ χαμηλά στρώματα. Το φύλο επηρεάζει περισσότερο τα υψηλά επίπεδα σύνταξης, αφού οι γυναίκες τείνουν να λαμβάνουν έως και 35% μικρότερο ποσό σε εκείνα τα επίπεδα. Η καταγωγή από Σκανδιναβική χώρα ασκεί μεγαλύτερη επιρροή στα χαμηλά στρώματα, στα οποία η διαφορά από τις άλλες χώρες είναι περίπου 360%. Ακόμα, όσον αφορά την απαίτηση χρήσης ηλεκτρονικού υπολογιστή, φαίνεται ότι όσο αυξάνονται οι συντάξεις τόσο αυξάνεται και η διαφορά μεταξύ εκείνων που υπήρχε η απαίτηση στην τελευταία τους εργασία και εκείνων που δεν υπήρχε τέτοιου είδους απαίτηση.

Εν κατακλείδι, η Παλινδρόμηση Πεμπτημορίων αποδείχτηκε ότι είναι καταλληλότερη μέθοδος στις μεταβλητές που εξετάστηκαν, εφόσον δεν πληρούνται τα κριτήρια της Γραμμικής Παλινδρόμησης, Εκτός αυτού, με αυτήν την μέθοδο ήταν εφικτή η διερεύνηση της συμπεριφοράς του ετήσιου ποσού σύνταξης σε όλο το εύρος της κατανομής, ώστε να ερμηνευτούν οι μεταβολές που βρίσκονται στα υψηλότερα επίπεδα σύνταξης σε σχέση με τα χαμηλότερα.

Προτάσεις για περαιτέρω έρευνα

Στην παρούσα εργασία δεν εξαντλήθηκαν οι δυνατότητες μελέτης και ανάλυσης των στοιχείων. Θέματα που μπορούν να ερευνηθούν είναι τα εξής:

- Να γίνουν μετασχηματισμοί στην εξαρτημένη μεταβλητή, ώστε να διερευνηθεί εάν ικανοποιούνται οι υποθέσεις της Γραμμικής Παλινδρόμησης και να γίνει μετέπειτα σύγκριση με την Παλινδρόμηση Πεμπτημορίων
- Σύγκριση των ποσοστιαίων σημείων της Παλινδρόμησης Πεμπτημορίων, ώστε να διαπιστωθεί εάν είναι στατιστικά σημαντικές οι διαφορές μεταξύ των χαμηλών και των υψηλών επιπέδων της εξαρτημένης μεταβλητής
- Παρουσία αλληλεπιδράσεων των ερμηνευτικών μεταβλητών στο υπόδειγμα, ώστε να ελεγχθεί εάν μεταβάλλουν την μεταβλητή απόκρισης

Παραρτήματα

Π1 Κωδικοποιήσεις μεταβλητών SHARE

Στον παρακάτω πίνακα παρουσιάζονται όλες οι μεταβλητές, οι οποίες χρησιμοποιήθηκαν για τον σκοπό της ανάλυσης. Επιπρόσθετα, φαίνεται σε ποια βάση δεδομένων βρίσκονται, το όνομα τους και η περιγραφή τους.

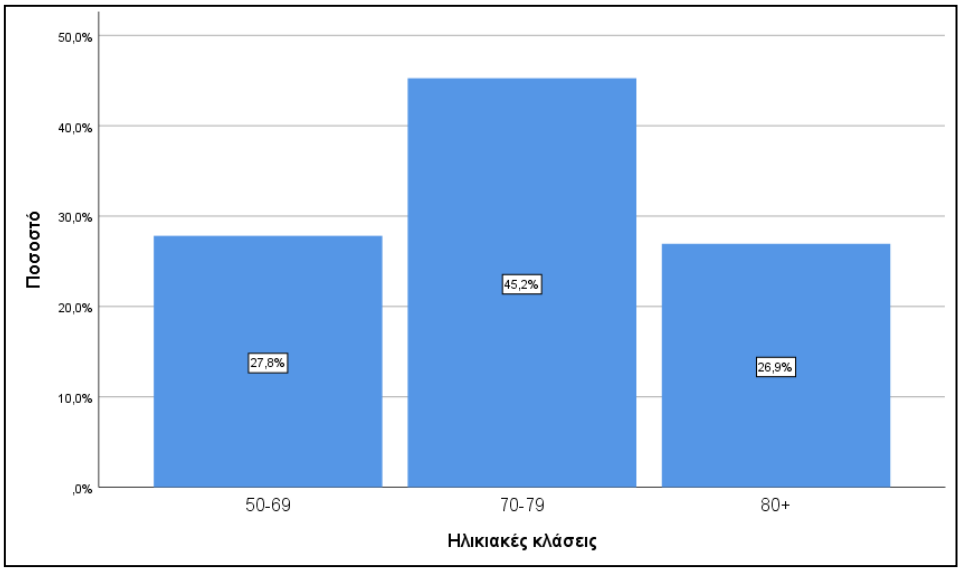
Πίνακας Π.1.1 *Πίνακας κωδικοποίησης μεταβλητών*

Μεταβλητή	Name	Dataset	Label
Ετήσιο ποσό σύνταξης	ypen1	sharew8_rel8-0-0_gv_imputations	Old age, early retirement, and survivor pensions
Φύλο	gender	sharew8_rel8-0-0_gv_imputations	Gender
Χώρα	country	sharew8_rel8-0-0_gv_imputations	Country identifier
Χρόνια εκπαίδευσης	yedu	sharew8_rel8-0-0_gv_imputations	Years of education
Επίπεδο εκπαίδευσης	isced	sharew8_rel8-0-0_gv_imputations	ISCED 1997 coding of education
Οικογενειακή κατάσταση	mstat	sharew8_rel8-0-0_gv_imputations	Marital status
Τρέχουσα εργασιακή κατάσταση	cjs	sharew8_rel8-0-0_gv_imputations	Current job situation
Παράλληλη εργασία	ep002_	sharew8_rel8-0-0_ep	Did nevertheless any paid work last four weeks
Υπάλληλος ή ελεύθερος επαγγελματίας	ep051_	sharew8_rel8-0-0_ep	Employee or a self employed in last job
Έτος συνταξιοδότησης	ep329_	sharew8_rel8-0-0_ep	Retirement year
Έτος γέννησης	dn003_	sharew8_rel8-0-0_dn	Year of birth
Απαίτηση χρήσης ηλεκτρονικού υπολογιστή	it002_	sharew8_rel8-0-0_it	Last job before retiring required using a computer

Π2 Διαγράμματα Περιγραφικής Ανάλυσης

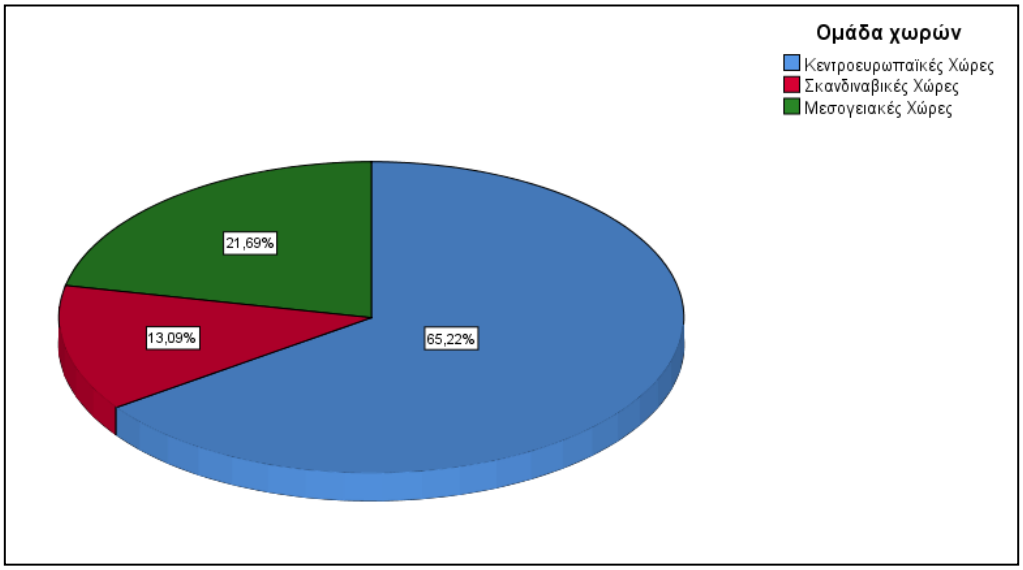
Στο παρακάτω διάγραμμα παρουσιάζονται τα ποσοστά που καταλαμβάνουν οι τρεις ηλικιακές ομάδες. Όπως έχει ήδη αναφερθεί, η κλάση «70-79» υπερέχει κατά πολύ, ενώ οι άλλες δύο φαίνεται και σχηματικά ότι είναι περίπου στο ίδιο επίπεδο.

Διάγραμμα Π.2.1 *Ραβδόγραμμα σχετικών συχνοτήτων για τις ηλικιακές ομάδες*



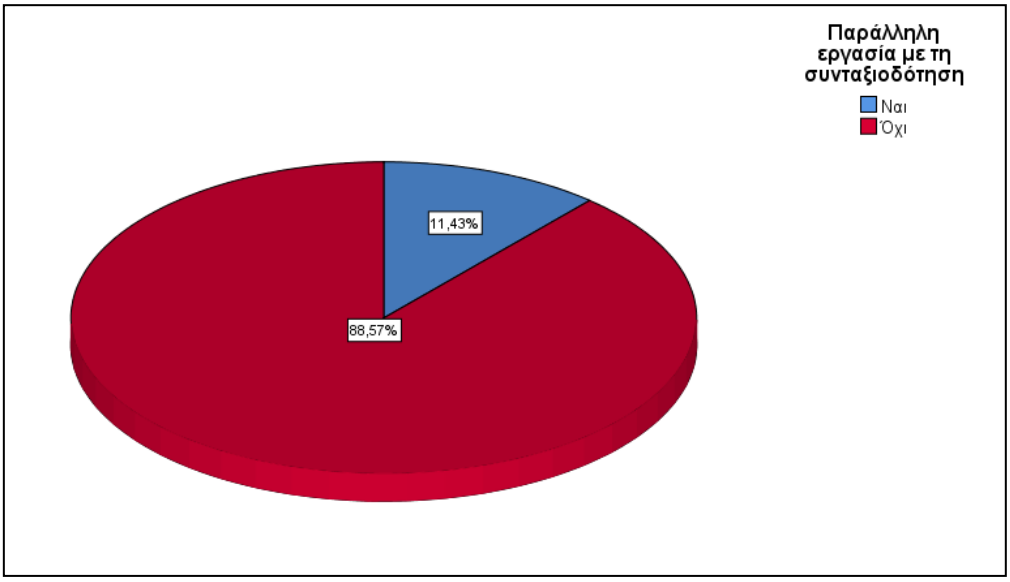
Το διάγραμμα πίτας επιβεβαιώνει ότι οι Κεντροευρωπαϊκές χώρες αποτελούν το μεγαλύτερο μέρος του δείγματος, ακολουθούν οι Μεσογειακές και το μικρότερο ποσοστό καταλαμβάνουν οι Σκανδιναβικές.

Διάγραμμα Π.2.2 *Διάγραμμα πίτας για τις ομάδες χωρών*



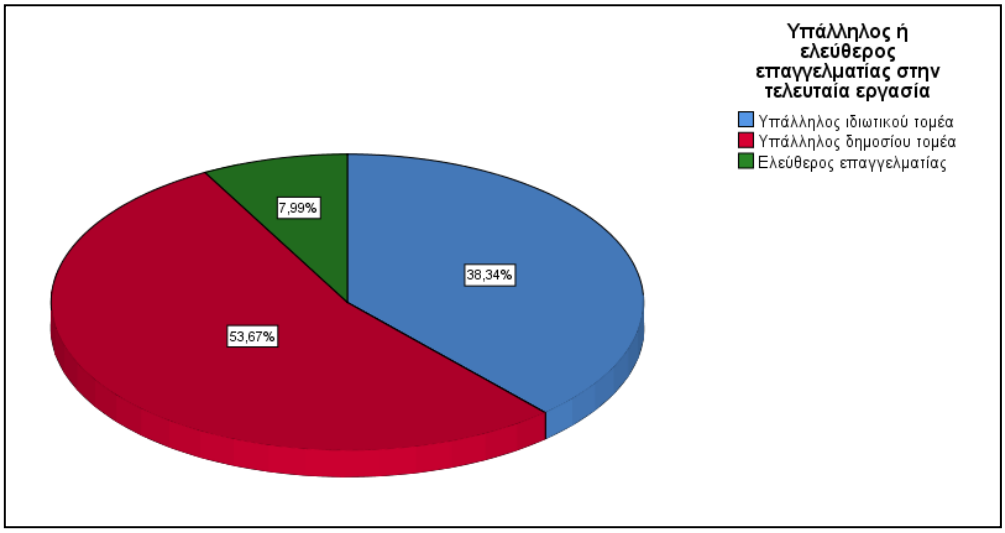
Στο παρακάτω διάγραμμα φαίνεται οπτικά ότι το μεγαλύτερο μέρος του δείγματος δεν συνεχίζει να εργάζεται μετά τη λήψη της σύνταξης.

Διάγραμμα Π.2. 3 *Διάγραμμα πίτας για τη μεταβλητή «Παράλληλη εργασία με τη συνταξιοδότηση»*



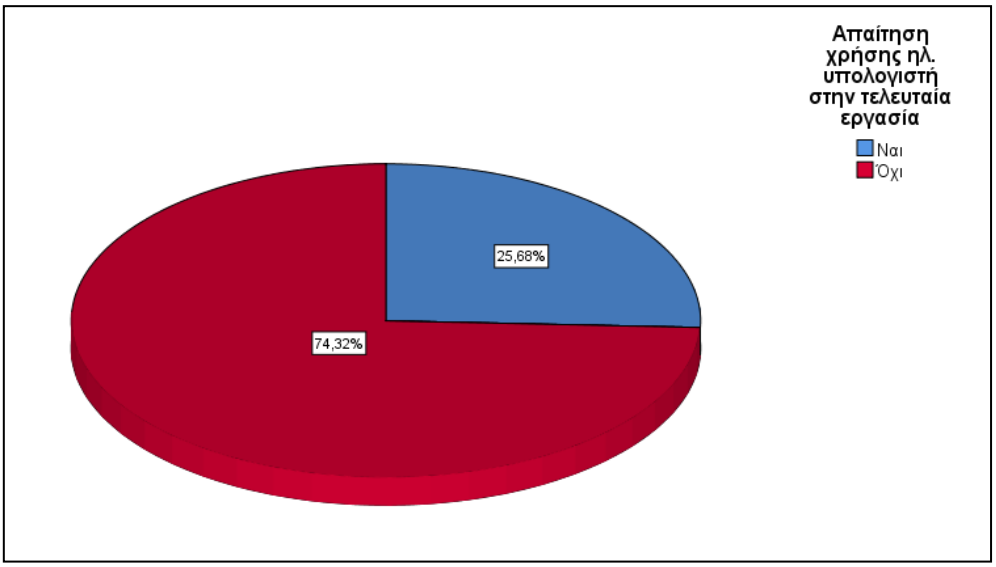
Στο διάγραμμα πίτας που ακολουθεί επιβεβαιώνεται οπτικά ότι το μεγαλύτερο μέρος του δείγματος εργαζόταν στον δημόσιο τομέα, ένα μικρότερο μέρος στον ιδιωτικό τομέα και ένα πολύ μικρό ποσοστό ήταν αυτοαπασχολούμενοι.

Διάγραμμα Π.2.4 *Διάγραμμα πίτας για τη μεταβλητή «Υπάλληλος ή ελεύθερος επαγγελματίας»*



Τέλος, παρατηρείται ότι περίπου το ένα τέταρτο του δείγματος στην τελευταία εργασία πριν την συνταξιοδότηση είχε την απαίτηση χρήσης ηλεκτρονικού υπολογιστή.

Διάγραμμα Π.2.5 *Διάγραμμα πίτας για τη μεταβλητή «Απαίτηση χρήσης ηλεκτρονικού υπολογιστή»*



Βιβλιογραφία

Ελληνική

Ζαφειρόπουλος, Κ. και Μυλωνάς Ν. (2018) Στατιστική με SPSS, Περιέχει Θεωρία Πιθανοτήτων. Θεσσαλονίκη: Τζιόλα.

Κατρακυλίδης, Κ., Κοντέος, Γ. και Σαριαννίδης, Ν. (2017) Εισαγωγή στη Σύγχρονη Οικονομετρία. Θεσσαλονίκη: Αλέξανδρος ΙΚΕ.

Ζαχαροπούλου, Χ. (2011) Στατιστική: μέθοδοι – εφαρμογές. Θεσσαλονίκη: Σοφία.

Κάτος, Α. (2004) Οικονομετρία: Θεωρία και Εφαρμογές. Θεσσαλονίκη: Ζυγός.

Συριόπουλος, Κ. και Φίλιππας, Δ. (2010) Οικονομετρικά Υποδείγματα και Εφαρμογές με το Eviews. Θεσσαλονίκη: Ανικούλα.

Γκανέτος, Η. (2007) Μικτά Μοντέλα και Παλινδρόμηση Ποσοστιαίων Σημείων (Μη εκδοθείσα διπλωματική εργασία). Πανεπιστήμιο Πειραιώς.

Κούτρας, Μ. (2020) Σημειώσεις μαθήματος «Ανάλυση παλινδρόμησης και ανάλυση διακύμανσης» (στο πλαίσιο του μεταπτυχιακού προγράμματος Εφαρμοσμένη Στατιστική του Πανεπιστημίου Πειραιώς)

Ξένη

Penda, I.A. (2017) 'The European National welfare states and the dissolution of the EU', Philosophy and Society, 29(2), p153-316.

UNESCO (2006[1997]) International Standard Classification of Education: ISCED 1997 (re-edition). Montreal: UNESCO Institute for Statistics.

Betti, G., Bettio, F. & Tinios, P. (2015) Unequal Ageing in Europe: Women's Independence and Pensions, New York: Palgrave Macmillan.

Nimon, K. & Oswald, F. (2013) 'Understanding the Results of Multiple Linear Regression: Beyond Standardized Regression Coefficients' Organizational Research Methods, 16(4), p650-674.

Lindsey, C. & Sheather, S. (2010) 'Variable selection in linear regression' The Stata Journal, 10(4), p650-669.

Angrist, J. & Pischke, J. (2008) Mostly Harmless Econometrics: An Empiricist's Companion. Princeton, NJ: Princeton University Press.

University of Virginia Library (2015) University of Virginia Library. Available at: https://library.virginia.edu/data/articles/getting-started-with-quantile-regression [Downloaded: 20/11/2023]

Koenker, R. & Hallock, K. (2001) ‘Quantile Regression’ Journal of Economic Perspectives, 15(4), p143-156.

Rodriguez, R & Yao, Y. (2017) ‘Five Things You Should Know about Quantile Regression’. Available at: https://support.sas.com/resources/papers/proceedings17/SAS0525-2017.pdf [Downloaded: 10/11/2023]

Rios-Avila, F. & Maroto, M. (2022) ‘Moving Beyond Linear Regression: Implementing and Interpreting Quantile Regression Models with Fixed Effects’ Sociological Methods & Research, p1-44 Available at: https://journals.sagepub.com/doi/epub/10.1177/00491241211036165 [Downloaded: 15/09/2023]

Koenker, R. (2004) ‘Quantile Regression for longitudinal data’ Journal of Multivariate Analysis, 91, p74-89.

Johar, M. & Katayama, H. (2012) ‘Quantile Regression Analysis of Body Mass and Wages’ Health Economics, 21, p597-611.

Waldmann, E. (2018) ‘Quantile Regression: A short story on how and why’ Statistical Modeling, 18(3-4), 203-218.

Le Cook, B. & Manning, W. (2013) ‘Thinking beyond the mean: a practical quide for using quantile regression methods for health services research’ Shanghai Archives of Psychiatry, 25(1), p55-59.

Ando, T. & Tsay, R. (2011) ‘Quantile regression models with factor-augmented predictors and information criterion’ Econometrics Journal, 14(1), p1-24.

Petscher, Y. & Logan, J. (2014) ‘Quantile Regression in the Study of Developmental Sciences’ Child Development, 85(3), p861-881.

Mingxiang, L. (2015) ‘Moving Beyond the Linear Regression Model: Advantages of the Quantile Regression Model’. Journal of Management, 41(1), p71-98.