ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ ΚΑΙ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΓΡΑΜΜΙΚΑ ΜΟΝΤΕΛΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Ονοματεπώνυμο: Δίπλας Χρήστος Α.Μ.: 16389
Κανέλλος Χρήστος Α.Μ.: 16408

Εποπτεύων Καθηγητής: Δρ. Παναγιώτα Βάθη Σαράβα

ΜΕΣΣΟΛΟΓΓΙ - 2019

Πρόλογος

Τα γραμμικά υποδείγματα παλινδρόμησης, όπως είναι ο τίτλος της παρούσας εργασίας, εκτιμούν τη γραμμική σχέση μιας μεταβλητής, η οποία ονομάζεται εξαρτημένη, με μια ή ένα σύνολο άλλων μεταβλητών που ονομάζονται ανεξάρτητες, με απώτερο σκοπό την ερμηνεία κάποιων φαινομένων, την εκτίμηση και την πρόβλεψη της συμπεριφοράς της εξαρτημένης μεταβλητής.

Η όλο και μεγαλύτερη άνθηση των επιστημών και οι αυξανόμενες ανάγκες της αγοράς, έχουν κάνει την ανάλυση παλινδρόμησης, στην οποία ανήκουν επιστημονικά τα γραμμικά υποδείγματα παλινδρόμησης, απαραίτητα σε ένα όλο και μεγαλύτερο φάσμα χρηστών των μεθοδολογιών της.

Όπως είναι γνωστό, στα οικονομικά, η χρήση των μεθοδολογιών της παλινδρόμησης είναι καθημερινή υπόθεση, καθώς οι μεταβλητές που έρχονται αντιμέτωποι οι οικονομολόγοι δημιουργούν αλληλεπιδράσεις, καθώς συσχετίζονται μεταξύ τους.

Έννοιες όπως ο πληθωρισμός, η ανεργία, ο δείκτης τιμών καταναλωτή, οι οποίες ακούγονται όχι μόνο ανάμεσα στους κύκλους των οικονομολόγων, αλλά αναφέρονται και στις ειδήσεις καθημερινά, είναι άρρηκτα συσχετισμένες μεταξύ τους. Έτσι, εάν μέσα από ένα γραμμικό υπόδειγμα παλινδρόμησης μπορούσαμε να ανακαλύψουμε μια σχέση που συνδέει κάποιες, ή όλες τις παραπάνω μεταβλητές, θα μπορούσαμε να κατανοήσουμε καλύτερα κάποια οικονομικά φαινόμενα, αλλά και να προβλέψουμε κάποια μελλοντικά τέτοια φαινόμενα.

Φυσικά, η ανάλυση παλινδρόμησης δεν είναι χρήσιμη μόνο στους οικονομολόγους, αλλά σε ένα ευρύ σύνολο επιστημόνων, όπως για παράδειγμα οι κοινωνιολόγοι, οι μετεωρολόγοι, οι μηχανικοί, οι επιστήμονες των υπολογιστών κ.α.

Σκοπός της παρούσας εργασίας είναι να εισάγει τον αναγνώστη στα υποδείγματα απλής και πολλαπλής παλινδρόμησης, όπως αυτά περιγράφονται και σε δεύτερο χρόνο να αναφέρει κάποιες από τις μεθοδολογίες που χρησιμοποιούνται για την εκτίμηση των υποδειγμάτων και τη χρήση τους στη διαδικασία της πρόβλεψης των τιμών της εξαρτημένης μεταβλητής.

Περίληψη

Στην παρούσα εργασία θα ασχοληθούμε με τα γραμμικά υποδείγματα παλινδρόμησης και κάποιες εφαρμογές τους στην ανάλυση απλής και πολλαπλής παλινδρόμησης. Θα προσπαθήσουμε να αναλύσουμε τη σχέση εισοδήματος και κατανάλωσης μέσα από τα δεδομένα 20 νοικοκυριών, ώστε να αναγνωρίσουμε την ύπαρξη της σχέσης μεταξύ των δυο μεταβλητών, αλλά και την ένταση της σχέσης τους.

Στο επόμενο κεφάλαιο γίνεται επέκταση της αναγνώρισης της σχέσης μεταξύ των μεταβλητών αυτών, όταν όμως μέσα στη σχέση συμπεριλαμβάνονται και οι δαπάνες στέγασης των 20 νοικοκυριών, όπου θα δείξουμε ότι η σχέση υπάρχει και είναι και αρκετά ισχυρή.

Στα κεφάλαια της Απλής Γραμμικής Παλινδρόμησης και Πολλαπλής Γραμμικής Παλινδρόμησης γίνεται χρήση του στατιστικού λογισμικού SPSS 25 ώστε να εφαρμοστούν όλες οι μεθοδολογίες που περιγράφονται στην εργασία και χρησιμοποιούνται στην επίλυση δυο επιλεγμένων ασκήσεων από το βιβλίο Εισαγωγή στην Οικονομετρία (Χρήστου, 2008).

Περιεχόμενα

Εισαγωγή

Σκοπός της παρούσας εργασίας είναι να φέρει τον αναγνώστη σε επαφή με μερικά από τα γραμμικά υποδείγματα απλής και πολλαπλής γραμμικής παλινδρόμησης, όπως επίσης και με μη γραμμικά υποδείγματα. Έτσι, αφού γίνει η εισαγωγή στο πρώτο κεφάλαιο με την ιστορική αναδρομή και την αναφορά σε κάποιες χρήσιμες έννοιες, στο επόμενο κεφάλαιο ο αναγνώστης εισάγεται στα απλά και πολλαπλά γραμμικά υποδείγματα παλινδρόμησης και σε μη γραμμικά υποδείγματα.

Στο τρίτο κεφάλαιο της εργασίας γίνεται αναφορά σε βασικές έννοιες της απλής γραμμικής παλινδρόμησης, όπως είναι η εξαρτημένη και η ανεξάρτητη μεταβλητή, οι συντελεστές παλινδρόμησης, τα κατάλοιπα, αλλά και έννοιες όπως ο συντελεστής γραμμικής συσχέτισης και ο συντελεστής προσδιορισμού της παλινδρόμησης, με εκτενή αναφορά στην ερμηνεία και τη σημασία τους. Γίνεται επίσης αναφορά σε ένα από τα εργαλεία της παλινδρόμησης, όπως είναι το διάγραμμα διασποράς.

Στο επόμενο κεφάλαιο ο αναγνώστης μπορεί να έρθει σε επαφή με την πολλαπλή παλινδρόμηση, όπου γίνεται επέκταση των εννοιών που αναφέρθηκαν στην απλή παλινδρόμηση. Επίσης, γίνεται αναφορά στο θεώρημα GAUSS-MARKOV και στο συντελεστή πολλαπλού προσδιορισμού και το διορθωμένο συντελεστή πολλαπλού προσδιορισμού.

Στα δυο αυτά κεφάλαια, δίνονται παραδείγματα εκτίμησης ενός απλού και ενός πολλαπλού γραμμικού υποδείγματος παλινδρόμησης. Δημιουργούμε, δηλαδή, ένα απλό γραμμικό υπόδειγμα παλινδρόμησης μεταξύ του εισοδήματος και της κατανάλωσης, ενώ στο επόμενο κεφάλαιο δημιουργούμε ένα πολλαπλό γραμμικό υπόδειγμα μεταξύ της κατανάλωσης ως εξαρτημένη μεταβλητή και του εισοδήματος και των στεγαστικών καταναλωτικών δαπανών, ως εξαρτημένες μεταβλητές.

Η εργασία κλείνει με κάποια χρήσιμα συμπεράσματα, τις βιβλιογραφικές αναφορές και τις διαδικτυακές πηγές, στις οποίες μπορεί ο αναγνώστης να ανατρέξει και να εμβαθύνει σε ένα χρήσιμο εργαλείο της ερευνητικής μεθοδολογίας, όπως είναι η ανάλυση παλινδρόμησης.

Κεφάλαιο 1. Στατιστική

1.1 Ιστορική αναδρομή της Στατιστικής

Η Στατιστική (Keller, 2010) είναι η επιστήμη η οποία χρησιμοποιεί δεδομένα, τα οποία συλλέγονται με δειγματοληπτικές μεθόδους όπως είναι αυτές των ερωτηματολογίων, είτε μέσω πειραμάτων κατάλληλα σχεδιασμένων, και με άλλες μεθόδους, ώστε να εξάγει αποτελέσματα τα οποία εφαρμόζονται σε πολλούς τομείς όπως τα οικονομικά, η διοίκηση επιχειρήσεων, το μάρκετινγκ, αλλά και επιστήμες όπως η ιατρική, η ψυχολογία, κ.α.

Σύμφωνα, μάλιστα, με τον Γερμανό φιλόσοφο Gottfried Achenwall, νομικό, ιστορικό, οικονομολόγο και στατιστικό, η λέξη «στατιστική» αφορά το σύνολο των πληροφοριών (δεδομένων) που χρειάζεται ένα κράτος για να λειτουργήσει (Χαραλάμπους, 2014).

Σύμφωνα με την ίδια εργασία (Χαραλάμπους, 2014), όπου γίνεται αναδρομή στα διάφορα ιστορικά σημεία της στατιστικής, η εφαρμογή μεθόδων της Στατιστικής συναντάται από τους αρχαίους χρόνους όπου γίνεται χρήση κάποιων από τα περιγραφικά στατιστικά όπως αυτά της μέσης τιμής και της επικρατούσας τιμής, αλλά και μιας μορφής δειγματοληψία και μιας μορφής απογραφή.

Στους πιο σύγχρονους χρόνους, γύρω στο δέκατο αιώνα παρουσιάζεται το πρώτο γράφημα, το οποίο θα χρησιμοποιούταν στα σχολεία των μοναστηριών και αφορούσε την κίνηση των πλανητών, ενώ τον επόμενο αιώνα καταγράφεται η αρχή των επίσημων στατιστικών στην Αγγλία, καθώς στο βιβλίο Domesday καταγράφεται η απογραφή για τον Ουίλιαμ τον Κατακτητή του καινούριου του βασιλείου.

Το 12^{ο} αιώνα ξεκινάει ο ετήσιος έλεγχος καθαρότητας των νομισμάτων που κυκλοφορούν από το Νομισματοκοπείο της Αγγλίας. Γίνεται ενός είδους δειγματοληψία καθώς για τον έλεγχό τους επιλέγεται τυχαία ένα ποσοστό νομισμάτων από το συνολικό αριθμό που κυκλοφορούν. Η μέθοδος αυτή χρησιμοποιείται ακόμα και σήμερα!

Αργότερα, το 16^{ο} αιώνα, γίνεται η μαθηματική θεμελίωση της Στατιστικής, καθώς ένας «τζογαδόρος», ο οποίος ήταν μαθηματικός, φυσικός και ιατρός, ο Geronymo Cardano υπολογίζει τις πιθανότητες διαφόρων ρίψεων δυο ζαριών. Τις πιθανότητες αυτές θα τις χρησιμοποιούσαν άλλοι παίχτες τυχερών παιχνιδιών… Τον επόμενο αιώνα, μέσα από την αλληλογραφία τους οι Pascal και Fermat, δημιουργούν τη Θεωρία Πιθανοτήτων, καθώς μελετούν τον καταμερισμό των στοιχημάτων στα τυχερά παιχνίδια. Λίγο πιο μετά, μέσα στον ίδιο αιώνα (17^{ος}) γράφεται και το πρώτο βιβλίο στη Θεωρία Πιθανοτήτων, από τον Huygens, ο τίτλος του οποίου είναι «Η λογική των τυχερών παιχνιδιών».

Το $17^{ο}$ αιώνα παρουσιάζεται και το πρώτο γνωστό γράφημα στατιστικών δεδομένων το οποίο δημιουργείται από τον Ολλανδό αστρονόμο και χαρτογράφο Michael van Langren. Το γράφημα αυτό αφορά το μέγεθος των σφαλμάτων των διαφορετικών εκτιμητών της απόστασης από την πόλη Τολέδο μέχρι τη Ρώμη.

Οι βάσεις των ασφαλειών ζωής και γενικότερα της ασφαλιστικής επιστήμης μπαίνουν το 1693, όπου ο Άγγλος αστρονόμος, μαθηματικός και μετεωρολόγος Edmund Halley δημιουργεί τους πρώτους πίνακες θνησιμότητας, οι οποίοι περιγράφουν τους ρυθμούς θανάτου ενός πληθυσμού, ανάλογα με την ηλικία.

Αργότερα, τον $18^{ο}$ αιώνα, ο Βολταίρος και ο de la Condamine, φίλος του μαθηματικός, δημιουργούν υψηλά κέρδη από την επένδυση σε μια λαχειοφόρο με ομόλογα, από την οποία παρατήρησαν ότι τα συνολικά κέρδη της λαχειοφόρου ήταν περισσότερα από το συνολικό κόστος των δελτίων. Μερικά χρόνια μετά, μέσα στον $18^{ο}$ αιώνα, ο Καζανόβα γίνεται θεματοφύλακας των Γαλλικών Κρατικών Λαχείων και θεωρείται ότι έχει συμβάλλει στην επινόηση του συστήματος των λαχείων. Λίγα χρόνια αργότερα αποδεικνύεται το λεγόμενο Θεώρημα του Bayes, που αφορά στον έλεγχο πεποιθήσεων και υποθέσεων, κάνοντας χρήση της δεσμευμένης πιθανότητας.

Εκτός των άλλων, μέσα στον $18^{ο}$ αιώνα, καταγράφεται και η πρώτη απογραφή πληθυσμού στις Η.Π.Α., εισάγονται για πρώτη φορά γραφήματα που παρουσιάζουν οικονομικά δεδομένα, από τον William Playfair, ενώ για πρώτη φορά γίνεται η χρήση της λέξης statistics στα Αγγλικά, σε δημοσιεύσεις του Sir John Sinclair, στις οποίες περιέχονται στατιστικά στοιχεία για τη ζωή στη Σκωτία.

Στις αρχές του $19^{ου}$ αιώνα, και συγκεκριμένα το 1805, παρουσιάζεται η μέθοδος ελαχίστων τετραγώνων από τον Andrien – Marie Legendre, η οποία όπως θα δούμε και στα επόμενα κεφάλαια, προσαρμόζει μια συνάρτηση (ευθεία) σε ένα σύνολο από δεδομένα (παρατηρήσεις). Τρία χρόνια μετά, παρουσιάζεται η Κανονική Κατανομή, από τον Gauss με τη συμβολή του Laplace, μια πολύ χρήσιμη συνεχής Κατανομή, στη Στατιστική και στη Θεωρία Πιθανοτήτων.

Στα μέσα του $19^{ου}$ αιώνα γίνεται η αρχή της ιατρικής στατιστικής, όταν ένας βρετανός επιδημιολόγος, ο William Farr δημιουργεί το επίσημο σύστημα καταγραφής των αιτιών θανάτου στην Αγγλία και την Ουαλία. Την ίδια περίοδο, και με μιας μορφής διαγράμματος πίτας (piechart), χρησιμοποιούνται στατιστικά ατυχημάτων του Πολέμου της Κριμαίας από την Florence Nightingale, με σκοπό να επηρεαστεί η κοινή γνώμη και το υπουργείο πολέμου της Αγγλίας.

Άλλη μια σημαντική ανακάλυψη παρουσιάζεται προς το τέλος του $19^{ου}$ αιώνα, καθώς το 1977 περιγράφεται η παλινδρόμηση από τον F. Galton, ενώ περίπου 10 χρόνια αργότερα ο ίδιος εισάγει και την έννοια της συσχέτισης.

Στα τελευταία χρόνια του ίδιου αιώνα εισάγεται και ο όρος της τυπικής απόκλισης, ενός σημαντικού περιγραφικού στατιστικού μέτρου, από τον Karl Pearson, ενώ αργότερα ο ίδιος αναπτύσσει και τους ελέγχους καλής προσαρμογής και ανεξαρτησίας χ-τετράγωνο.

Στις αρχές του $20^{ου}$ αιώνα αρχίζει η ανάπτυξη των χρηματοοικονομικών μαθηματικών, μετά την παρατήρηση του Louis Bachelier, ο οποίος δείχνει ότι οι διακυμάνσεις των τιμών των μετοχών στο χρηματιστήριο συμπεριφέρονται όπως τα μόρια τυχαίας κίνησης Brown.

Λίγα χρόνια μετά παρουσιάζεται και το t-test, από τον χημικό Gossett, μέσα από μικρά δείγματα μπύρας, για να εξασφαλιστεί ότι έχουν εξίσου καλή γεύση. Αργότερα και κατά τη διάρκεια του πρώτου παγκοσμίου πολέμου, αναπτύσσονται στατιστικοί νόμοι πρόβλεψης εναέριων μαχών από τον Lanchester, όπου παρατηρείται ότι με διπλασιασμό του μεγέθους των δυναμικού ο στρατός ξηράς γίνεται δυο φορές πιο δυνατός, ενώ η αεροπορία γίνεται τέσσερις φορές πιο ισχυρή.

Τον ίδιο αιώνα, μερικά χρόνια αργότερα, ο Fisher παρουσιάζει τρόπους ώστε να μπορεί να αποφασίζει ο ερευνητής εάν τα αποτελέσματα του πειράματος που εκτέλεσε είναι σημαντικά!

Μέσα στα χρόνια του πρώτου παγκοσμίου πολέμου ο Neyman παρουσιάζει τα διαστήματα εμπιστοσύνης, τα οποία έχουν πολύ σημαντική εφαρμογή στη στατιστική συμπερασματολογία, ενώ ο Alan Turing κάνοντας χρήση Μπεϋζιανής στατιστικής και τον πρώτο προγραμματιζόμενο ηλεκτρονικό υπολογιστής, τον Colossus, καταφέρνει να σπάσει το Γερμανικό κώδικα Enigma.

Στα μέσα του $20^{ου}$ αιώνα, παρουσιάζεται η σχέση που έχουν το κάπνισμα με τον καρκίνο των πνευμόνων, από τους Richard Doll που ήταν γιατρός και τον Bradford Hill, που ήταν βιοστατιστικός. Μετά από μερικά χρόνια, αναπτύσσεται ένα εργαλείο για τους γιατρούς, ο εκτιμητής Kaplan–Meier, που δίνει τη δυνατότητα στους γιατρούς με έναν απλό στατιστικό τρόπο να κρίνουν ποιες θεραπευτικές αγωγές και ποια φάρμακα έχουν καλύτερη αποδοτικότητα. Την ίδια περίοδο, ο Taguchi με τις στατιστικές του μεθόδους καταφέρνει μεγαλύτερη βελτίωση στην ποιότητα των εξαρτημάτων των αυτοκινήτων και των ηλεκτρονικών.

Στις πιο σύγχρονες εποχές, στο τέλος δηλαδή του $20^{ου}$ αιώνα και στις αρχές του $21^{ου}$ αιώνα, έχουμε τις σημαντικές ανακαλύψεις της στατιστικής γλώσσας προγραμματισμού R, το 1993, η οποία έχει αναπτυχθεί πολύ τα τελευταία χρόνια και είναι ένα σημαντικό εργαλείο για τη στατιστική ανάλυση, ενώ το 1997 πρωτοεμφανίζεται ο όρος «Μεγάλα Δεδομένα» (Big Data), που έχει αναπτυχθεί πολύ τελευταία.

Εν κατακλείδι, η εφαρμογή της στατιστικής και των στατιστικών μεθόδων έχει εισχωρήσει πολύ δυναμικά σε πολλές εκφάνσεις της επιστήμης, καθώς στο CERN επιβεβαιώθηκε με πιθανότητα μια στα 3,5 εκατομμύρια, η ύπαρξη του σωματιδίου του Higgs. Αλλά όχι μόνο αυτό, καθώς στον αθλητισμό, όπως είδαμε στην τέχνη του πολέμου κ.λπ., οι εφαρμογές της στατιστικής είναι πολύ χρήσιμες.

Η Στατιστική είναι μια επιστήμη τα εργαλεία της οποίας χρησιμοποιούνται από τους αρχαίους χρόνους, μέχρι και σήμερα. Όπως παρουσιάζεται στην εργασία η εξέλιξη της επιστήμης της Στατιστικής είναι ραγδαία τα τελευταία χρόνια και επιτρέπει τη χρήση της σε πολλούς τομείς, το οποίο σαν γεγονός μας αφήνει πολλές υποσχέσεις ότι στο μέλλον θα ακολουθήσουν πολύ σημαντικά αποτελέσματα τα οποία θα οφείλουν την ύπαρξή τους στη Στατιστική.

1.2 Βασικές έννοιες Στατιστικής

1.2.1. Πληθυσμός – Δείγμα

Από τις πιο βασικές έννοιες που συναντάει κανείς στη στατιστική, είναι αυτές του πληθυσμού και του δείγματος (Keller, 2010).

Σαν πληθυσμό ορίζουμε το σύνολο όλων των οντοτήτων των οποίων θέλουμε να μελετήσουμε και να αναλύσουμε ορισμένα από τα χαρακτηριστικά του.

Ως δείγμα θεωρείται ένα υποσύνολο του πληθυσμού, το οποίο συλλέγεται κατά τυχαίο τρόπο, ώστε να μελετηθούν και να αναλυθούν τα χαρακτηριστικά του. Ο τρόπος συλλογής των στοιχείων που θα απαρτίζουν το δείγμα που προέρχεται από τον πληθυσμό, ονομάζεται δειγματοληψία, και μπορεί να είναι απλή τυχαία δειγματοληψία, η στρωματοποιημένη τυχαία δειγματοληψία, η συστηματική δειγματοληψία, κλπ. Τις μεθόδους δειγματοληψίας δεν θα τις αναλύσουμε περισσότερο, καθώς ξεφεύγουν από τις ανάγκες της εργασίας.

Κάποια από τα παραδείγματα που θα μπορούσαν να περιγράψουν έναν πληθυσμό είναι το σύνολο των ατόμων που διαμένουν σε κάποια χώρα (πληθυσμός) για τους οποίους ενδιαφερόμαστε να μελετήσουμε και να αναλύσουμε το ετήσιο εισόδημά τους (χαρακτηριστικό), είτε την ετήσια κατανάλωση (χαρακτηριστικό), ενώ ως δείγμα του πειράματος αυτού είναι το υποσύνολο των ατόμων που θα εκλεχθούν ώστε να εξεταστούν αναφορικά με το εισόδημά τους και την ετήσια κατανάλωσή τους.

Ένα άλλο παράδειγμα είναι αυτό του ελέγχου των ανασφάλιστων αυτοκινήτων μιας χώρας. Το σύνολο των αυτοκινήτων που κυκλοφορούν στη χώρα αυτή είναι ο πληθυσμός που ελέγχεται, ενώ η ύπαρξη νόμιμης ασφάλειας για κυκλοφορία είναι το χαρακτηριστικό που εξετάζουμε. Ως δείγμα αυτού του πειράματος είναι το υποσύνολο των αυτοκινήτων που επιλέγονται για να ελεγχθούν εάν διαθέτουν νόμιμη ασφάλιση.

Από τα παραπάνω παραδείγματα είναι προφανής η διάκριση μεταξύ πληθυσμού και δείγματος. Συνήθως είναι πολύ δύσκολο, είτε πολυδάπανο να συλλεχθεί και να εξεταστεί ως προς το χαρακτηριστικό του ο πληθυσμός και για το λόγο αυτό επιλέγεται ένα υποσύνολο του πληθυσμού, το δείγμα, το οποίο είναι και ευκολότερο να εξεταστεί ως προς τα χαρακτηριστικά που μας ενδιαφέρουν.

1.2.2. Τύποι Δεδομένων

Τρεις είναι οι βασικοί τύποι δεδομένων που συλλέγονται και χρησιμοποιούνται σε αναλύσεις.

1. Διαστρωματικά Δεδομένα (CrossSectionData)

Τα Διαστρωματικά Δεδομένα είναι οι τιμές κάποιων χαρακτηριστικών που συλλέγονται σε συγκεκριμένη χρονική στιγμή και συνήθως χρησιμοποιούνται ερωτηματολόγια είτε καταμετρήσεις (απογραφές).

Ένα παράδειγμα Διαστρωματικών Δεδομένων μπορεί να είναι η συλλογή των δεδομένων μισθού και ετών εκπαίδευσης του συνόλου των εργαζομένων μιας εταιρείας.

Τα δεδομένα του μισθού και της εκπαίδευσης συλλέχθηκαν κάποια συγκεκριμένη χρονική στιγμή (π.χ. το Νοέμβριο του 2016).

2. Χρονοσειρές (Time Series)

Τα δεδομένα χρονοσειρών δείχνουν την πορεία των τιμών κάποιου συγκεκριμένου χαρακτηριστικού στην πορεία του χρόνου. Ο χρόνος αφορά τακτικές χρονικές στιγμές, δηλαδή ημερήσιες τιμές, εβδομαδιαίες, μηνιαίες, τριμηνιαίες, ετήσιες, κλπ τιμές του χαρακτηριστικού που εξετάζεται.

Ένα παράδειγμα χρονοσειράς μπορεί να είναι η εξέλιξη του μισθού του συνόλου των εργαζομένων της εταιρείας που εξετάσαμε παραπάνω και οι τιμές του μισθού να έχουν συλλεχθεί ετήσια.

3. Διαχρονικά – Διαστρωματικά Δεδομένα (PanelData)

Τα Δεδομένα Panelείναι ένας συνδυασμός των δυο πρώτων κατηγοριών δεδομένων, διαστρωματικών και χρονοσειρών.

Ένα παράδειγμα Panelδεδομένων, μπορεί να είναι οι καταναλωτικές δαπάνες του iνοικοκυριού στο έτος t. Με iσυμβολίζεται το κάθε νοικοκυριό που εξετάζεται ξεχωριστά, ενώ με tσυμβολίζεται το κάθε έτος στο οποίο συλλέγονται τα δεδομένα.

1.2.3. Συσχέτιση

Η έννοια της συσχέτισης που θα μας απασχολήσει και στα επόμενα κεφάλαια, αφορά τη σχέση αιτίας και αιτιατού που αναπτύσσεται μεταξύ δυο χαρακτηριστικών (μεταβλητών), είτε αντίστροφα.

Ένα παράδειγμα που δείχνει τη συσχέτιση μεταξύ δυο μεταβλητών, είναι η σχέση μεταξύ του ύψους του πατέρα το οποίο συσχετίζεται με το ύψος των απογόνων του, δηλαδή αν γνωρίζουμε το ύψος του πατέρα, να μπορούμε να γνωρίζουμε, ίσως με κάποια απόκλιση, το ύψος των παιδιών του.

Κεφάλαιο 2. Παλινδρόμηση

2.1 Εισαγωγή στην παλινδρόμηση

Η παλινδρόμηση είναι η μέθοδος με την οποία ερευνάται εμπειρικά η σχέση που υπάρχει, ανάμεσα σε μια μεταβλητή, η οποία ονομάζεται εξαρτημένη και σε ένα σύνολο από μεταβλητές, οι οποίες ονομάζονται ανεξάρτητες ή επεξηγηματικές (Χρήστου, 2008).

Από την παλινδρόμηση προσπαθούμε να προσδιορίσουμε τη συναρτησιακή μορφή της σχέσης ανάμεσα στην εξαρτημένη και τις ανεξάρτητες μεταβλητές, με την οποία μπορούμε, ανάλογα με τις τιμές των ανεξάρτητων μεταβλητών, να προσδιορίσουμε την τιμή της εξαρτημένης μεταβλητής. Αυτή η διαδικασία γίνεται από την προσαρμογή των δεδομένων του δείγματος που έχουν συλλεχθεί, πάνω στη συνάρτηση.

Πρόκειται για μια στατιστική σχέση εξάρτησης, με κάποια συναρτησιακή σχέση, και όχι για μια προσδιοριστική (ντετερμινιστική) σχέση, στην οποία οι μεταβλητές που συμμετέχουν στο υπόδειγμα είναι τυχαίες μεταβλητές, που προσδιορίζονται από κάποια κατανομή πιθανότητας. Η σχέση αυτή μεταξύ της εξαρτημένης μεταβλητής και των επεξηγηματικών μεταβλητών θα θεωρείται δεδομένη κάθε φορά που θα γίνεται αναφορά σε αυτή, παρότι πρόκειται για τυχαίες μεταβλητές.

Η συναρτησιακή σχέση που δημιουργείται έχει δυο κομμάτια. Το πρώτο κομμάτι είναι το ερμηνευτικό κομμάτι, που ονομάζεται και συστηματικό, εκείνο δηλαδή που δίνει τις πληροφορίες και μπορεί να εκτιμήσει την εξαρτημένη μεταβλητή μέσα από τις επεξηγηματικές, ή ανεξάρτητες μεταβλητές.

Το δεύτερο κομμάτι είναι αυτό που δεν βοηθάει στην ερμηνεία της εξαρτημένης μεταβλητής. Είναι το μη συστηματικό κομμάτι και είναι η διαφορά των τιμών της εξαρτημένης μεταβλητής οι οποίες παρατηρούνται στο δείγμα, από τις εκτιμήσεις της και ονομάζεται υπόλοιπο, σφάλμα ή κατάλοιπα.

Σε αντίθεση με τις επεξηγηματικές μεταβλητές, ο όρος των καταλοίπων, όπως και ο όρος της εξαρτημένης μεταβλητής, είναι στοχαστικός. Γι' αυτό και οι σχέσεις που περιγράφονται μέσα από τα υποδείγματα παλινδρόμησης είναι στατιστικές και όχι προσδιοριστικές.

Η ύπαρξη του όρου των καταλοίπων οφείλεται σε τυχαίους παράγοντες, οι οποίοι είναι οι εξής:

1. Λάθος εξειδίκευση του υποδείγματος.

   Αρκετές φορές, κάποιος ερευνητής που διενεργεί την ανάλυση εκτελώντας παλινδρομήσεις, υποθέτει ότι το υπόδειγμα που πρέπει να εκτιμηθεί είναι γραμμικό, είτε είναι ένα υπόδειγμα απλής παλινδρόμησης κλπ., καθώς τα περισσότερα φαινόμενα ερμηνεύονται από γραμμικές σχέσεις και από σχέσεις μεταξύ μιας εξαρτημένης και μιας ανεξάρτητης μεταβλητής.

   Στην περίπτωση, όμως, που αυτές οι υποθέσεις του ερευνητή είναι λανθασμένες, τότε το σφάλμα στην εκτίμηση της εξαρτημένης μεταβλητής θα είναι αυξημένο.

2. Παράλειψη εισαγωγής επεξηγηματικών μεταβλητών στο υπόδειγμα.

   Η επεξηγηματική ικανότητα ενός υποδείγματος παλινδρόμησης θα είναι αυξημένη, σε περίπτωση που εισαχθούν στο υπόδειγμα ανεξάρτητες μεταβλητές που έχουν επίδραση πάνω στην εξαρτημένη μεταβλητή του συγκεκριμένου υποδείγματος.

   Αντίθετα, στην περίπτωση που οι μεταβλητές αυτές, ενώ είναι απαραίτητες, παραλειφθούν να εισαχθούν στο υπόδειγμα, τότε αναμένεται να μειωθεί η επεξηγηματική ικανότητα του υποδείγματος και αναμένεται να αυξηθεί ο όρος των καταλοίπων.

3. Από παράγοντες που δεν μπορούν να μετρηθούν ποσοτικά

   Συχνά, κάποιοι από τους παράγοντες, που παίζουν ρόλο στην ερμηνεία της εξαρτημένης μεταβλητής δεν είναι δυνατό να ποσοτικοποιηθούν. Ένα παράδειγμα είναι οι προσδοκίες των καταναλωτών για ένα προϊόν, οι οποίες μπορούν να περιγραφούν ποιοτικά, αλλά δεν υπάρχει κάποιος τρόπος να δημιουργηθεί μια μεταβλητή που να τις περιγράφει και ποσοτικά.

   Γενικά, όταν οι επεξηγηματικές μεταβλητές δεν μπορούν να εκτιμηθούν ποσοτικά, εισάγονται στον όρο των καταλοίπων και μειώνουν την επεξηγηματική ικανότητα του υποδείγματος.

4. Τυχαίοι έκτακτοι (έντονοι) παράγοντες.

   Στον όρο των σφαλμάτων εισάγονται και εκείνοι οι παράγοντες που μπορεί να συμβούν έκτακτα και οι οποίοι επηρεάζουν την εξαρτημένη μεταβλητή. Ένα παράδειγμα θα μπορούσε να είναι μια σοβαρή οικονομική κρίση, ένα κραχ του χρηματιστηρίου, κλπ.

Η μέθοδος της παλινδρόμησης χρησιμοποιείται σε πολλές επιστήμες για την εκτίμηση της σχέσης μεταξύ μεταβλητών. Τα οικονομικά, η μετεωρολογία, οι φυσικές επιστήμες κλπ. είναι μερικές από αυτές τις επιστήμες.

2.2 Απλή παλινδρόμηση

Έχουμε ήδη αναφέρει ότι με την έννοια της παλινδρόμησης, στη στατιστική εννοούμε τη μελέτη της σχέσης που δημιουργείται μεταξύ ενός αριθμού μεταβλητών και μιας μεταβλητής, την οποία χαρακτηρίσαμε ως εξαρτημένη.

Με τον όρο απλή, χαρακτηρίζουμε την παλινδρόμηση, στην οποία μελετάμε τη σχέση ανάμεσα σε μια μεταβλητή, της εξαρτημένης και μόνο μιας ανεξάρτητης μεταβλητής. Αυτό θα γίνει με τη δημιουργία μαθηματικών υποδειγμάτων που θα περιγράφουν τη σχέση των δυο μεταβλητών.

Ένα υπόδειγμα απλής παλινδρόμησης έχει τη μορφή συναρτησιακής σχέσης μιας εξαρτημένης μεταβλητής με μια ανεξάρτητη, ως εξής:

$$Y = f(X)$$

Ο σκοπός της παλινδρόμησης είναι η πρόβλεψη των τιμών της εξαρτημένης μεταβλητής, από τις τιμές της ανεξάρτητης μεταβλητής, κάθε φορά που θα προκύπτει η μελέτη των δυο αυτών μεταβλητών στο μέλλον.

Ένα παράδειγμα απλής παλινδρόμησης στα οικονομικά είναι η εκτίμηση της σχέσης μεταξύ της τιμής του καφέ (εξαρτημένη μεταβλητή) και της κατανάλωσης του καφέ σε φλιτζάνια καφέ ανά άτομο και ανά ημέρα (ανεξάρτητη μεταβλητή).

Η εκτίμηση της σχέσης της μέσης θερμοκρασίας των πόλεων μιας χώρας η οποία εξαρτάται από το γεωγραφικό μήκος (ή το γεωγραφικό πλάτος) των πόλεων, το οποίο φυσικά παίζει το ρόλο ανεξάρτητης μεταβλητής, είναι ένα παράδειγμα παλινδρόμησης από την επιστήμη της μετεωρολογίας.

Ακόμα ένα παράδειγμα, από τα χρηματοοικονομικά, είναι η εκτίμηση της σχέσης των αποδόσεων των τιμών μιας μετοχής στο χρηματιστήριο ως εξαρτημένη μεταβλητή και των αποδόσεων των τιμών του γενικού δείκτη του χρηματιστηρίου στο οποίο διαπραγματεύεται η μετοχή, ως ανεξάρτητη μεταβλητή.

2.3 Πολλαπλή παλινδρόμηση

Αν και η βάση της παλινδρόμησης είναι η απλή, παρόλα αυτά η περίπτωση της απλής παλινδρόμησης είναι ειδική περίπτωση της πολλαπλής παλινδρόμησης, όταν ο αριθμός των επεξηγηματικών μεταβλητών είναι μία.

Έτσι, μπορούμε να πούμε ότι η πολλαπλή παλινδρόμηση, μελετά τη σχέση μεταξύ μιας μεταβλητής, της εξαρτημένης, με δυο ή περισσότερες ανεξάρτητες μεταβλητές. Η αύξηση αυτή των επεξηγηματικών μεταβλητών αυξάνει και την ερμηνευτική ικανότητα του υποδείγματος, προσδιορίζοντας με μεγαλύτερη ακρίβεια τις τιμές της εξαρτημένης μεταβλητής.

Όπως έχουμε περιγράψει ήδη και στην περίπτωση της απλής παλινδρόμησης, οι όροι της εξαρτημένης μεταβλητής Y και ο όρος των καταλοίπων είναι στοχαστικοί όροι, δηλαδή είναι τυχαίες μεταβλητές.

Αντίθετα, όλες οι ανεξάρτητες μεταβλητές που συμπεριλαμβάνονται στο υπόδειγμα είναι μη στοχαστικές, καθώς μπορούμε να θεωρήσουμε τις τιμές δεδομένες. Η στοχαστική σχέση που προσδιορίζεται από την παλινδρόμησης, όμως, παραμένει καθώς παραμένει να υπάρχει ο όρος των καταλοίπων, που όπως είπαμε παραπάνω, είναι στοχαστικός.

Ο διαταρακτικός όρος, ή αλλιώς ο όρος του σφάλματος στην πολλαπλή παλινδρόμηση, δημιουργείται για τους ίδιους λόγους τους οποίους περιγράψαμε και στην απλή παλινδρόμηση.

Ένα παράδειγμα πολλαπλής παλινδρόμησης είναι η εκτίμηση της σχέσης της αποταμίευσης των κατοίκων μιας χώρας, ως εξαρτημένη μεταβλητή των ανεξάρτητων μεταβλητών του εισοδήματος, των δαπανώ στέγασης και των καταναλωτικών δαπανών. Είναι ένα παράδειγμα, το οποίο προέρχεται από τα οικονομικά.

2.4 Γραμμική Παλινδρόμηση

Η γραμμικότητα, ως ιδιότητα της παλινδρόμησης, μπορεί να αναφέρεται σε δυο περιπτώσεις. Η πρώτη περίπτωση είναι να αναφέρεται στις παραμέτρους. Στην περίπτωση αυτή η μορφή των υποδειγμάτων μπορεί να είναι η επόμενη:

1. $Y = a + \beta X + \varepsilon$
2. $Y = a + \beta \sqrt{X} + \varepsilon$
3. $Y = a + \beta \frac{1}{X} + \varepsilon$
4. $Y = a + \beta X^2 + \varepsilon$
5. $Y = \beta_0 + \beta_1 X_1^1 + \beta_2 X_2^2 + \beta_3 X_3^3 + \cdots + \varepsilon$

Όπως φαίνεται στις παραπάνω περιπτώσεις, τα υποδείγματα απλής παλινδρόμησης, όπως είναι τα υποδείγματα 2-4 και το υπόδειγμα πολλαπλής παλινδρόμησης, όπως είναι το υπόδειγμα 5, δεν είναι γραμμικά ως προς τις μεταβλητές που υπάρχουν στο υπόδειγμα, αλλά είναι γραμμικά ως προς τις παραμέτρους τους.

Το υπόδειγμα 1, απλής παλινδρόμησης τυχαίνει να είναι γραμμικό και ως προς τις παραμέτρους του, αλλά και ως προς τις μεταβλητές του.

Σε άλλη περίπτωση, τα υποδείγματα μπορεί να είναι γραμμικά ως προς τις μεταβλητές και όχι (απαραίτητα) ως προς τις παραμέτρους τους, όπως είναι για παράδειγμα τα επόμενα υποδείγματα:

1. $Y = a + \beta X + \varepsilon$
2. $Y = a + \sqrt{\beta} X + \varepsilon$
3. $Y = a + \frac{1}{\beta} X + \varepsilon$
4. $Y = a + \beta^2 X + \varepsilon$
5. $Y = \beta_0 + \beta_1^1 X_1 + \beta_2^2 X_2 + \beta_3^3 X_3 + \cdots + \varepsilon$

Τα παραπάνω υποδείγματα απλής παλινδρόμησης 2-4 όπως και το υπόδειγμα πολλαπλής παλινδρόμησης 5, φαίνεται είναι γραμμικά ως προς τις μεταβλητές τους, αλλά όχι και ως προς τις παραμέτρους τους. Το πρώτο υπόδειγμα απλής παλινδρόμησης, όπως και πριν, μπορεί να χαρακτηριστεί γραμμικό και ως προς τις παραμέτρους του και ως προς τη μεταβλητή του.

Με τον όρο γραμμικότητα, στην ανάλυση παλινδρόμησης, αναφερόμαστε στα υποδείγματα τα οποία είναι γραμμικά ως προς τις παραμέτρους τους και όχι απαραίτητα και ως προς τις μεταβλητές τους. Τέτοια υποδείγματα μπορούν να εκτιμηθούν.

Σε ορισμένες περιπτώσεις, υπάρχει δυνατότητα εκτίμησης μη γραμμικών ως προς τις παραμέτρους τους υποδειγμάτων, τα οποία με κατάλληλο μετασχηματισμό μπορούν να μετατραπούν σε γραμμικά. Στη συνέχεια, μπορεί να χρησιμοποιηθεί η μεθοδολογία της ανάλυσης παλινδρόμησης.

2.5 Μη γραμμική παλινδρόμηση

Πολλές φορές η συναρτησιακή σχέση που συνδέει δυο μεταβλητές, την εξαρτημένη μεταβλητή Y και την ανεξάρτητη μεταβλητή X, δεν είναι γραμμική, δηλαδή δεν είναι της μορφής:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$$

Κάποιες από τις μη γραμμικές σχέσεις, με κατάλληλους μετασχηματισμούς, μπορούν να μετατραπούν σε γραμμικές σχέσεις.
Παρακάτω, θα αναφέρουμε μερικά από αυτά τα υποδείγματα.

2.5.1 Πολυωνυμική Μορφή Υποδειγμάτων

Πολυωνυμικά ονομάζονται τα μη γραμμικά υποδείγματα της μορφής:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k + u_i$$

Το παραπάνω υπόδειγμα μετατρέπεται σε γραμμικό υπόδειγμα, εάν χρησιμοποιήσουμε το μετασχηματισμό:

$$X_i^k = X_{ki}^*$$

Μετά την εφαρμογή του μετασχηματισμού μπορούμε να χρησιμοποιούμε τις μεθόδους εκτίμησης των παραμέτρων ενός γραμμικού συστήματος, όπως θα τις περιγράψουμε παρακάτω.

Τέτοια υποδείγματα μπορούν να χρησιμοποιηθούν για να περιγράψουν καμπύλες οριακού κόστους, οι οποίες έχουν ως εξαρτημένη μεταβλητή Y το οριακό κόστος και ως ανεξάρτητες μεταβλητές X το κάθε προϊόν.

2.5.2 Αντίστροφη Μορφή Υποδειγμάτων

Τα υποδείγματα που ανήκουν στην κατηγορία των υποδειγμάτων με αντίστροφη μορφή είναι τα εξής:

$$Y_i = \beta_0 + \beta_1 \frac{1}{X_i} + u_i$$

Η εφαρμογή του μετασχηματισμού:

$$\frac{1}{X} = X^*$$

Μετατρέπει τα υποδείγματα αυτά σε γραμμικά υποδείγματα, όπου με τη χρήση των μεθόδων των γραμμικών υποδειγμάτων μπορούμε να εκτιμήσουμε τους συντελεστές παλινδρόμησης.

Τα υποδείγματα αυτά είναι χρήσιμα στη μελέτη οικονομικών φαινομένων τέτοιων ώστε όταν η τιμή της ανεξάρτητης μεταβλητής X αυξάνεται, τότε η μέση τιμή της εξαρτημένης μεταβλητής πλησιάζει ένα επίπεδο τιμών, χωρίς όμως να μπορεί να πιάσει αυτό το επίπεδο.

Παραδείγματα τέτοιων υποδειγμάτων είναι η καμπύλη Phillips που περιγράφει την αντίστροφη σχέση μεταξύ του ρυθμού μεταβολής των μισθών και του επιπέδου της ανεργίας, όπως επίσης και οι καμπύλη Engel, που περιγράφει την αντίστροφη σχέση μεταξύ εισοδήματος και δαπανών σε ορισμένες κατηγορίες αγαθών.

2.5.3 Συνάρτηση Σταθερών Ελαστικοτήτων

Τα μη γραμμικά υποδείγματα στα οποία οι μεταβλητές X_i εμφανίζονται πολλαπλασιαστικά και ο κάθε ένας από τους συντελεστές β_i είναι σταθερός και ίσος με την ελαστικότητα της εξαρτημένης μεταβλητής Y, ως προς τη μεταβλητή X_i, είναι της μορφής:

$$Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} \dots X_k^{\beta_k} u_i$$

Η παραπάνω σχέση ονομάζεται συνάρτηση σταθερών ελαστικοτήτων.

Η τιμή του συντελεστή β_i υπολογίζεται ως εξής:

$\beta_i = \frac{\partial Y}{\partial X_i}\frac{X_i}{Y}$, για κάθε i = 1,2,….,k

Τα υποδείγματα σταθερών ελαστικοτήτων μετατρέπονται σε γραμμικά, χρησιμοποιώντας τους εξής λογαριθμικούς μετασχηματισμούς:

$$Y_i^* = logY_i$$
$$X_i^* = logX_i$$
$$\beta_0^* = log\beta_0$$
$$u_i^* = \log(u_i)$$

Μετά την εφαρμογή των μετασχηματισμών, η μορφή του υποδείγματος σταθερών ελαστικοτήτων μετατρέπεται σε γραμμική, ως εξής:

$$Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} \dots X_k^{\beta_k} u_i \Rightarrow$$
$$logY_i = \log\left(\beta_0 X_1^{\beta_1} X_2^{\beta_2} \dots X_k^{\beta_k} u_i\right) \Rightarrow$$
$$logY_i = \log(\beta_0) + \log\left(X_1^{\beta_1}\right) + \log\left(X_2^{\beta_2}\right) + \cdots + \log\left(X_k^{\beta_k}\right) + \log(u_i) \Rightarrow$$
$$Y_i^* = \beta_0^* + X_1^* + X_2^* + \cdots + X_k^* + u_i^*$$

Ειδική περίπτωση της συνάρτησης σταθερών ελαστικοτήτων αποτελεί η συνάρτηση Cobb- Douglas, η οποία έχει μια εξαρτημένη μεταβλητή Υ, που περιγράφει το προϊόν, και οι δυο ανεξάρτητες μεταβλητές X_1, που περιγράφει την εργασία και η μεταβλητή X_2, που περιγράφει το κεφάλαιο.

2.5.4 Συνάρτηση Εκθετικής Μορφής

Η Εκθετική Συνάρτηση είναι της μορφής:

$$Y_i = e^{\beta_0+\beta_1 X+u}$$

Η συναρτήσεις αυτής της μορφής μετατρέπονται σε γραμμικές συναρτήσεις, εάν εφαρμόσουμε τον εξής μετασχηματισμό:

$$Y_i^* = lnY_i$$

$$\ln\left(e^{\beta_0+\beta_1 X+u}\right) = \beta_0 + \beta_1 X + u$$

Τέτοιου είδους υποδείγματα χρησιμοποιούνται για την περιγραφή φαινομένων προόδου, όπως για παράδειγμα η πρόοδος της επιστήμης, η αύξηση του αριθμού των επιστημόνων κλπ.

2.5.5 Συνάρτηση Λογιστικής Μορφής

Τα υποδείγματα Λογιστικής καμπύλης έχουν την εξής μορφή:

$Y_i = \frac{\gamma}{1+e^{\beta_0+\beta_1 X}}$, με γ>0 και β_1<0

Η λογιστική καμπύλη χρησιμοποιείται για να περιγράψει φαινόμενα ανάπτυξης, εξάπλωσης, μεγέθυνσης κλπ. ενός προϊόντος, είτε γενικά ενός κλάδου.

Κεφάλαιο 3. Απλά γραμμικά μοντέλα παλινδρόμησης

3.1. Βασικές έννοιες

Όπως έχουμε ήδη αναφέρει και στα προηγούμενα κεφάλαια, η έννοια της απλής γραμμικής παλινδρόμησης αφορά την εκτίμηση ενός υποδείγματος, το οποίο έχει γραμμική μορφή, δηλαδή πρόκειται για την προσαρμογή των δεδομένων πάνω σε ευθεία, έχοντας ως δείγμα n παρατηρήσεων από ένα ζεύγος μιας ανεξάρτητης μεταβλητής και μιας εξαρτημένης μεταβλητής.

Ας θεωρήσουμε για παράδειγμα ένα δείγμα 20 παρατηρήσεων από τις μεταβλητές του ετήσιου εισοδήματος 20 νοικοκυριών (ανεξάρτητη μεταβλητή X) και της ετήσιας κατανάλωσης των ίδιων 20 νοικοκυριών(εξαρτημένη μεταβλητή Y), ενώ με u_i συμβολίζεται τα κατάλοιπα του υποδείγματος, όπως έχουμε περιγράψει και παραπάνω (Χρήστου, 2008).

Μέσα από αυτό το δείγμα θέλουμε να εκτιμήσουμε ένα γραμμικό υπόδειγμα της μορφής:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Αναζητούμε την εκτίμηση των συντελεστών της παλινδρόμησης, β_0και β_1.

Στον επόμενο πίνακα παρουσιάζονται οι 20 τιμές του δείγματος των δυο μεταβλητών εισοδήματος (X) και κατανάλωσης (Y). Οι τιμές αφορούν ετήσια δεδομένα και δίνονται σε χιλιάδες ευρώ.

Κατανάλωση (Y)	**Εισόδημα (X)**
5,00 €	7,00 €
5,20 €	7,60 €
5,90 €	8,10 €
6,00 €	8,40 €
6,30 €	8,50 €
5,70 €	9,20 €
6,50 €	9,30 €
6,50 €	10,00 €
7,00 €	10,60 €
6,80 €	10,90 €
6,80 €	11,30 €
7,20 €	11,50 €
7,60 €	12,40 €
8,00 €	13,00 €
8,50 €	13,30 €
8,60 €	14,00 €
9,00 €	14,00 €
9,00 €	14,40 €
10,00 €	16,00 €
12,00 €	17,00 €

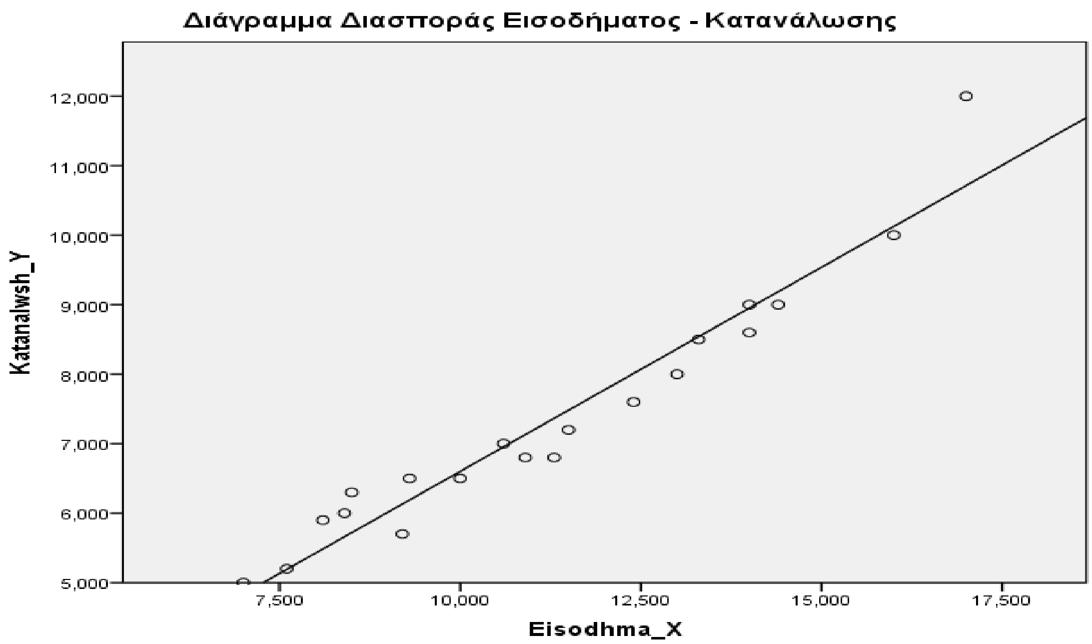
Πίνακας 1. Πίνακας τιμών Κατανάλωσης και Εισοδήματος 20 νοικοκυριών για την εκτίμηση απλού υποδείγματος γραμμικής παλινδρόμησης.

Στο υπόδειγμα παλινδρόμησης που θα προσπαθήσουμε να εκτιμήσουμε, θεωρούμε ότι η μοναδική μεταβλητή που επηρεάζει την κατανάλωση είναι το εισόδημα του κάθε νοικοκυριού.

3.2 Διάγραμμα διασποράς

Το διάγραμμα διασποράς είναι το γράφημα το οποίο χρησιμοποιείται για να απεικονίσει τη σχέση, εάν υπάρχει, μεταξύ δυο μεταβλητών X και Y. Στον οριζόντιο άξονα του γραφήματος τοποθετούνται οι τιμές της ανεξάρτητης μεταβλητής X, ενώ στον κατακόρυφο άξονα τοποθετούνται οι τιμές της εξαρτημένης μεταβλητής Y και κάθε σημείο του γραφήματος απεικονίζει ένα ζευγάρι τιμών (X, Y).

Το διάγραμμα διασποράς των τιμών του Πίνακα 1, για την Κατανάλωση και το Εισόδημα είναι το εξής:



Εικόνα 1. Διάγραμμα Διασποράς μεταξύ των μεταβλητών του Εισοδήματος και της Κατανάλωσης των 20 νοικοκυριών του δείγματος

Μπορούμε να διαπιστώσουμε από το γράφημα της Εικόνας 1 ότι για τις μεγαλύτερες τιμές εισοδήματος, προκύπτουν και μεγαλύτερες τιμές κατανάλωσης. Άρα όσο αυξάνει το ετήσιο εισόδημα των νοικοκυριών, αυξάνει και η ετήσια κατανάλωσή τους. Ταυτόχρονα, στις μικρότερες τιμές εισοδήματος παρατηρούνται και μικρότερες τιμές κατανάλωσης.

Σύμφωνα με τις παραπάνω παρατηρήσεις και με τη μορφή του διαγράμματος διασποράς που φαίνεται να έχει θετική κλίση, μπορούμε να πούμε ότι οι δυο μεταβλητές που εξετάζουμε, το εισόδημα και η κατανάλωση παρουσιάζουν θετική γραμμική συσχέτιση, καθώς μπορούμε να παρατηρήσουμε ότι τα σημεία του γραφήματος κινούνται λίγο πάνω και λίγο κάτω από την ευθεία γραμμή που έχουμε σχηματίσει.

Το διάγραμμα διασποράς θα έδειχνε την αρνητική συσχέτιση μεταξύ των δυο μεταβλητών εάν για μεγαλύτερες τιμές εισοδήματος παρατηρούνταν μικρότερες τιμές κατανάλωσης, και αντίστροφα, εάν στις μικρότερες τιμές εισοδήματος παρατηρούσαμε μεγαλύτερες τιμές κατανάλωσης. Ταυτόχρονα, στο γράφημα θα εμφανίζονταν τα σημεία να κινούνται πάνω και κάτω από την ευθεία με αρνητική κλίση.

Τέλος, εάν τα σημεία δεν εμφάνιζαν κάποια σαφή πορεία, θετική είτε αρνητική, αλλά σχημάτιζαν κάποιο ασαφές νέφος, είτε τα σημεία δεν κινούνταν πάνω σε μια ευθεία, τότε θα θεωρούσαμε ότι οι δυο μεταβλητές του γραφήματος δεν συσχετίζονται γραμμικά και θα τις θεωρούσαμε μεταξύ τους ασυσχέτιστες.

3.2.1. Συντελεστής Συσχέτισης

Από το διάγραμμα διασποράς μπορούμε να παρατηρήσουμε την ύπαρξη συσχέτισης και να την χαρακτηρίσουμε θετική είτε αρνητική. Για να υπολογίσουμε την ισχύ της συσχέτισης των δυο μεταβλητών χρησιμοποιούμε ένα χρήσιμο μέτρο το οποίο ονομάζεται συντελεστής συσχέτισης.

Ο συντελεστής συσχέτισης, που θα υπολογίσουμε με τα δεδομένα του δείγματος, συμβολίζεται με rκαι δίνεται από τον εξής τύπο:

$$r_{xy} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Ο συντελεστής συσχέτισης παίρνει τιμές από -1 μέχρι 1, $-1 \leq r_{xy} \leq 1$. Για τιμές του συντελεστή μεγαλύτερες του 0, η συσχέτιση μεταξύ των δυο μεταβλητών είναι θετική και όσο πιο κοντά στη μονάδα πλησιάζουν οι τιμές του συντελεστή τόσο πιο ισχυρή θετική είναι η συσχέτιση των δυο μεταβλητών.

Αντίθετα, οι αρνητικές τιμές του συντελεστή δείχνουν την αρνητική συσχέτιση των δυο μεταβλητών και όσο πιο κοντά στο -1 πλησιάζουν οι τιμές τόσο πιο ισχυρή αρνητική συσχέτιση έχουν οι δυο μεταβλητές.

Στο παράδειγμά μας, για να υπολογίσουμε το συντελεστή συσχέτισης χρησιμοποιούμε το SPSS το οποίο μας δίνει τον εξής πίνακα:

Συσχετίσεις			
		Κατανάλωση_Y	Εισόδημα_X
Κατανάλωση_Y	Συσχέτιση Pearson	1	,968**
	P-τιμή ελέγχου		,000
	Μέγεθος δείγματος	20	20
Εισόδημα_X	Συσχέτιση Pearson	,968**	1
	P-τιμή ελέγχου	,000	
	Μέγεθος δείγματος	20	20
**. Η συσχέτιση είναι στατιστικά σημαντική στο επίπεδο σημαντικότητας 0,01			

Πίνακας 2. Υπολογισμός συντελεστή συσχέτισης μεταξύ των μεταβλητών Εισοδήματος και Κατανάλωσης

Από τον Πίνακα 2 που παρουσιάζεται παραπάνω προκύπτει ότι ο συντελεστής συσχέτισης υπολογίζεται σε 0,968. Αυτή η τιμή του συντελεστή δείχνει μια αρκετά ισχυρή θετική συσχέτιση, καθώς η τιμή του συντελεστή είναι θετική (έχει θετικό πρόσημο) και η τιμή του είναι πολύ κοντά στη μονάδα.

Τα αποτελέσματα που προέκυψαν από τον υπολογισμό του συντελεστή συσχέτισης ήταν τα αναμενόμενα, καθώς από το διάγραμμα διασποράς των δυο μεταβλητών φαίνεται ήδη η θετική συσχέτιση των μεταβλητών. Αυτό που δεν φαίνεται είναι η ισχύς της συσχέτισης, η οποία τελικά υπολογίζεται να είναι αρκετά ισχυρή.

3.3 Μέθοδος ελαχίστων τετραγώνων

Έστω το υπόδειγμα απλής παλινδρόμησης, της μορφής:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Σε ένα δείγμα n παρατηρήσεων, με ένα αντίστοιχο διάγραμμα διασποράς όπως το γράφημα της Εικόνας 1, έχουμε σαν αντικειμενικό στόχο τη δημιουργία ευθειών παλινδρόμησης τέτοιες ώστε να περνούν ανάμεσα από τα σημεία των παρατηρήσεων. Η μέθοδος ελαχίστων τετραγώνων είναι εκείνη που θα δώσει την ευθεία παλινδρόμησης η οποία θα περνάει ανάμεσα από τα σημεία με τέτοιο τρόπο ώστε η απόσταση των σημείων από την ευθεία να είναι ελάχιστη. Άρα το άθροισμα τετραγώνων των καταλοίπων θα είναι το ελάχιστο (Χρήστου, 2008).

Επίσης, η μέθοδος ελαχίστων τετραγώνων είναι η πιο συνηθισμένη μέθοδος που θα δώσει την εκτίμηση των παραμέτρων του υποδείγματος απλής παλινδρόμησης, β_0 και β_1, καθώς σαν μέθοδος είναι σχετικά απλή στη χρήση της και οι εκτιμητές που προκύπτουν από τους υπολογισμούς της μεθόδου έχουν ιδιότητες που είναι επιθυμητές, όπως θα συζητήσουμε και παρακάτω, στο θεώρημα Gauss-Markov.

Η απλότητα της μεθόδου αφορά την εφαρμογή της, καθώς υπάρχουν έτοιμοι τύποι υπολογισμού των εκτιμήσεων των παραμέτρων β_0 και β_1, οι οποίοι είναι οι εξής:

$$\beta_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

και

$$\beta_0 = \overline{Y} - \beta_1 \overline{X}$$

Όπου $\overline{X}$ και $\overline{Y}$ οι μέσες τιμές των X_i και Y_i αντίστοιχα, οι οποίες δίνονται από τους εξής τύπους:

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

και

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

3.4 Απλό γραμμικό υπόδειγμα

3.4.1 Εκτίμηση Συντελεστών Παλινδρόμησης – Εκτίμηση Απλού Γραμμικού Υποδείγματος Παλινδρόμησης

Στην παράγραφο αυτή της εργασίας, θα χρησιμοποιήσουμε τη μέθοδο ελαχίστων τετραγώνων όπως περιεγράφηκε παραπάνω, για να εκτιμήσουμε τις παραμέτρους της ευθείας παλινδρόμησης, θα ερμηνεύσουμε την εκτίμηση των παραμέτρων και θα χρησιμοποιήσουμε την εκτίμηση του γραμμικού υποδείγματος για να κάνουμε πρόβλεψη.

Με τη χρήση του SPSS, μπορούμε να υπολογίσουμε την εκτίμηση των παραμέτρων της παλινδρόμησης στο παράδειγμα του εισοδήματος και της κατανάλωσης των 20 νοικοκυριών. Οι εκτιμήσεις των συντελεστών β_0 και β_1, δίνονται στον επόμενο πίνακα.

Συντελεστές

Υπόδειγμα		Εκτίμηση Συντελεστών B	Εκτίμηση Συντελεστών Τυπικό Σφάλμα	Τυποποιημένοι Συντελεστές B	t-στατιστική	P-τιμή Ελέγχου
1	Σταθερά	,727	,416		1,746	,098
	Εισόδημα_X	,587	,036	,968	16,456	,000

Πίνακας 3. Υπολογισμός εκτίμησης συντελεστών του απλού γραμμικού υποδείγματος, της ανεξάρτητης μεταβλητής του εισοδήματος και της εξαρτημένης μεταβλητής της κατανάλωσης.

Σύμφωνα με τον πίνακα 3, η εκτίμηση των συντελεστών είναι η εξής:

$\beta_0 = 0{,}727$

και

$\beta_1 = 0{,}587$

Αντίστοιχα, η εκτίμηση της ευθείας παλινδρόμησης του απλού γραμμικού υποδείγματος εισοδήματος – κατανάλωσης που εξετάζουμε είναι το εξής:

$$Y_i = 0{,}727 + 0{,}587X_i$$

3.4.2 Πρόβλεψη με το Απλό Γραμμικό Υπόδειγμα

Μπορούμε τώρα να χρησιμοποιήσουμε την εξίσωση που εκτιμήσαμε για να κάνουμε πρόβλεψη, για όλες τις τιμές της ανεξάρτητης μεταβλητής του εισοδήματος, ώστε να μας επιστρέφονται οι εξαρτημένες μεταβλητές της κατανάλωσης. Η πρόβλεψη μπορεί να γίνει και για τιμές της μεταβλητής του εισοδήματος X εκτός του δείγματος, δηλαδή στο συγκεκριμένο παράδειγμα μπορούμε να δώσουμε τιμές μικρότερες από τη μικρότερη τιμή του δείγματος που είναι 5.000 € και τιμές μεγαλύτερες από την μέγιστη τιμή του δείγματος που είναι 12.000 €.

Για παράδειγμα, κάποιο νοικοκυριό που έχει ετήσιο εισόδημα 3.000 € (τιμή μικρότερη της ελάχιστης τιμή εισοδήματος του δείγματος), δηλαδή τιμή της μεταβλητής του εισοδήματος X ίση με 3, η οποία αντιστοιχεί σε ετήσια κατανάλωση

Y $= 0{,}727 + 0{,}587 * 3 = 0{,}727 + 1{,}761 = 2{,}488$.

Δηλαδή, κάποιο νοικοκυριό με ετήσιο εισόδημα 3.000 € θα καταναλώνει ετησίως 2.488 €.

Αντίστοιχα, κάποιο νοικοκυριό με ετήσιο εισόδημα 20.000 € (τιμή μεγαλύτερη της μέγιστης τιμή εισοδήματος του δείγματος), δηλαδή τιμή της μεταβλητής του εισοδήματος X ίση με 20, η οποία αντιστοιχεί σε ετήσια κατανάλωση

Y $= 0{,}727 + 0{,}587 * 20 = 0{,}727 + 11{,}74 = 12{,}467$.

Δηλαδή, κάποιο νοικοκυριό με ετήσιο εισόδημα 20.000 € θα καταναλώνει ετησίως 12.467 €.

Στον επόμενο πίνακα παρουσιάζονται οι πραγματικές (παρατηρούμενες) τιμές σε σχέση με τις τιμές που προβλέπονται από το υπόδειγμα που εκτιμήθηκε. Ενδιαφέρον παρουσιάζει και η τελευταία στήλη στην οποία παρουσιάζονται τα κατάλοιπα, δηλαδή η διαφορά της παρατηρούμενης κατανάλωσης για το κάθε επίπεδο εισοδήματος, από την κατανάλωση που προβλέπει το υπόδειγμα.

Παρατηρούμενη Κατανάλωση (Y)	Εισόδημα (X)	Προβλεπόμενη Κατανάλωση ($\hat{Y}$)	Κατάλοιπα ($\hat{u}$)
5,00 €	7,00 €	4,84 €	0,16 €
5,20 €	7,60 €	5,19 €	0,01 €
5,90 €	8,10 €	5,48 €	0,42 €
6,00 €	8,40 €	5,66 €	0,34 €
6,30 €	8,50 €	5,72 €	0,58 €
5,70 €	9,20 €	6,13 €	-0,43 €
6,50 €	9,30 €	6,19 €	0,31 €
6,50 €	10,00 €	6,60 €	-0,10 €
7,00 €	10,60 €	6,95 €	0,05 €
6,80 €	10,90 €	7,13 €	-0,33 €
6,80 €	11,30 €	7,36 €	-0,56 €
7,20 €	11,50 €	7,48 €	-0,28 €
7,60 €	12,40 €	8,01 €	-0,41 €
8,00 €	13,00 €	8,36 €	-0,36 €
8,50 €	13,30 €	8,53 €	-0,03 €
8,60 €	14,00 €	8,95 €	-0,35 €
9,00 €	14,00 €	8,95 €	0,05 €
9,00 €	14,40 €	9,18 €	-0,18 €
10,00 €	16,00 €	10,12 €	-0,12 €
12,00 €	17,00 €	10,71 €	1,29 €

Πίνακας 4. Τιμές Παρατηρούμενης Κατανάλωσης, Προβλεπόμενης Κατανάλωσης και Καταλοίπων για τις διάφορες τιμές του Εισοδήματος.

Από τον πίνακα 4 μπορούμε να παρατηρήσουμε ότι οι τιμές που προβλέπει το εισόδημα είναι αρκετά κοντά με τις τιμές που παρατηρούνται από το δείγμα των 20 νοικοκυριών.

Παράλληλα, από το υπόδειγμα μπορούμε να πάρουμε τιμές και εκτός δείγματος, όπως έχουμε ήδη αναφέρει, για να τις χρησιμοποιήσουμε σαν πρόβλεψη.

3.4.3 Ερμηνεία Συντελεστών Απλού Γραμμικού Υποδείγματος

Οι συντελεστές του απλού γραμμικού παίζουν σημαντικό ρόλο, όπως είδαμε και έχουν ξεχωριστή ερμηνεία ο καθένας.

Πιο συγκεκριμένα, ο σταθερός όρος της παλινδρόμησης, ο συντελεστής β_0 δείχνει την τιμή που παίρνει η εξαρτημένη μεταβλητή Υ, όταν η ανεξάρτητη μεταβλητή της παλινδρόμησης πάρει την τιμή μηδέν.

Στο δικό μας παράδειγμα, ο συντελεστής β_0 δείχνει την τιμή της κατανάλωσης, όταν ένα νοικοκυριόδεν έχει καθόλου εισόδημα. Η τιμή του συντελεστή β_0 είναι ίση με 0,727, άρα παρότι ένα νοικοκυριό δεν έχει καθόλου εισοδήματα, θα πρέπει να καταναλώσει σε ένα έτος 727 €.

Αντίστοιχα, ο συντελεστής της ανεξάρτητης μεταβλητής β_1 δείχνει τη μεταβολή της εξαρτημένης μεταβλητής, όταν η ανεξάρτητη μεταβλητή αυξηθεί κατά μια μονάδα. Η μεταβολή της εξαρτημένης μεταβλητής εξαρτάται από το πρόσημο του β_1. Όταν το πρόσημο του β_1 είναι θετικό, τότε όταν η ανεξάρτητη μεταβλητή αυξάνεται, αυξάνεται και η εξαρτημένη μεταβλητή, ενώ όταν το πρόσημο του β_1 είναι αρνητικό, τότε η ανεξάρτητη μεταβλητή μειώνεται.

Στο παράδειγμα που μελετάμε, η τιμή του συντελεστή β_1 δείχνει τη μεταβολή στην ετήσια κατανάλωση, όταν το ετήσιο εισόδημα ενός νοικοκυριού αυξηθεί κατά μια μονάδα.

Άρα, όταν το εισόδημα αυξηθεί κατά μια μονάδα (1.000 €), η ετήσια κατανάλωση θα αυξηθεί κατά 0,587 μονάδες (587 €).

3.5 Υποθέσεις απλού γραμμικού υποδείγματος

Ας θεωρήσουμε το απλό γραμμικό υπόδειγμα της μορφής

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Η γραμμική αύτη σχέση διέπεται από συγκεκριμένες υποθέσεις, οι οποίες είναι οι εξής (Χρήστου, 2008):

1. Η αναμενόμενη τιμή των καταλοίπων είναι μηδέν. Δηλαδή, η τιμή των καταλοίπων, δηλαδή των αποστάσεων της εκτίμησης της ανεξάρτητης μεταβλητής από την πραγματική τιμή της ανεξάρτητης μεταβλητής μπορεί να πάρει και αρνητικές και θετικές τιμές, αλλά κατά μέσο όρο η τιμή αυτών των αποστάσεων είναι μηδέν.
   Συγκεκριμένα,

   $E(u_i) = 0$

2. Η διακύμανση των καταλοίπων παραμένει σταθερή και ίση με σ^2. Αυτή είναι η υπόθεση της ομοσκεδαστικότητας. Όταν παραβιάζεται η υπόθεση αυτή παρουσιάζεται ετεροσκεδαστικότητα στα κατάλοιπα. Πρόκειται για ένα πρόβλημα το οποίο θα αναλύσουμε σε επόμενο κεφάλαιο της εργασίας.
   Η υπόθεση της ομοσκεδαστικότητας συμβολίζεται ως εξής:

   $Var(u_i) = \sigma^2$.

3. Η συνδιακύμανση των καταλοίπων, για κατάλοιπα διαφορετικών χρονικών στιγμών ισούται με μηδέν. Δηλαδή, σε διαφορετικές χρονικές στιγμές, οι όροι των καταλοίπων δεν συσχετίζονται μεταξύ τους.
   Ο συμβολισμός της υπόθεσης αυτής είναι ο εξής:

   $Cov\ (u_i, u_j) = 0, \forall\ i \neq j$

Για τις τιμές των καταλοίπων σε ίδιες χρονικές στιγμές, δηλαδή όταν $i = j$, τότε η συνδιακύμανση των καταλοίπων γίνεται Cov (u_i,u_i), η οποία ισούται με τη διακύμανση των καταλοίπων και η οποία, λόγω της προηγούμενης υπόθεσης πρέπει να παραμένει σταθερή και ίση με σ^2.
Ο συμβολισμός είναι ο εξής:

$\forall\ i \neq j$, Cov (u_i,u_j) =Cov (u_i,u_i) =Var (u_i) = σ^2.

4. Η εκτίμηση των παραμέτρων του απλού γραμμικού υποδείγματος, β_0καιβ_1, παραμένει σταθερή σε όλη τη διάρκεια της μελέτης. Αυτό σημαίνει ότι μετά την εφαρμογή της μεθόδου ελαχίστων τετραγώνων και αφού εκτιμηθεί η τιμή των δυο παραμέτρων, θεωρούμε ότι τιμές αυτές δεν μπορεί να μεταβληθούν.

5. Ο διαταρακτικός όρος (κατάλοιπα) u_i, είναι τυχαία μεταβλητή, η οποία ακολουθεί την κανονική κατανομή, με μέση τιμή μηδέν και διακύμανση σ^2.

$$u_i \sim N(0, \sigma^2)$$

6. Οι τιμές που παίρνει η επεξηγηματική μεταβλητή X παραμένουν σταθερές και μεταξύ τους δεν είναι όλες ίσες. Άρα, η επεξηγηματική μεταβλητή δεν είναι στοχαστική, αλλά ντετερμινιστική.

3.6 Κανονικές εξισώσεις - Ευθεία ελαχίστων τετραγώνων (ευθεία παλινδρόμησης)

3.6.1 Κανονικές Εξισώσεις

Η μέθοδος ελαχίστων τετραγώνων προκύπτει από το σκεπτικό ότι η ευθεία ελαχίστων τετραγώνων θα περνάει ανάμεσα στα σημεία με τέτοιο τρόπο ώστε να ελαχιστοποιούνται οι αποστάσεις των σημείων από την ευθεία, δηλαδή τα κατάλοιπα, και για την ακρίβεια από την ελαχιστοποίηση του αθροίσματος των τετραγώνων των καταλοίπων.

Πιο συγκεκριμένα, η συνάρτηση που ελαχιστοποιούμε είναι η εξής:

$$\Phi = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (Y_i - \widehat{Y_\iota})^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_\iota)^2$$

Η ελαχιστοποίηση της παραπάνω συνάρτησης Φ γίνεται παραγωγίζοντας την Φ, ως προς τους δυο αγνώστους, $\beta_0 και\ \beta_1$ και θέτοντας την πρώτη παράγωγό της ίση με το μηδέν. Οι δυο εξισώσεις που προκύπτουν ονομάζονται Κανονικές Εξισώσεις.

Η παράγωγος της Φ ως προς β_0 είναι η εξής:

$$\frac{\partial \Phi}{\partial \beta_0} = 0 \Rightarrow -2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_\iota) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_\iota) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (Y_i) - \sum_{i=1}^{n} \beta_0 - \sum_{i=1}^{n} \beta_1 X_\iota = 0$$

$$\Rightarrow \sum_{i=1}^{n} (Y_i) = n\beta_0 + \beta_1 \sum_{i=1}^{n} X_\iota$$

Η τελευταία εξίσωση που προέκυψε μετά από τις πράξεις είναι η πρώτη Κανονική Εξίσωση.

Μάλιστα, εάν διαιρέσουμε κατά μέλη με το μέγεθος του δείγματος n, θα προκύψει ο τύπος υπολογισμού του συντελεστή β_0.

$$\sum_{i=1}^{n}(Y_i) = n\beta_0 + \beta_1 \sum_{i=1}^{n} X_\iota \Rightarrow$$

$$\frac{\sum_{i=1}^{n}(Y_i)}{n} = \frac{n\beta_0}{n} + \frac{\beta_1 \sum_{i=1}^{n} X_\iota}{n} \Rightarrow$$

$$\bar{Y} = \beta_0 + \beta_1 \bar{X} \Rightarrow$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

Η παράγωγος της Φ ως προς β_1 είναι η εξής:

$$\frac{\partial \Phi}{\partial \beta_1} = 0 \Rightarrow -2\sum_{i=1}^{n} X_\iota(Y_i - \beta_0 - \beta_1 X_\iota) = 0$$

$$\Rightarrow \sum_{i=1}^{n} X_\iota(Y_i - \beta_0 - \beta_1 X_\iota) = 0$$

$$\Rightarrow \sum_{i=1}^{n} X_\iota Y_i - \beta_0 \sum_{i=1}^{n} X_\iota - \beta_1 \sum_{i=1}^{n} X_i^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n} X_\iota Y_i = \beta_0 \sum_{i=1}^{n} X_\iota + \beta_1 \sum_{i=1}^{n} X_i^2$$

Η τελευταία εξίσωση είναι η Δεύτερη Κανονική Εξίσωση.

Εάν στην εξίσωση αυτή αντικαταστήσουμε στο β_0την έκφραση που βρήκαμε από την πρώτη Κανονική εξίσωση και επιλύσουμε ως προς το β_1 θα προκύψει ο τύπος υπολογισμού του β_1, ο οποίος είναι ο εξής:

$$\beta_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

3.6.2 Ιδιότητες της Ευθείας Παλινδρόμησης

Η εκτίμηση της ευθείας παλινδρόμησης που προκύπτει μετά τη χρήση της μεθόδου ελαχίστων τετραγώνων, παρουσιάζει τις παρακάτω ιδιότητες:

1. Το σημείο που δημιουργείται από τους μέσους της ανεξάρτητης και της εξαρτημένης μεταβλητής, επαληθεύεται από την ευθεία παλινδρόμησης.
   Δηλαδή η ευθεία περνάει από το σημείο $(\bar{X}, \bar{Y})$ και ισχύει:

$$\bar{Y} = \beta_0 + \beta_1 \bar{X}$$

2. Το άθροισμα των τιμών των εκτιμήσεων της εξαρτημένης μεταβλητής που προκύπτουν από τη μέθοδο ελαχίστων τετραγώνων, είναι ίσο με το άθροισμα των πραγματικών τιμών της εξαρτημένης μεταβλητής.
   Δηλαδή,

$$\sum_{i=1}^{n} \hat{Y} = \sum_{i=1}^{n} Y$$

3. Από την παραπάνω σχέση προκύπτει ότι τα κατάλοιπα αθροίζουν στο μηδέν, καθώς:

$$\sum_{i=1}^{n} \hat{Y} = \sum_{i=1}^{n} Y \Rightarrow$$

$$\sum_{i=1}^{n} \hat{Y} - \sum_{i=1}^{n} Y = 0 \Rightarrow$$

$$\sum_{i=1}^{n} \left(\hat{Y} - Y\right) = 0 \Rightarrow$$

$$\sum_{i=1}^{n} \hat{u}_i = 0$$

4. Το άθροισμα των γινομένων των καταλοίπων επί τις τιμές της ανεξάρτητης μεταβλητής είναι ίσο με μηδέν. Δηλαδή,

$$\sum_{i=1}^{n} \hat{u}_i X_i = 0$$

Η σχέση αυτή προκύπτει εάν πολλαπλασιάσουμε τα κατάλοιπα επί τη μεταβλητή X, ως εξής:

$$\hat{u}_i = Y_i - \widehat{Y_\iota}$$

$$\hat{u}_i = Y_i - \beta_0 - \beta_1 X_\iota \Rightarrow$$

$$\hat{u}_i X = Y_i X - \beta_0 X - \beta_1 {X_\iota}^2 \Rightarrow$$

Εάν, τώρα αθροίσουμε τα δύο μέλη, σύμφωνα με τη σχέση:

$$\sum_{i=1}^{n} \hat{u}_i X_i = \sum_{i=1}^{n} Y_\iota X_i - \beta_0 \sum_{i=1}^{n} X_\iota - \beta_1 \sum_{i=1}^{n} {X_i}^2 \Rightarrow$$

$$\sum_{i=1}^{n} \hat{u}_i X_i = 0$$

Το αποτέλεσμα αυτό προκύπτει από τη δεύτερη κανονική εξίσωση, καθώς το δεύτερο μέλος που είναι η δεύτερη κανονική εξίσωση γνωρίζουμε ότι είναι ίσο με μηδέν. Άρα, και το πρώτο μέλος λόγω της ισότητας θα είναι ίσο με μηδέν.

5. Το άθροισμα των γινομένων των καταλοίπων επί τις τιμές των εκτιμήσεων της εξαρτημένης μεταβλητής είναι ίσο με μηδέν. Δηλαδή,

$$\sum_{i=1}^{n} \hat{u}_i \widehat{Y}_\iota = 0$$

Η σχέση αυτή προκύπτει από τη χρήση της τρίτης και τέταρτης ιδιότητας της ευθείας παλινδρόμησης, ως εξής:

$$\sum_{i=1}^{n} \hat{u}_i \widehat{Y}_\iota = \sum_{i=1}^{n} \hat{u}_i (\beta_0 - \beta_1 X_\iota) =$$

$= \beta_0 \sum_{i=1}^{n} \hat{u}_i - \beta_1 \sum_{i=1}^{n} \hat{u}_i \, X_\iota = 0.$

3.7 Θεώρημα Gauss-Markov

Σύμφωνα με το θεώρημα Gauss–Markov, οι εκτιμητές του απλού γραμμικού υποδείγματος που θα προκύψουν από τη χρήση της μεθόδου παλινδρόμησης έχουν μερικές πολύ χρήσιμες ιδιότητες, οι οποίες είναι οι παρακάτω:

1. Οι εκτιμητές είναι γραμμικές συναρτήσεις των τιμών της εξαρτημένης μεταβλητής Y.

   Δηλαδή, ο εκτιμητής ελαχίστων τετραγώνων, που συμβολίζεται με $\hat{\theta}$, θα είναι της μορφής:

   $$\hat{\theta} = \sum_{i=1}^{n} a_i Y_i$$

   Όπου a_iείναι σταθερές για όλα τα i.

2. Οι εκτιμητές έχουν την ιδιότητα της αμεροληψίας, δηλαδή η αναμενόμενη τιμή του εκτιμητή για τον κάθε συντελεστή είναι ίση με την πραγματική του συντελεστή.

   Για τους εκτιμητές ελαχίστων τετραγώνων του απλού γραμμικού υποδείγματος, ισχύει ότι:

   $$E(\hat{\beta}_0) = \beta_0$$

   $$E(\hat{\beta}_1) = \beta_1$$

3. Ανάμεσα σε όλους τους εκτιμητές που έχουν τις δυο παραπάνω ιδιότητες της γραμμικότητας και της αμεροληψίας, οι εκτιμητές ελαχίστων τετραγώνων έχουν τη μικρότερη διακύμανση, που δίνεται από τους παρακάτω τύπους:

   $$Var(\hat{\beta}_0) = \sigma^2 \frac{\sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

   $$Var(\hat{\beta}_1) = \sigma^2 \frac{1}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

Η συνδιακύμανση των δυο συντελεστών μεταξύ τους δίνεται από την εξής σχέση:

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = -\sigma^2 \frac{\bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Συνοπτικά, βάσει του θεωρήματος Gauss–Markov, οι εκτιμητές είναι Άριστοι (Best), Γραμμικοί (Linear) και Αμερόληπτοι (Unbiased) Εκτιμητές (Estimators). Δηλαδή είναι B.L.U.E.

3.8 Συντελεστής προσδιορισμού

3.8.1 Αθροίσματα Τετραγώνων

Έχουμε ήδη αναλύσει το γεγονός ότι η πραγματική τιμή της εξαρτημένης μεταβλητής Y μπορεί να γραφεί σαν το άθροισμα του επεξηγηματικού μέρους του απλού γραμμικού υποδείγματος και του μη επεξηγηματικού μέρους το οποίο έχουμε ονομάσει υπόλοιπο.
Αναλυτικά, μπορούμε να γράψουμε τα παραπάνω ως εξής:

$$Y_i = (\hat{\beta}_o + \hat{\beta}_1 X_i) + u_i \Rightarrow$$

$$Y_i = \hat{Y} + u_i$$

Η παραπάνω εξίσωση είναι ήδη γνωστή ως το απλό γραμμικό υπόδειγμα, ενώ η δεύτερη εξίσωση είναι και αυτή γνωστή καθώς μπορούμε να γράψουμε την πραγματική τιμή της εξαρτημένης μεταβλητής ως το άθροισμα της εκτίμησης της ανεξάρτητης μεταβλητής συν ένα σφάλμα. Εναλλακτικά, θα μπορούσαμε να πούμε ότι η διαφορά της πραγματικής τιμής της εξαρτημένης μεταβλητής από την εκτίμηση της εξαρτημένης μεταβλητής είναι ίση με το υπόλοιπο.

Αν στην τελευταία εξίσωση αφαιρέσουμε κατά μέλη την αναμενόμενη τιμή της εξαρτημένης μεταβλητής $\overline{Y}$ και στη συνέχεια υψώσουμε στο τετράγωνο, θα έχουμε την παρακάτω σχέση:

$$Y_i = \hat{Y} + u_i \Rightarrow$$

$$Y_i - \overline{Y} = \hat{Y} + u_i - \overline{Y} \Rightarrow$$

$$(Y_i - \overline{Y})^2 = (\hat{Y} - \overline{Y} + u_i)^2$$

Αθροίζοντας την τελευταία σχέση κατά μέλη, έχουμε:

$$(Y_i - \overline{Y})^2 = (\hat{Y} - \overline{Y} + u_i)^2 \Rightarrow$$

$$\sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \sum_{i=1}^{n} (\hat{Y} - \overline{Y} + u_i)^2$$

Στη συνέχεια, κάνοντας το ανάπτυγμα τετραγώνων στη δεύτερη σχέση, θεωρώντας ως τα δυο μέλη του αναπτύγματος τα $(\widehat{\mathrm{Y}} - \overline{\mathrm{Y}})$ και $\mathrm{u_i}$, έχουμε:

$$\sum_{i=1}^{n}(\mathrm{Y_i} - \overline{\mathrm{Y}})^2 = \sum_{i=1}^{n}(\widehat{\mathrm{Y}} - \overline{\mathrm{Y}} + \mathrm{u_i})^2 \Rightarrow$$

$$\sum_{i=1}^{n}(\mathrm{Y_i} - \overline{\mathrm{Y}})^2 = \sum_{i=1}^{n}(\widehat{\mathrm{Y}} - \overline{\mathrm{Y}})^2 + \sum_{i=1}^{n}\mathrm{u_i}^2 + \sum_{i=1}^{n}2(\widehat{\mathrm{Y}}_\mathrm{i} - \overline{\mathrm{Y}}\,)\mathrm{u_i}$$

Ο τελευταίος όρος του αθροίσματος, σύμφωνα με τις ιδιότητες της ευθείας παλινδρόμησης, μπορούμε να δείξουμε ότι είναι ίσος με το μηδέν, καθώς:

$$\sum_{i=1}^{n}2\big(\widehat{\mathrm{Y}}_\mathrm{i} - \overline{\mathrm{Y}}\big)\mathrm{u_i} = 2\left(\sum_{i=1}^{n}\widehat{\mathrm{Y}}_\mathrm{i}\,\mathrm{u_i} - \sum_{i=1}^{n}\overline{\mathrm{Y}}\,\mathrm{u_i}\right) =$$

$$2\left(\sum_{i=1}^{n}\widehat{\mathrm{Y}}_\mathrm{i}\,\mathrm{u_i} - \overline{\mathrm{Y}}\sum_{i=1}^{n}\mathrm{u_i}\right) = 0$$

Άρα, μετά από αυτές τις πράξεις, η εξίσωση γίνεται:

$$\sum_{i=1}^{n}(\mathrm{Y_i} - \overline{\mathrm{Y}})^2 = \sum_{i=1}^{n}(\widehat{\mathrm{Y}} - \overline{\mathrm{Y}})^2 + \sum_{i=1}^{n}\mathrm{u_i}^2 + \sum_{i=1}^{n}2\big(\widehat{\mathrm{Y}}_\mathrm{i} - \overline{\mathrm{Y}}\big)\mathrm{u_i} \Rightarrow$$

$$\sum_{i=1}^{n}(\mathrm{Y_i} - \overline{\mathrm{Y}})^2 = \sum_{i=1}^{n}(\widehat{\mathrm{Y}} - \overline{\mathrm{Y}})^2 + \sum_{i=1}^{n}\mathrm{u_i}^2$$

Η παραπάνω εξίσωση αποτελείται από τρία αθροίσματα. Το πρώτο άθροισμα, το $\sum_{i=1}^{n}(\mathrm{Y_i} - \overline{\mathrm{Y}})^2$, ονομάζεται συνολικό άθροισμα τετραγώνων (Total Sum of Squares– TSS) και δείχνει τη συνολική μεταβλητότητα της εξαρτημένης μεταβλητής Y.

Το δεύτερο άθροισμα, το $\sum_{i=1}^{n}(\hat{Y}-\bar{Y})^2$, το οποίο ονομάζεται άθροισμα τετραγώνων της παλινδρόμηση, (Sum of Squares Regression–SSR), δείχνει τη μεταβλητότητα της εξαρτημένης μεταβλητής Υ, η οποία ερμηνεύεται από το υπόδειγμα παλινδρόμησης.

Τέλος, το τρίτο άθροισμα, το $\sum_{i=1}^{n} u_i{}^2$, το οποίο ονομάζεται άθροισμα τετραγώνων των καταλοίπων, (Sum of Squares Error – SSE), δείχνει τη μεταβλητότητα της εξαρτημένης μεταβλητής Υ, η οποία δεν ερμηνεύεται από το υπόδειγμα παλινδρόμησης και εξαρτάται από άλλους τυχαίους παράγοντες.

Από τα παραπάνω, μπορούμε να γράψουμε την εξής σχέση:

SST = SSR + SSE

Η σχέση αυτή δείχνει ότι το συνολικό άθροισμα τετραγώνων ισούται με το άθροισμα τετραγώνων της παλινδρόμησης και με το άθροισμα τετραγώνων των καταλοίπων. Αν διαιρέσουμε κατά μέλη την παραπάνω σχέση με το συνολικό άθροισμα τετραγώνων, SST έχουμε:

$$\frac{\mathrm{SST}}{\mathrm{SST}}=\frac{\mathrm{SSR}}{\mathrm{SST}}+\frac{\mathrm{SSE}}{\mathrm{SST}}$$

Ο λόγος $\frac{\mathrm{SSR}}{\mathrm{SST}}$ που δείχνει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής Υ, που ερμηνεύεται από το υπόδειγμα παλινδρόμησης, ονομάζεται συντελεστής προσδιορισμού και συμβολίζεται με R^2.

Ισχύει, δηλαδή:

$$R^2=\frac{\mathrm{SSR}}{\mathrm{SST}}=1-\frac{\mathrm{SSE}}{\mathrm{SST}}$$

Ο συντελεστής προσδιορισμού παίρνει τιμές ανάμεσα στο μηδέν και στο 1,

$0 \leq R^2 \leq 1$. Όσο πιο κοντά στο μηδέν είναι η τιμή του συντελεστή τόσο πιο μικρό ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής επεξηγεί το υπόδειγμα, ενώ αντίθετα, όσο πιο κοντά στο 1 βρίσκεται η τιμή του συντελεστή προσδιορισμού, τόσο πιο μεγάλο ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής επεξηγείται από το υπόδειγμα παλινδρόμησης.

Ειδικά στην περίπτωση της απλής παλινδρόμησης ισχύει και μια ακόμα σχέση που συνδέει το συντελεστή προσδιορισμού με το συντελεστή συσχέτισης ως εξής:

$$R^2 = r^2$$

Σύμφωνα με την παραπάνω σχέση, το τετράγωνο του συντελεστή συσχέτισης ισούται με το συντελεστή προσδιορισμού.

Η σχέση αυτή ισχύει μόνο στην περίπτωση της απλής παλινδρόμησης και όχι στην περίπτωση της πολλαπλής παλινδρόμησης.

3.8.2 Παράδειγμα Υπολογισμού Συντελεστή Προσδιορισμού

Επιστρέφοντας στο παραπάνω παράδειγμα, των μεταβλητών του εισοδήματος X και της κατανάλωσης Y των 20 νοικοκυριών, για τον υπολογισμό των αθροισμάτων τετραγώνων και του συντελεστή προσδιορισμού, χρησιμοποιούμε τους παρακάτω πίνακες.

ΠΙΝΑΚΑΣ ΑΝΑΥΣΗΣ ΔΙΑΚΥΜΑΝΣΗΣ						
Υπόδειγμα		Αθροίσματα Τετραγώνων	Βαθμοί Ελευθερίας	Μέσα Τετράγωνα	F-Στατιστική	P-τιμή Ελέγχου
1	Παλινδρόμηση	53,946	1	53,946	270,806	,000
	Κατάλοιπα	3,586	18	,199		
	Ολικό	57,532	19			

Πίνακας 5. Τιμές Αθροισμάτων Τετραγώνων Παλινδρόμησης, Καταλοίπων και Συνολικό, στο παράδειγμα της απλής παλινδρόμησης.

Σύμφωνα με τον πίνακα 5, το άθροισμα τετραγώνων της παλινδρόμησης είναι 53,946, το άθροισμα τετραγώνων των καταλοίπων είναι 3,586 ενώ το συνολικό άθροισμα τετραγώνων είναι 57,532.

Στον επόμενο πίνακα παρουσιάζεται η τιμή του συντελεστή προσδιορισμού της παλινδρόμησης του υποδείγματος εισοδήματος – κατανάλωσης.

Σύνοψη Υποδείγματος				
Υπόδειγμα	Συντελεστής Συσχέτισης	Συντελεστής Προσδιορισμού	Προσαρμοσμένος Συντελεστής Προσδιορισμού	Τυπικό Σφάλμα της Εκτίμησης
1	,968	,938	,934	,44633

Πίνακας 6. Υπολογισμός τιμής Συντελεστή Προσδιορισμού R^2του παραδείγματος.

Από τον πίνακα 6 προκύπτει ότι ο συντελεστής προσδιορισμού της παλινδρόμησης είναι 0,938. Είναι μια τιμή του συντελεστή πολύ κοντά στη μονάδα, άρα μπορούμε να συμπεράνουμε ότι η μεταβλητή του εισοδήματος στο υπόδειγμα απλής παλινδρόμησης επεξηγεί αρκετά καλά τη μεταβλητότητα της εξαρτημένης μεταβλητής της κατανάλωσης.

Επίσης, μπορούμε να παρατηρήσουμε ότι ο συντελεστής προσδιορισμού είναι ίσος με το τετράγωνο του συντελεστή συσχέτισης που δίνεται 0,968, καθώς $0{,}968^2 = 0{,}638$.

3.9 Επίλυση Άσκησης Απλής Γραμμικής Παλινδρόμησης Με Χρήση του SPSS

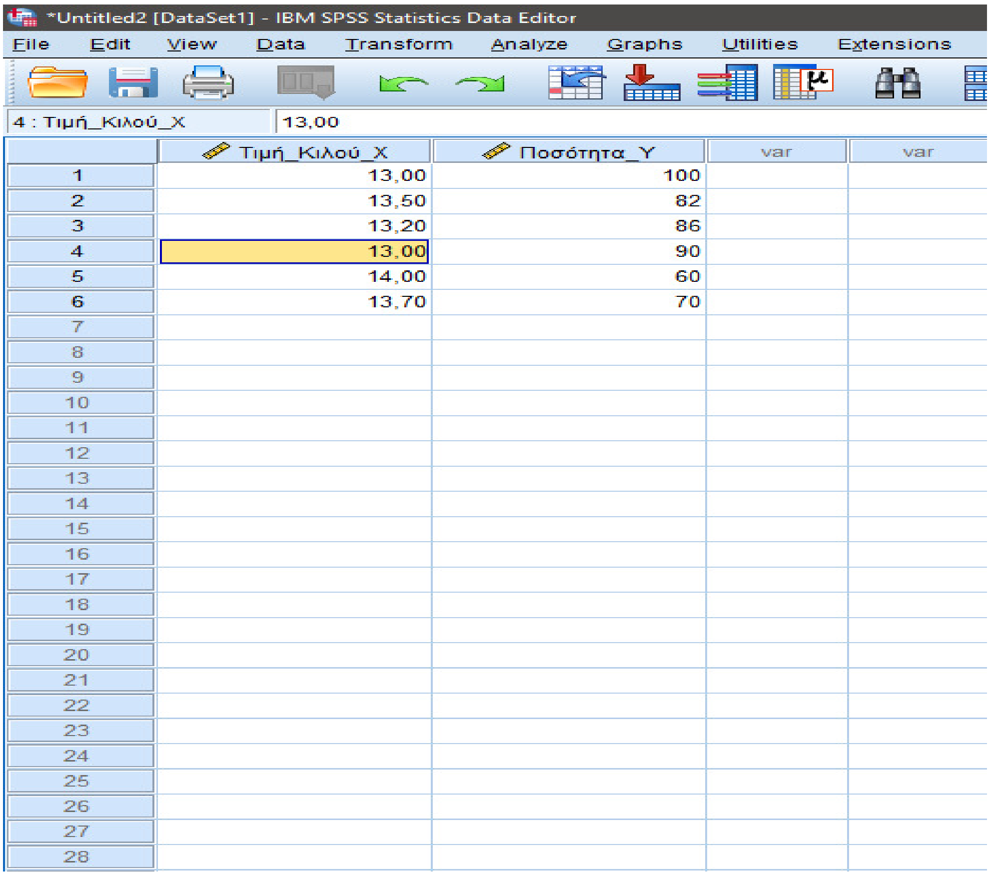
Με την πάροδο των χρόνων και με την ανάπτυξη της τεχνολογίας και κυρίως των Ηλεκτρονικών Υπολογιστών, έγινε πιο απλή και πιο ταχύτερη η διενέργεια μιας μελέτης η οποία απαιτεί τη χρήση της μεθόδου Ελαχίστων Τετραγώνων.

Στα επόμενα βήματα θα δείξουμε πως μπορεί κάποιος χρησιμοποιώντας το στατιστικό πρόγραμμα SPSS να εφαρμόσει τις απαιτούμενες μεθόδους, προσπαθώντας να επιλύσουμε την άσκηση 6 από το βιβλίο, Εισαγωγή στην Οικονομετρία (Χρήστου, 2008).

Το πρώτο βήμα που θα πρέπει να κάνουμε στο SPSS είναι να περάσουμε τα δεδομένα στην καρτέλα του Data View του SPSS.

Στη συνέχεια, στην καρτέλα του Variable View μπορούμε να αλλάξουμε τα ονόματα των μεταβλητών, σε Ποσότητα σε Κιλά (Y) και Τιμή Κιλού (X). Στην ίδια καρτέλα θα αλλάξουμε τον αριθμό των δεκαδικών ψηφίων που θα δέχεται η μεταβλητή Y σε 0.

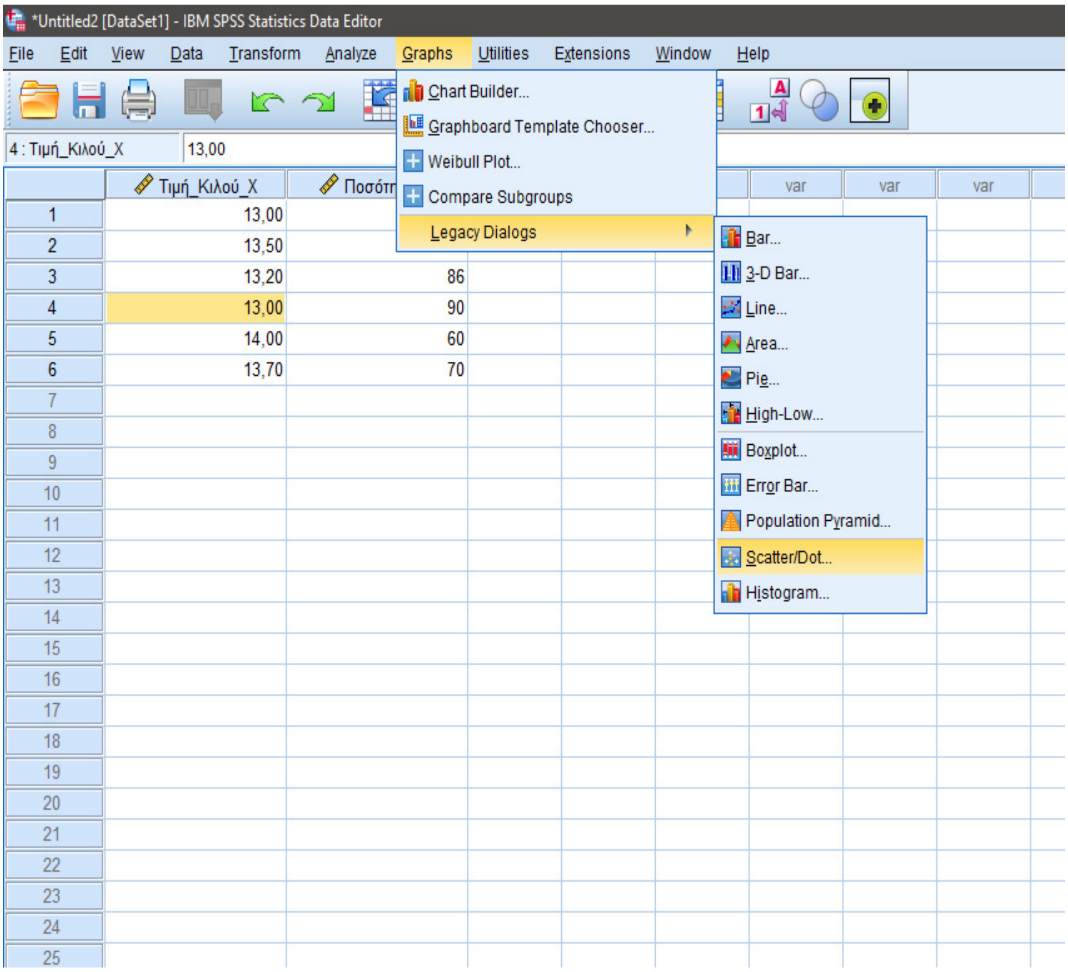
Τα αποτελέσματα παρουσιάζονται στην επόμενη εικόνα με τα δεδομένα.



Εικόνα 2. Παρατηρήσεις Ζήτησης ενός Αγαθού (Τιμή - Ποσότητα)

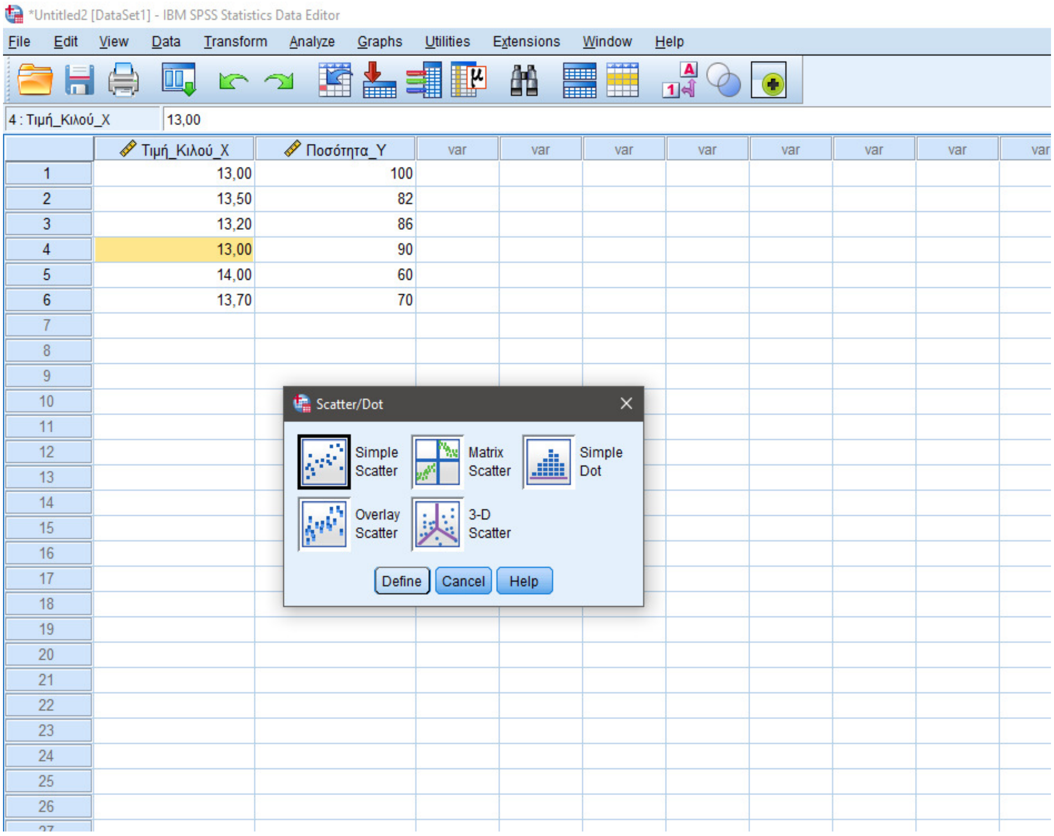
Στη συνέχεια κάνοντας χρήση του SPSS, θα φτιάξουμε ένα γράφημα διασποράς των δεδομένων.

Ακολουθώντας την εξής διαδικασία Graphs → Legacy Dialogs → Scatter/Dot όπως στην επόμενη εικόνα.



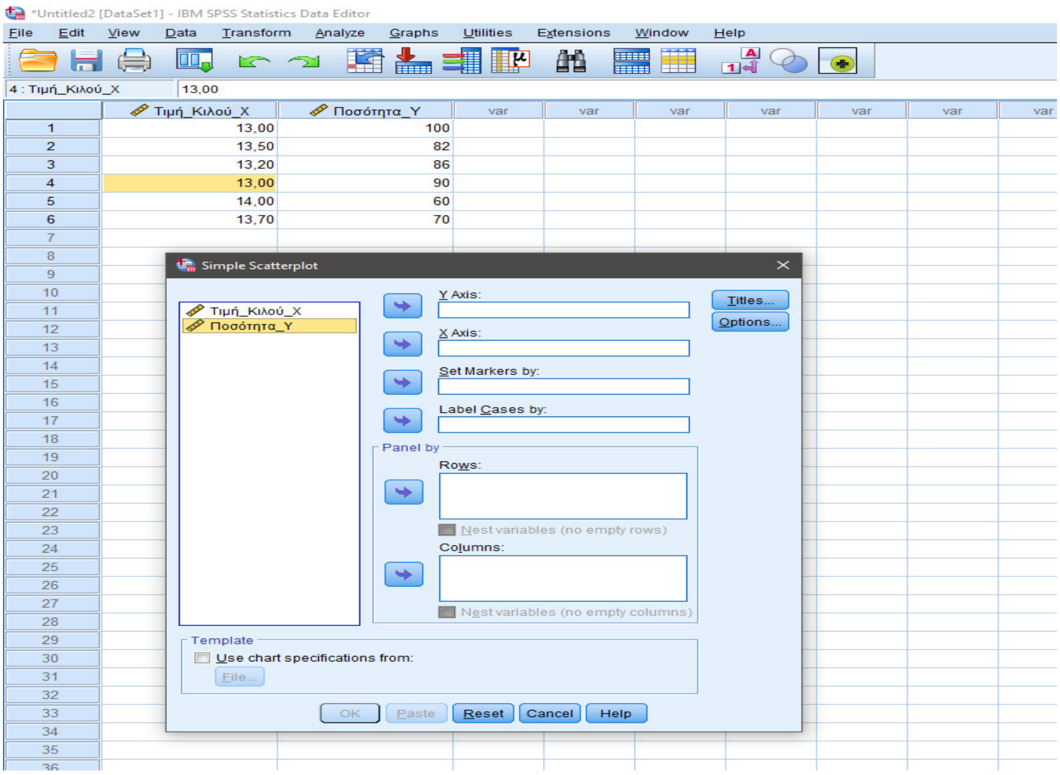
Εικόνα 3.Δημιουργία Γραφήματος Διασποράς των Δεδομένων

Στο παράθυρο που ανοίγει επιλέγουμε Simple Scatter και πατάμε Define.



Εικόνα 4. Επιλογή Απλού Γραφήματος Διασποράς των Δεδομένων

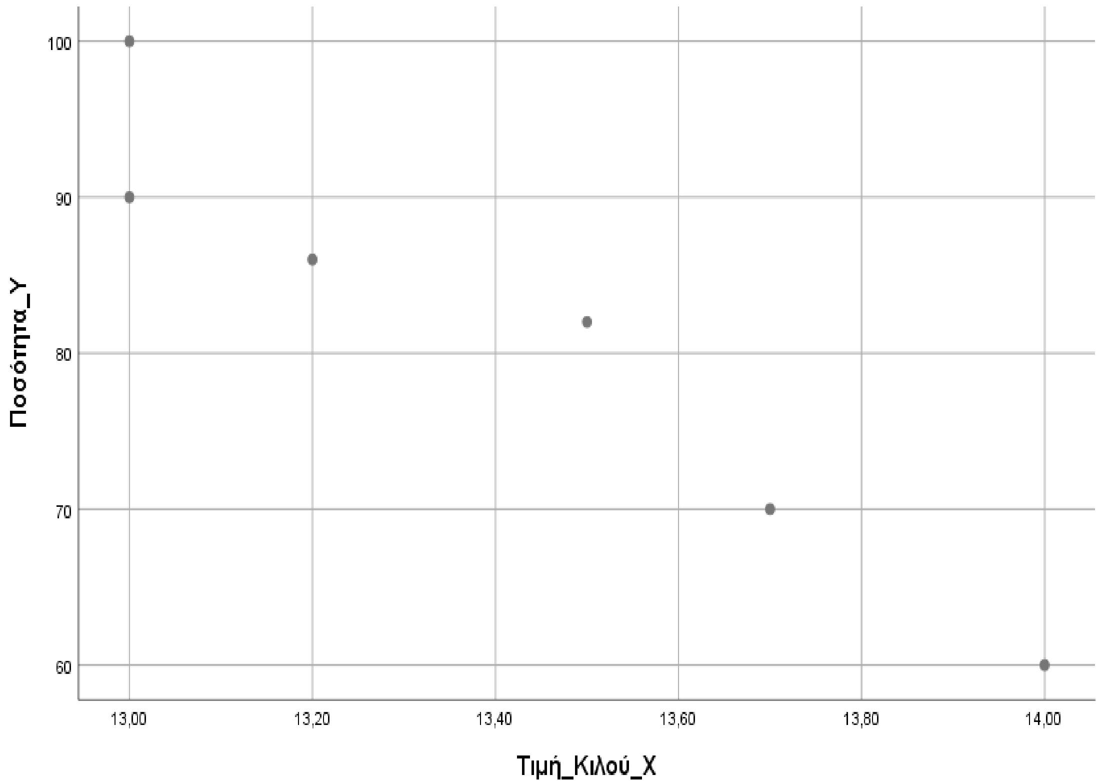
Πατώντας Define στο παράθυρο που φαίνεται στην Εικόνα 4, προκύπτει το παράθυρο που εμφανίζεται στην Εικόνα 5.



Εικόνα 5. Επιλογή Μεταβλητών από τα Δεδομένα

Στην Εικόνα 5 επιλέγουμε τη μεταβλητή εκείνη που θα παρουσιάζεται στον άξονα X και τη μεταβλητή που θα παρουσιάζεται στον άξονα Y. Έτσι, επιλέγουμε την Τιμή_Κιλού_X και την εισάγουμε στον άξονα X και την Ποσότητα_Y και την εισάγουμε στον άξονα Y.

Έτσι, πατώντας το OK στο ίδιο παράθυρο, προκύπτει το επόμενο Γράφημα.



Εικόνα 6. Γράφημα Διασποράς μεταξύ των Μεταβλητών X και Y

Όπως έχουμε ήδη αναφέρει, ο λόγος για τον οποίο δημιουργούμε το γράφημα αυτό είναι για να δείξουμε ότι δημιουργείται μια γραμμική σχέση μεταξύ των δυο μεταβλητών.

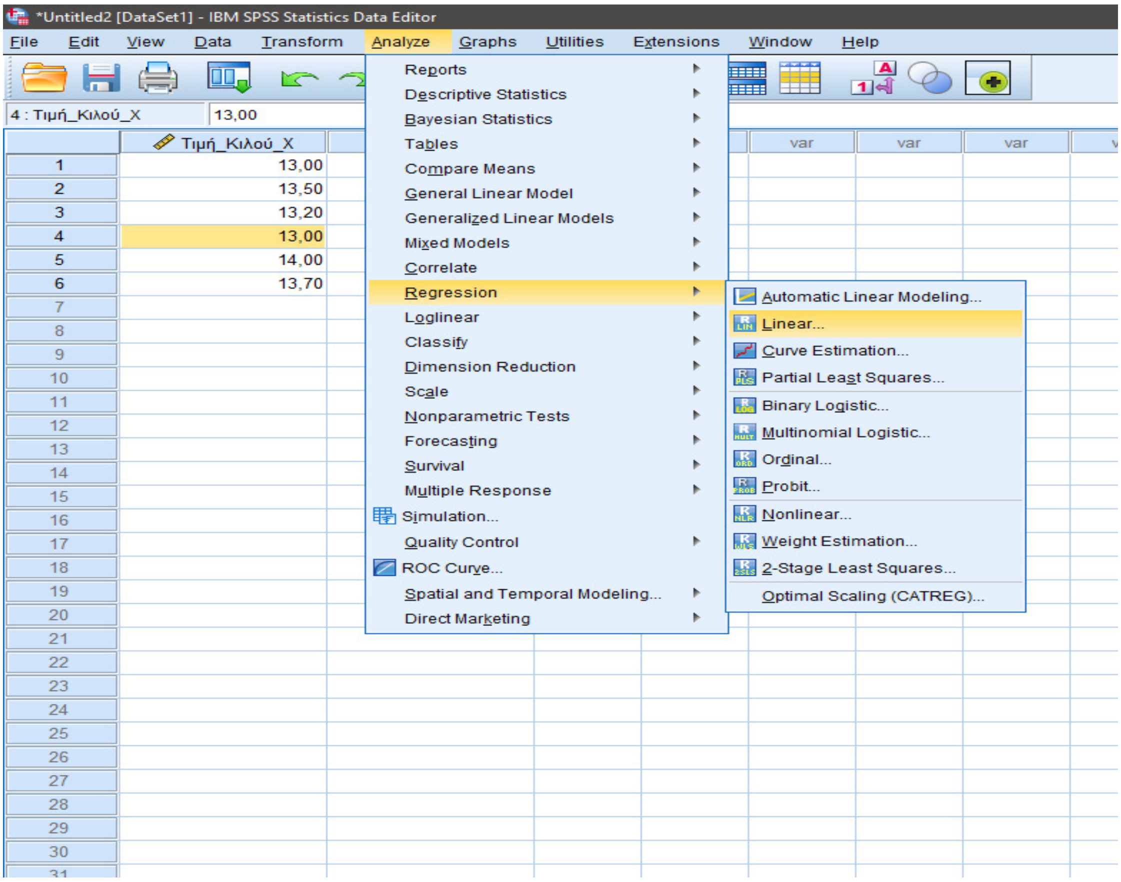
Είναι εμφανές ότι στο γράφημα της Εικόνας 6 αυτό συμβαίνει και μάλιστα υπάρχει αρνητική σχέση μεταξύ της μεταβλητής X και της μεταβλητής Y, καθώς όπως φαίνεται, όσο αυξάνεται η Τιμή Κιλού του Αγαθού Α, τόσο μειώνεται η ζητούμενη Ποσότητα, γεγονός το οποίο είναι απόλυτα λογικό να συμβαίνει.

Το παραπάνω γράφημα δεν ήταν ζητούμενο από την Άσκηση 6 την οποία προσπαθούμε να επιλύσουμε, αλλά είναι απαραίτητο να γίνει εφόσον θέλουμε να κάνουμε χρήση της μεθόδου ελαχίστων τετραγώνων και άρα θα πρέπει να ελέγξουμε εάν υπάρχει γραμμική σχέση μεταξύ των μεταβλητών X και Y που θα εφαρμόσουμε τη μεθοδολογία.

Το ερώτημα α) ζητάει να εκτιμηθούν οι συντελεστές β_0 και β_1 σε ένα υπόδειγμα με εξαρτημένη μεταβλητή τη ζητούμενη ποσότητα του αγαθού Α και με ανεξάρτητη μεταβλητή την τιμή του κιλού του αγαθού Α, το οποίο είναι το εξής:

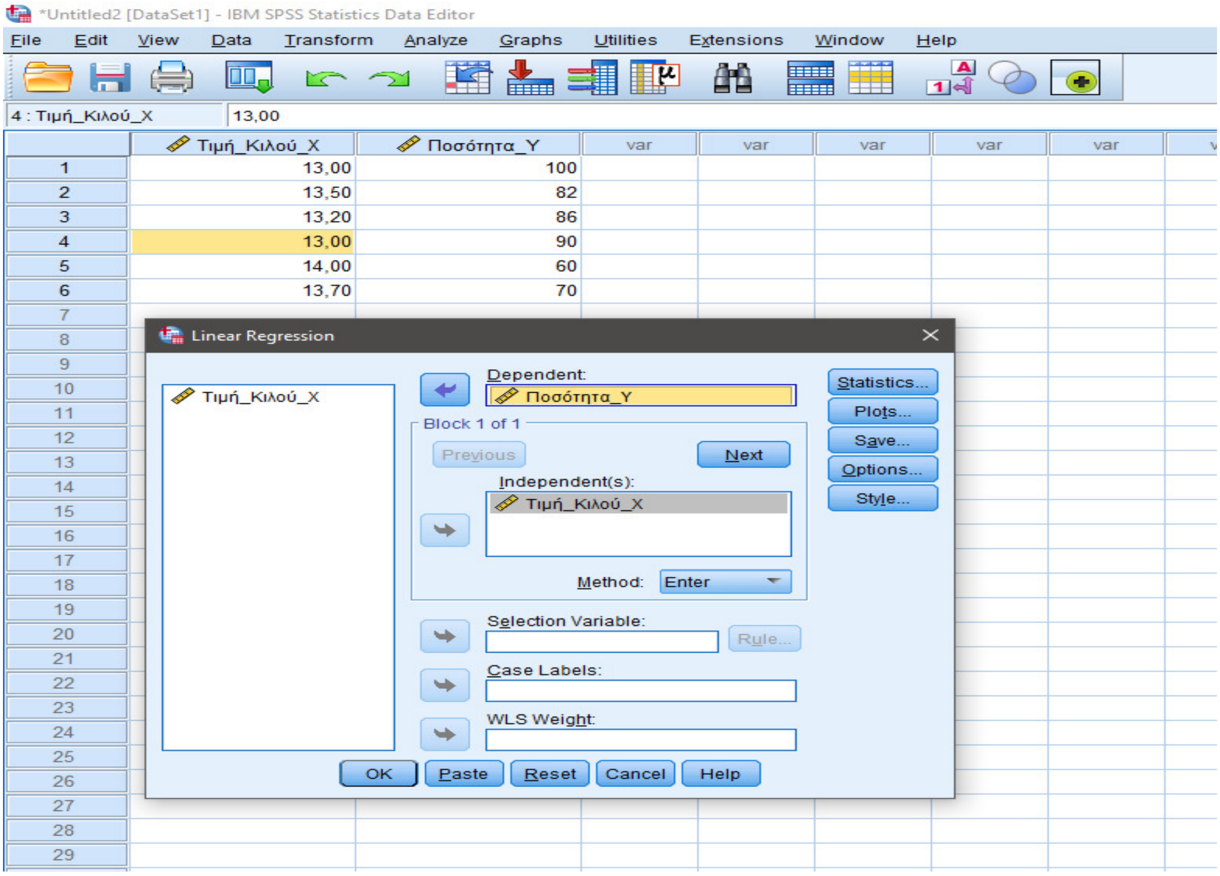
$$\text{Ποσότητα\_Y} = \beta_0 + \beta_1 \text{Τιμή\_Κιλού\_X}$$

Κάνοντας χρήση του SPSS και πιο συγκεκριμένα ακολουθώντας τη διαδικασία Analyze → Regression → Linear, όπως φαίνεται στην Εικόνα 7



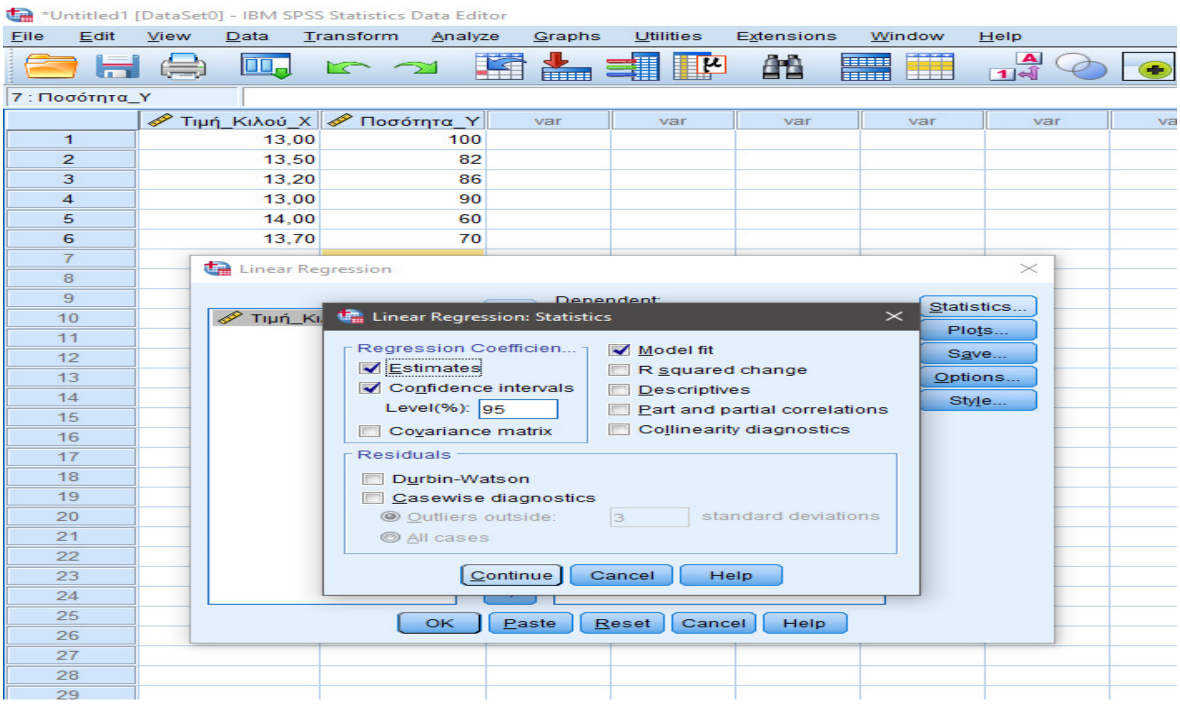
Εικόνα 7. Διαδικασία μεθόδου Γραμμικής Παλινδρόμησης στο SPSS

Στο παράθυρο που ανοίγει στη συνέχεια, το οποίο φαίνεται στην επόμενη εικόνα, μπορούμε να εισάγουμε τις μεταβλητές που θα χρησιμοποιήσουμε και μπορούμε να τις χαρακτηρίσουμε και ως εξαρτημένη και ανεξάρτητη. Στη συγκεκριμένη περίπτωση θα εισάγουμε στο πεδίο της εξαρτημένης μεταβλητής την Ποσότητα_Y και στο πεδίο της ανεξάρτητης μεταβλητής την Τιμή_Κιλού_X.



Εικόνα 8. Εισαγωγή Εξαρτημένης και Ανεξάρτητης Μεταβλητής στο υπόδειγμα Παλινδρόμησης

Στο κουμπί Statistics θα τσεκάρουμε την επιλογή Confidence intervals Level % 95, και στη συνέχεια επιλέγουμε Continue, όπως φαίνεται στην Εικόνα 9.



Εικόνα 9. Επιλογή Κατασκευής Διαστημάτων Εμπιστοσύνης για τους Συντελεστές Παλινδρόμησης

Πατώντας το OK μπορούμε να χρησιμοποιήσουμε τον επόμενο πίνακα ώστε να απαντήσουμε την ερώτηση αναφορικά με την εκτίμηση των συντελεστών της παλινδρόμησης.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95,0% Confidence Interval for B Lower Bound	95,0% Confidence Interval for B Upper Bound
1	(Constant)	538,894	63,342		8,508	,001	363,028	714,761
	Τιμή_Κιλού_X	-34,146	4,725	-,964	-7,226	,002	-47,266	-21,027

a. Dependent Variable: Ποσότητα_Y

Πίνακας 7. Πίνακας συντελεστών Παλινδρόμησης

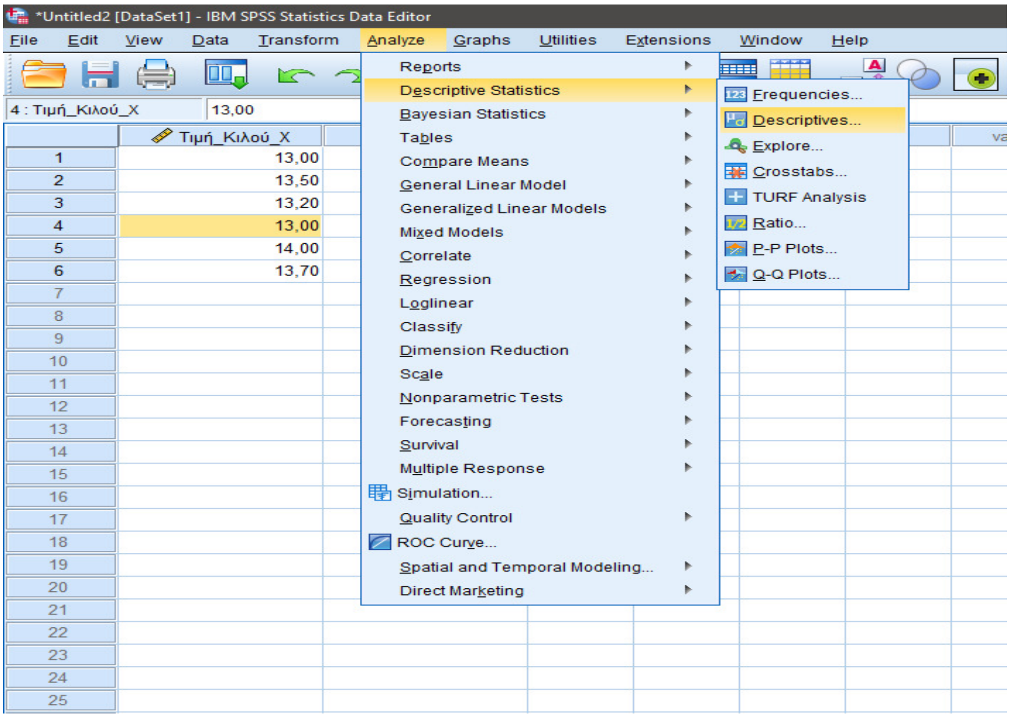
Από τον Πίνακα 7 μπορούμε να πούμε ότι ο σταθερός όρος της παλινδρόμησης β_0 είναι 538,894, ενώ ο συντελεστής της μεταβλητής X, β_1 παίρνει την τιμή -34,146.

Στο ερώτημα β) της άσκησης καλούμαστε να υπολογίσουμε την ελαστικότητα ζήτησης στο σημείο των μέσων $(\bar{X},\bar{Y})$

Θα χρησιμοποιήσουμε τον τύπο της ελαστικότητας ζήτησης ο οποίος είναι:

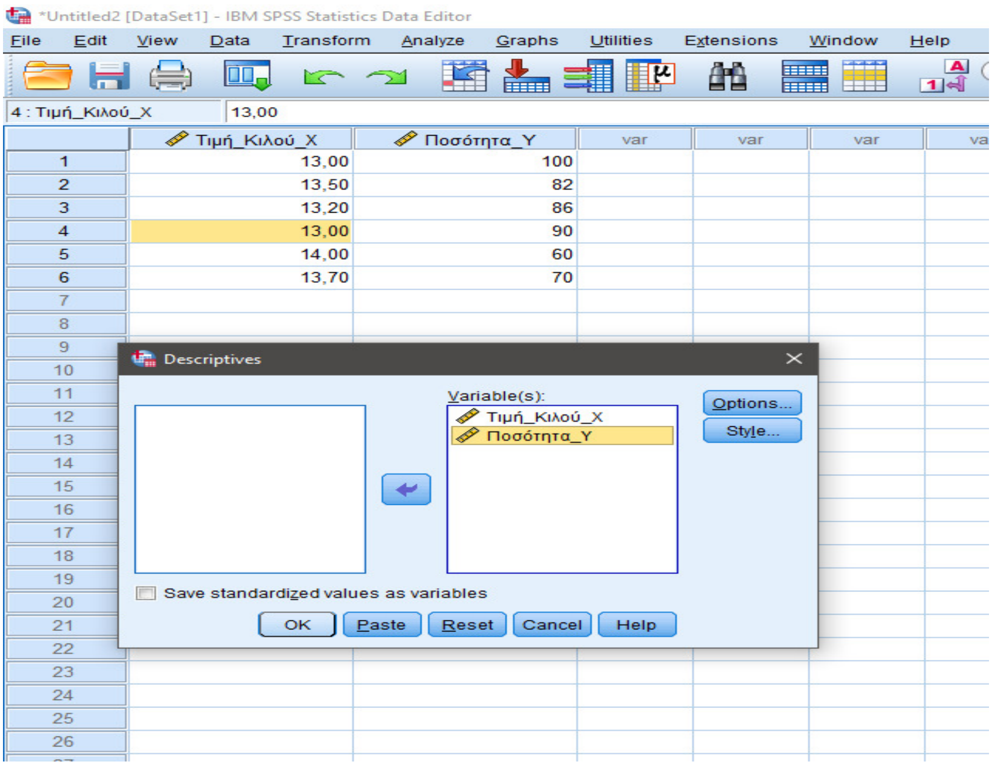
$$\varepsilon_{XY} = \beta_1 \frac{\bar{X}}{\bar{Y}}$$

Αφού χρησιμοποιήσουμε το SPSS για να υπολογίσουμε τις μέσες τιμές των μεταβλητών X και Y, κάνοντας χρήση της διαδικασίας Analyze → Descriptive Statistics → Descriptives



Εικόνα 10. Διαδικασία Υπολογισμού Περιγραφικών Στατιστικών των Μεταβλητών Χ και Υ

Αφού χρησιμοποιήσουμε την παραπάνω διαδικασία, ανοίγει το επόμενο παράθυρο



Εικόνα 11. Επιλογή Μεταβλητών Υπολογισμού Περιγραφικών Στατιστικών

Πατώντας το ΟΚ στο παράθυρο επιλογής μεταβλητών προκύπτει ο επόμενος πίνακας

Περιγραφικά Στατιστικά					
	N	Ελάχιστο	Μέγιστο	Μέση Τιμή	Τυπική Απόκλιση
Τιμή_Κιλού_X	6	13,00	14,00	13,4000	,40497
Ποσότητα_Y	6	60	100	81,33	14,348
Έγκυρες Παρατηρήσεις	6				

Πίνακας 8. Περιγραφικά Στατιστικά των Μεταβλητών της Τιμής_Κιλού_X και της Ποσότητας_Y

Οπότε παρατηρούμε ότι η μέση τιμή της ζητούμενης ποσότητας είναι 81,33, ενώ για τη μεταβλητή της τιμής του κιλού είναι 13,4.

Μπορούμε να υπολογίσουμε την ελαστικότητα σύμφωνα με τον τύπο

$$\varepsilon_{XY} = \beta_1\frac{\overline{X}}{\overline{Y}} = -34{,}146\frac{13{,}4}{81{,}33} = -5{,}626$$

Η ελαστικότητα ζήτησης δείχνει την ποσοστιαία μεταβολή της ποσότητας αναλογικά με την ποσοστιαία μεταβολή της τιμής. Εδώ μπορούμε να παρατηρήσουμε ότι η ζήτηση είναι ελαστική, καθώς η ελαστικότητα ζήτησης είναι μεγαλύτερη από τη μονάδα κατά απόλυτη τιμή. Αυτό σημαίνει ότι η ποσοστιαία μεταβολή της ποσότητας είναι μεγαλύτερη από την ποσοστιαία μεταβολή της τιμής.

Το πρόσημο της ελαστικότητας ζήτησης είναι αρνητικό. Αυτό ερμηνεύεται από το γεγονός ότι όσο μεγαλώνει η τιμή ανά κιλό του αγαθού Α, τόσο θα μειώνεται η ζητούμενη ποσότητα για το αγαθό Α.

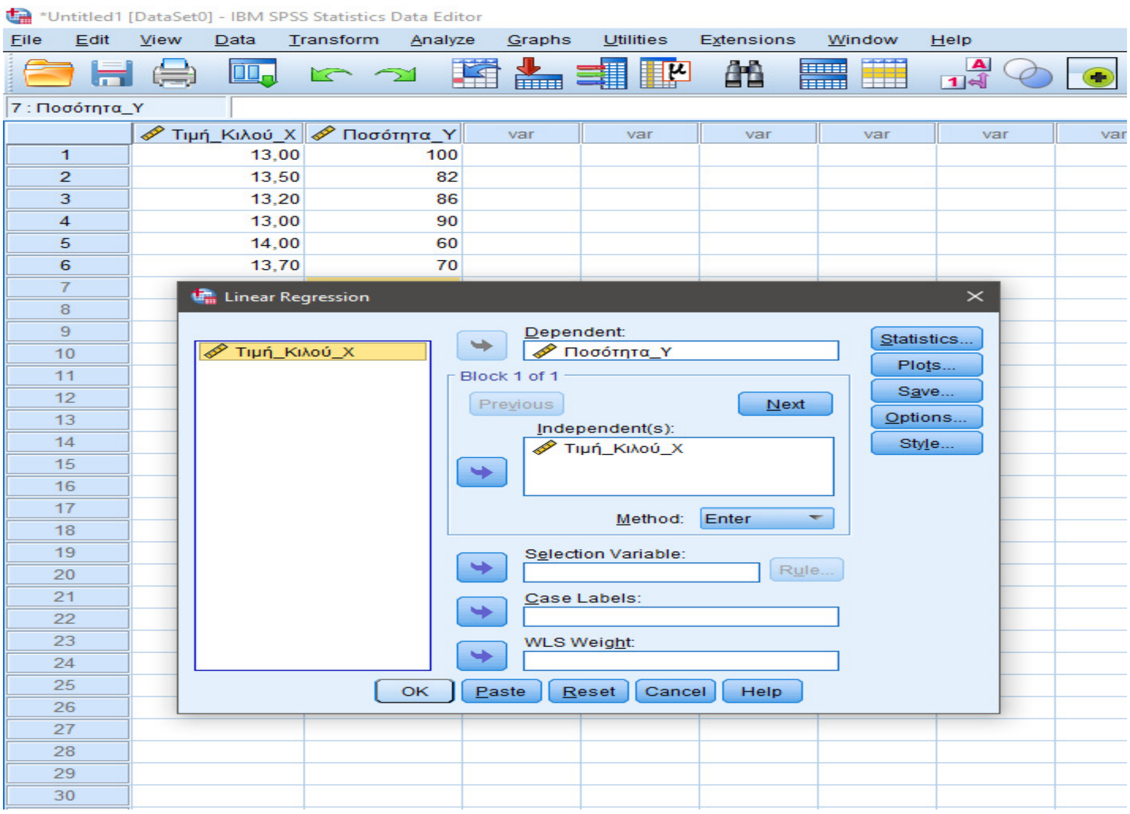
Με λίγα λόγια, σύμφωνα με την τιμή που υπολογίσαμε για την ελαστικότητα ζήτησης στο σημείο των μέσων, δείχνει ότι η ποσοστιαία αύξηση της τιμής ανά κιλό του αγαθού Α θα οδηγήσει σε μια ποσοστιαία μεγαλύτερη μείωση της ζητούμενης ποσότητας για το αγαθό Α.

Στο ερώτημα γ) θα χρειαστεί να αναφερθούμε ποιοτικά στα αποτελέσματα από οικονομική άποψης.

Μπορούμε να πούμε ότι ακόμα και εάν η τιμή ανά κιλό του αγαθού Α ήταν μηδέν, δηλαδή το αγαθό Α προσφερόταν στην αγορά δωρεάν, η ζητούμενη ποσότητα για το αγαθό Α θα ήταν β_0 = 538,894 κιλά.

Στην περίπτωση που η τιμή ανά κιλό του αγαθού Α αυξηθεί κατά μια μονάδα τότε η ζητούμενη ποσότητα για το αγαθό Α θα μειωθεί κατά β_1 = -34,146, το οποίο είναι λογικό, καθώς γνωρίζουμε από την οικονομική θεωρία ότι όσο αυξάνεται η τιμή ενός αγαθού τόσο μειώνεται η ζητούμενη ποσότητα του ίδιου αγαθού. Άρα τα αποτελέσματα που έχουμε υπολογίσει επαληθεύουν την οικονομική θεωρία.

Για να απαντήσουμε το ερώτημα δ) και να αξιολογήσουμε τα αποτελέσματα βάσει στατιστικών κριτηρίων, θα πρέπει να ακολουθήσουμε τη διαδικασία Analyze → Regression → Linear και στο παράθυρο που θα ανοίξει θα εισάγουμε τη μεταβλητή Ποσότητα_Υ στο πεδίο της εξαρτημένης μεταβλητής και τη μεταβλητή Τιμή_Κιλού_Χ στο πεδίο της ανεξάρτητης μεταβλητής και πατάμε ΟΚ, όπως φαίνεται στην επόμενη εικόνα.



Εικόνα 12. Εισαγωγή ανεξάρτητης και εξαρτημένης μεταβλητής στα αντίστοιχα πεδία

Μετά τα παραπάνω προκύπτουν οι πίνακες που θα μας φανούν χρήσιμοι για την ανάλυση, αλλά αυτό που μας ενδιαφέρει κυρίως είναι ο επόμενος πίνακας.

Σύνοψη Υποδείγματος

Υπόδειγμα	Συντελεστής Συσχέτισης	Συντελεστής Προσδιορισμού	Προσαρμοσμένος Συντελεστής Προσδιορισμού	Τυπικό Σφάλμα της Εκτίμησης
1	,964	,929	,911	4,279

Πίνακας 9. Υπολογισμός Συντελεστή Προσδιορισμού του Υποδείγματος Παλινδρόμησης

Όπως προκύπτει από τον παραπάνω πίνακα, η τιμή του συντελεστή προσδιορισμού είναι 0,964 και δείχνει τη μεταβλητότητα της ζητούμενης ποσότητας του αγαθού Α που ερμηνεύεται από τη μεταβλητότητα της τιμής ανά κιλό του αγαθού Α. Άρα το 96,4 % της ζητούμενης ποσότητας ερμηνεύεται από την τιμή/ κιλό και άρα μπορούμε να πούμε ότι το υπόδειγμα είναι ικανοποιητικό σε πολύ μεγάλο βαθμό.

Οι ζητούμενες τιμές των τυπικών σφαλμάτων δίνονται από τον πίνακα 7, όπου η τιμή του τυπικού σφάλματος του σταθερού όρου είναι 63,342, ενώ η τιμή του τυπικού σφάλματος του συντελεστή της τιμής ανά κιλό του αγαθού Α είναι 4,725.

Μπορούμε να εξετάσουμε διαδοχικά εάν οι συντελεστές είναι μηδενικοί.

Εξετάζουμε πρώτα για το σταθερό όρο.

Ορίζουμε τη μηδενική υπόθεση και την εναλλακτική υπόθεση, οι οποίες είναι:

H_0: $\beta_0 = 0$

H_1: $\beta_0 \neq 0$

Από τον πίνακα 7 μπορούμε να απορρίψουμε τη μηδενική υπόθεση ότι ο σταθερός όρος είναι μηδέν, καθώς η p-value = 0,001του ελέγχου είναι μικρότερη από το επίπεδο σημαντικότητας α=5%.

Στη συνέχεια εξετάζουμε για το συντελεστή της ανεξάρτητης μεταβλητής της τιμής ανά κιλό.

Ορίζουμε τη μηδενική υπόθεση και την εναλλακτική υπόθεση, οι οποίες είναι:

H_0: $\beta_1 = 0$

H_1: $\beta_1 \neq 0$

Από τον πίνακα 7 μπορούμε να απορρίψουμε τη μηδενική υπόθεση ότι ο συντελεστής β_1 είναι μηδέν, καθώς η p-value = 0,002 του ελέγχου είναι μικρότερη από το επίπεδο σημαντικότητας α=5%.

Συνολικά, μπορούμε να πούμε ότι τα αποτελέσματα που έχουν προκύψει είναι ικανοποιητικά και το υπόδειγμα που εξετάζουμε είναι ένα αρκετά καλό υπόδειγμα.

Με την ίδια διαδικασία του SPSS, (Analyze → Regression → Linear), μπορούμε να απαντήσουμε και το ερώτημα ε), το οποίο ζητάει την κατασκευή του πίνακα ανάλυσης διακύμανσης.

Ο πίνακας ανάλυσης διακύμανσης ANOVAείναι ο εξής:

ΠΙΝΑΚΑΣ ΑΝΑΛΥΣΗΣ ΔΙΑΚΥΜΑΝΣΗΣ

	Υπόδειγμα	Άθροισμα Τετραγώνων	Βαθμοί Ελευθερίας	Μέσα Τετράγωνα	F-Στατιστική	P-τιμή Ελέγχου
1	Παλινδρόμηση	956,098	1	956,098	52,220	,002[b]
	Κατάλοιπα	73,236	4	18,309		
	Ολικό	1029,333	5			

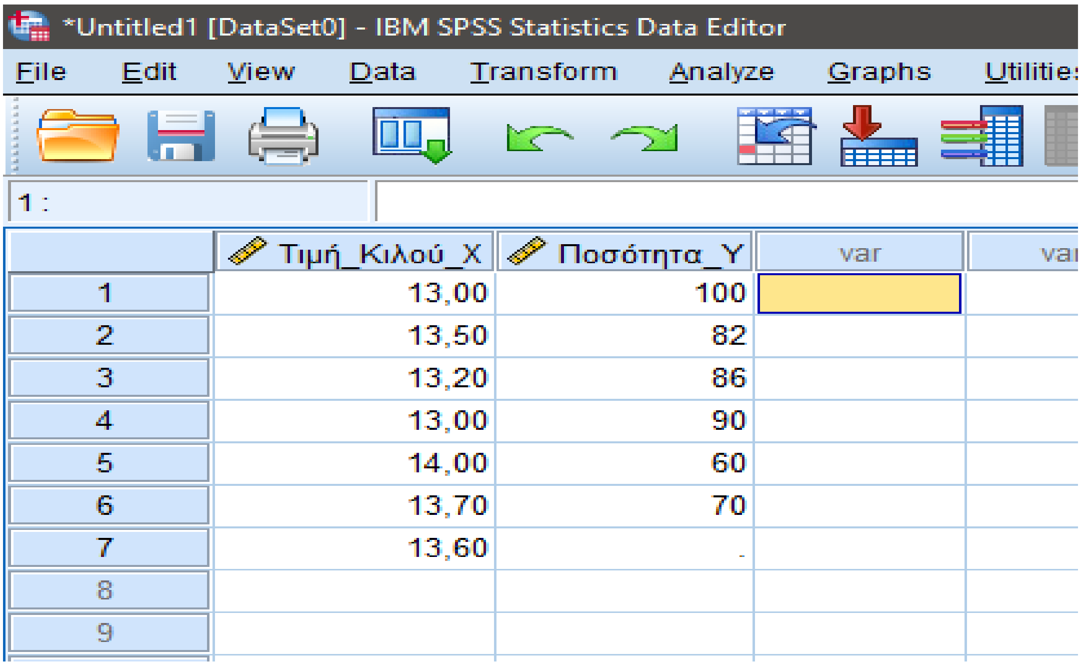
Πίνακας 10. Πίνακας Ανάλυσης Διακύμανσης

Στο ερώτημα στ) ζητείται να κατασκευάσουμε ένα 95% διάστημα εμπιστοσύνης για το συντελεστή β_1.

Επιστρέφοντας στον Πίνακα 7 μπορούμε να διαπιστώσουμε ότι το διάστημα εμπιστοσύνης για το συντελεστή β_1 είναι το (-47,266, -21,027).

Για την απάντηση του ζ) ερωτήματος θα πρέπει να δουλέψουμε ως εξής:

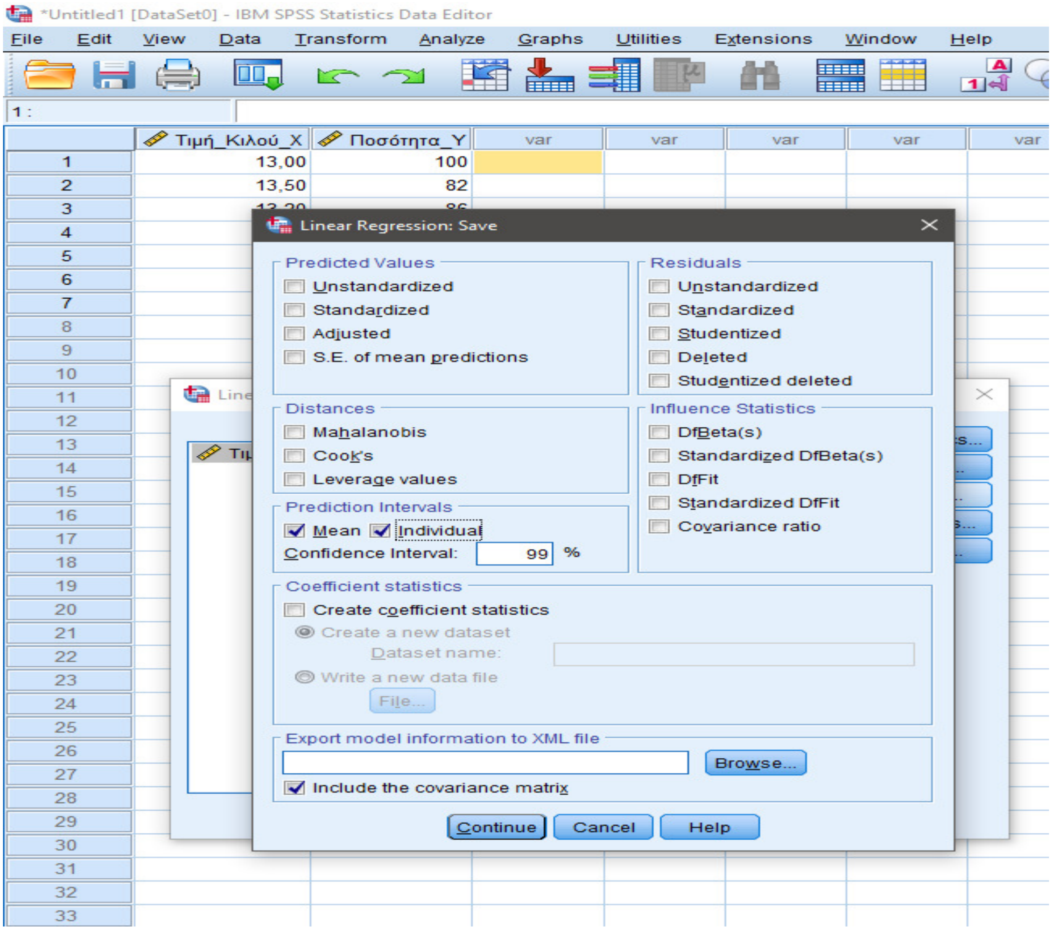
Θα πρέπει να εισάγουμε τη ζητούμενη τιμή 13,6 ως τιμή για τη μεταβλητή της τιμής κιλού X, έτσι ώστε να φαίνεται όπως στην επόμενη εικόνα.



Εικόνα 13. Εισαγωγή της Ζητούμενης Τιμής 13,6 στις Τιμές της Μεταβλητής X

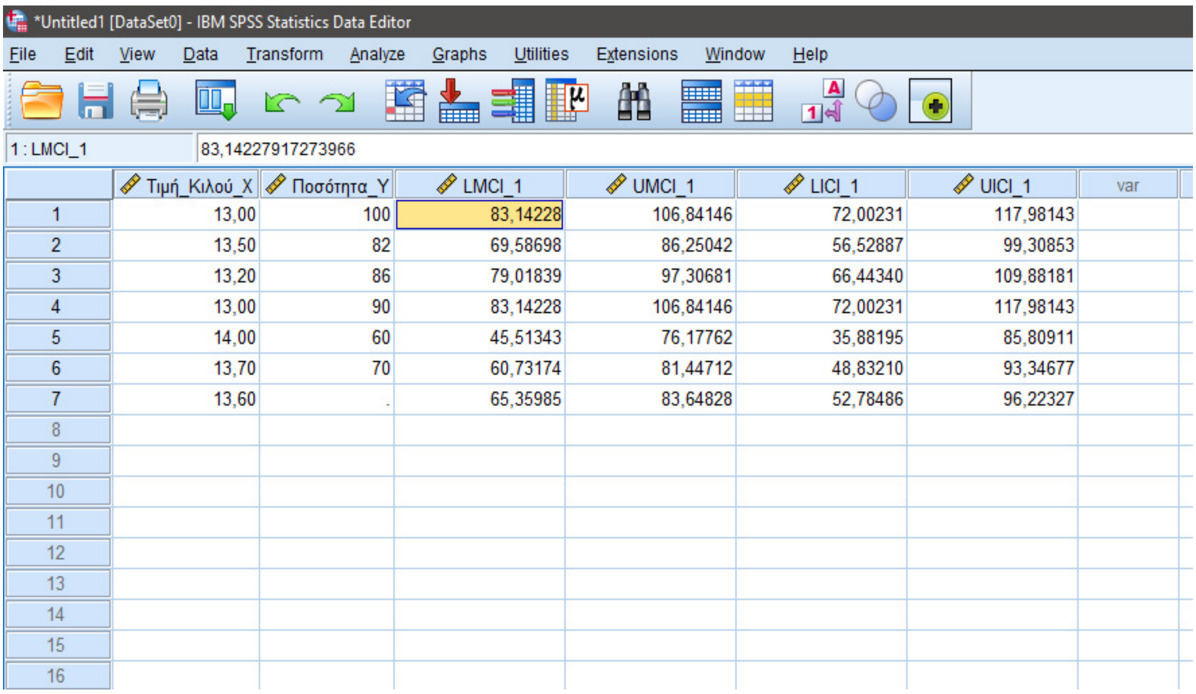
Στη συνέχεια ακολουθούμε την εξής διαδικασία Analyze → Regression → Linear (Όπως φαίνεται στην Εικόνα 7) και στο παράθυρο που ανοίγει στο menu Save (Εικόνα 11) στο τμήμα για τα Prediction Intervals τσεκάρουμε και τις δυο επιλογές Mean και Individuals και αλλάζουμε την τιμή του Confidence Interval σε 99 %. Στη συνέχεια πατάμε Continue και τέλος το OK.

Τα παραπάνω παρουσιάζονται στην επόμενη Εικόνα.



Εικόνα 14. Επιλογή Διαστημάτων Εμπιστοσύνης για τη Μέση Τιμή και για την Προβλεπόμενη Τιμή

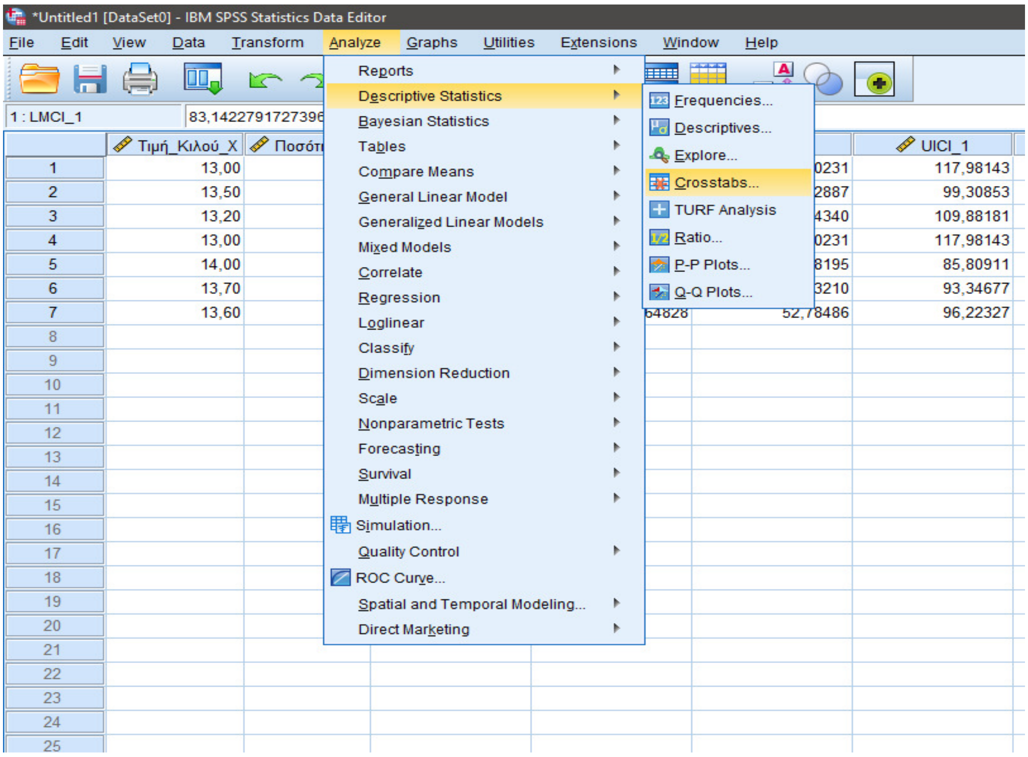
Στην Εικόνα 15 παρουσιάζονται οι νέοι υπολογισμοί.



Εικόνα 15. Υπολογισμοί Διαστημάτων Εμπιστοσύνης Μέσης και Προβλεπόμενης Τιμής

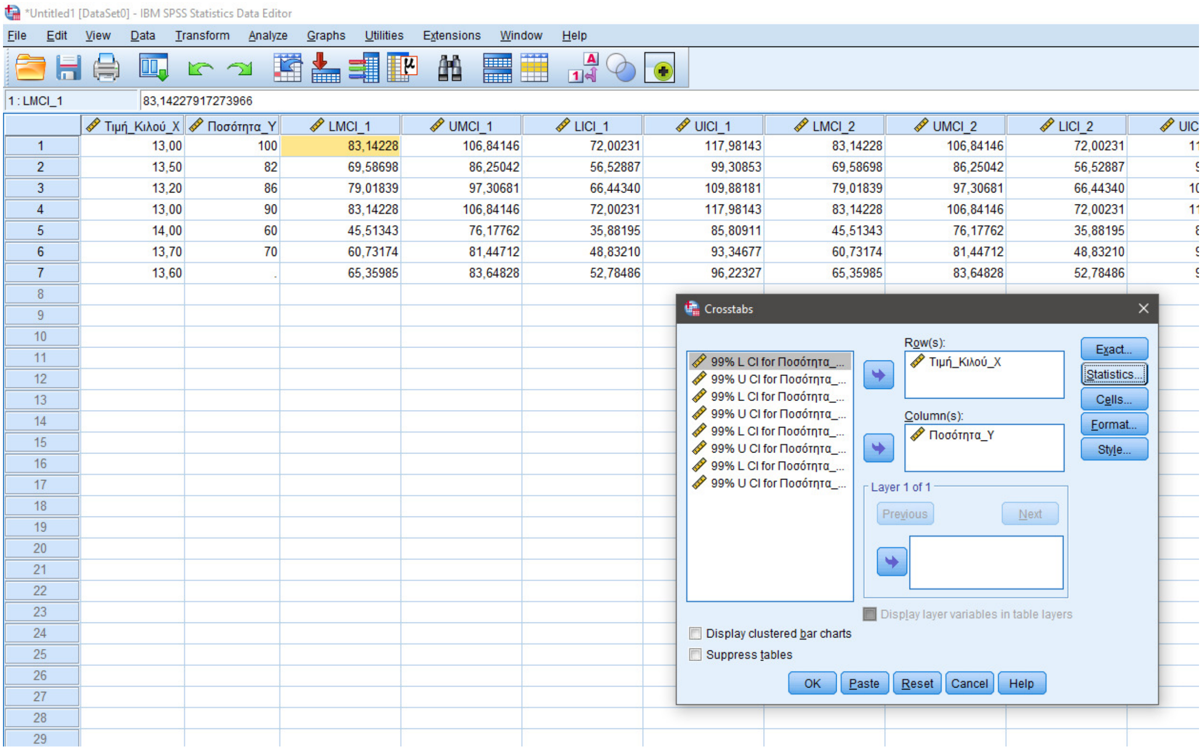
Άρα το 99% διάστημα εμπιστοσύνης για τη μέση ζητούμενη ποσότητα όταν η τιμή είναι 13,6 είναι το (65,35985, 83,64828), ενώ το 99% διάστημα εμπιστοσύνης για τη συγκεκριμένη ζητούμενη ποσότητα όταν η τιμή είναι 13,6 είναι το (52,78486, 96,22327).

Για να κάνουμε τον έλεγχο στατιστικής σημαντικότητας του συντελεστή συσχέτισης όπως ζητάει το ερώτημα η) ακολουθούμε τη διαδικασία Analyze → Descriptive Statistics → Crosstabs, όπως φαίνεται στην επόμενη εικόνα



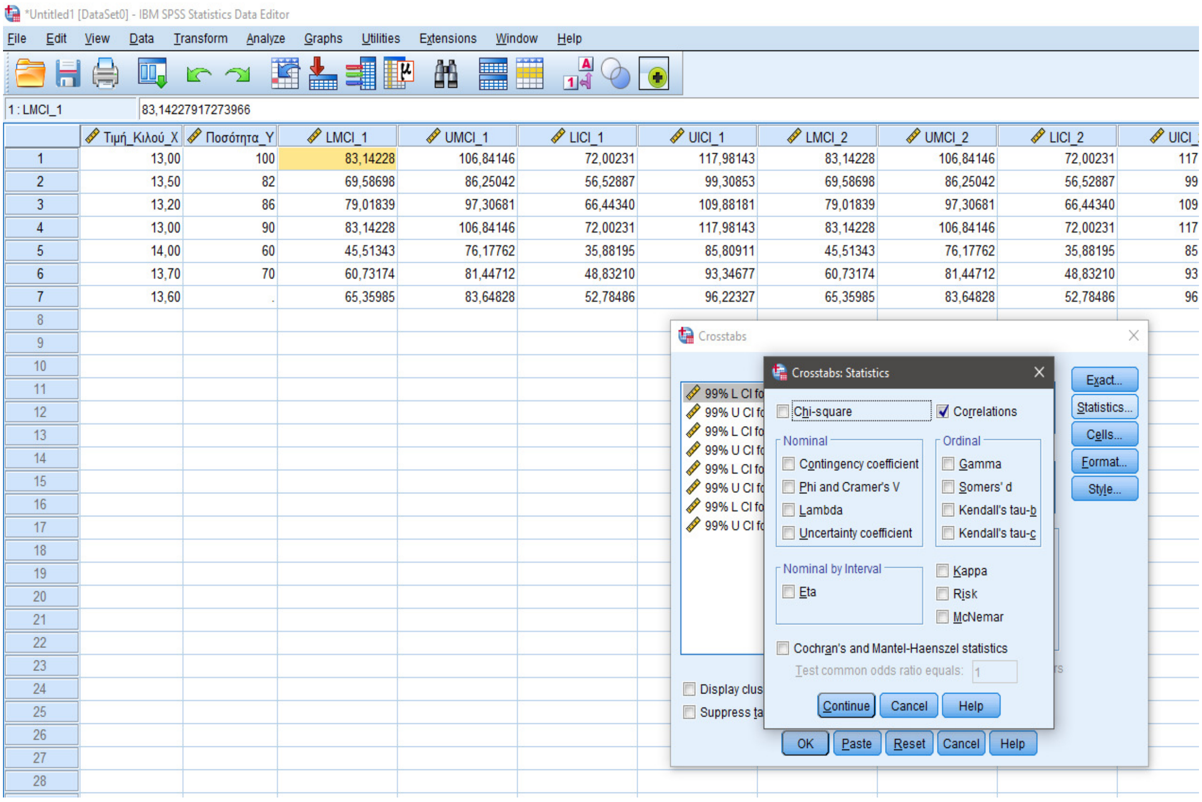
Εικόνα 16. Διαδικασία Ελέγχου Στατιστικής Σημαντικότητας Συντελεστή Συσχέτισης

Στο παράθυρο που ανοίγει επιλέγουμε τη μεταβλητή Τιμή_Κιλού_X στο πεδίο Row(s) και τη μεταβλητή Ποσότητα_Y στο πεδίο Column(s) και στη συνέχεια επιλέγουμε το menu Statistics, όπως φαίνεται στην εικόνα



Εικόνα 17. Επιλογή Μεταβλητών για τον Έλεγχο Στατιστικής Σημαντικότητας του Συντελεστή Συσχέτισης

Στο menu του Statistics τσεκάρουμε την επιλογή Correlations και πατάμε Continue, όπως φαίνεται στην επόμενη εικόνα.



Εικόνα 18. Menu Statistics και επιλογή Correlations

Πατώντας OK προκύπτει ο επόμενος πίνακας

ΠΙΝΑΚΑΣ ΣΥΜΜΕΤΡΙΚΩΝ ΜΕΤΡΗΣΕΩΝ					
		Τιμή	Ασυμπτωτικό Τυπικό Σφάλμα	Προσεγγιστική Τιμή-Τ	Προσεγγιστική Σημαντικότητα
Ανά Διάστημα	Συντελεστής Συσχέτισης Pearson's R	-,964	,021	-7,226	,002[c]
Έγκυρες Παρατηρήσεις		6			

Πίνακας 11. Υπολογισμός Συντελεστή Συσχέτισης και Έλεγχος Στατιστικής Σημαντικότητας του Συντελεστή Συσχέτισης

Όπως φαίνεται στον Πίνακα 11, η τιμή του συντελεστή συσχέτισης (Pearson's R) είναι -0,964 , ενώ μπορούμε να απορρίψουμε τη μηδενική υπόθεση ότι η τιμή του συντελεστή είναι μηδέν, καθώς η p-value του ελέγχου είναι 0,002, τιμή που είναι μικρότερη από το επίπεδο στατιστικής σημαντικότητας α=5%.

Άρα ο συντελεστής συσχέτισης είναι στατιστικά σημαντικός.

Η ερώτηση θ) ζητάει τις υποθέσεις που θα πρέπει να ισχύουν ώστε να απαντηθούν όλες οι ερωτήσεις. Αυτές οι υποθέσεις που πρέπει να ισχύουν έχουν ήδη παρατεθεί στην ενότητα 3.5. και επιγραμματικά είναι οι εξής (Χρήστου, 2008):

1. Η αναμενόμενη τιμή των καταλοίπων είναι μηδέν. $E(u_i) = 0$
2. Η διακύμανση των καταλοίπων παραμένει σταθερή και ίση με σ2. (υπόθεση ομοσκεδαστικότητας). $Var(u_i) = \sigma^2$.
3. Η συνδιακύμανση των καταλοίπων, για κατάλοιπα διαφορετικών χρονικών στιγμών ισούται με μηδέν. $Cov(u_i,u_j) = 0, \forall\, i \neq j$
4. Η εκτίμηση των παραμέτρων του απλού γραμμικού υποδείγματος, β_0καιβ_1, παραμένει σταθερή σε όλη τη διάρκεια της μελέτης.
5. Ο διαταρακτικός όρος (κατάλοιπα) u_i , είναι τυχαία μεταβλητή, η οποία ακολουθεί την κανονική κατανομή, με μέση τιμή μηδέν και διακύμανση σ^2. $u_i \sim N(0, \sigma^2)$
6. Οι τιμές που παίρνει η επεξηγηματική μεταβλητή X παραμένουν σταθερές και μεταξύ τους δεν είναι όλες ίσες. Άρα, η επεξηγηματική μεταβλητή δεν είναι στοχαστική, αλλά ντετερμινιστική.

Κεφάλαιο 4. Πολλαπλά γραμμικά μοντέλα παλινδρόμησης

4.1 Βασικές έννοιες

Στο προηγούμενο κεφάλαιο αναλύσαμε τα υποδείγματα απλής παλινδρόμησης, υποδείγματα τα οποία περιέχουν μια εξαρτημένη μεταβλητή και μια ανεξάρτητη μεταβλητή. Όμως, τα περισσότερα οικονομικά υποδείγματα δεν εκτιμούν τη συσχέτιση μιας εξαρτημένης με μια ανεξάρτητη μεταβλητή, αλλά τη σχέση μιας μεταβλητής με ένα σύνολο μεταβλητών. Αυτά είναι τα υποδείγματα πολλαπλής παλινδρόμησης.

Στο κεφάλαιο αυτό θα αναλύσουμε εκείνα τα γραμμικά υποδείγματα τα οποία περιέχουν μια εξαρτημένη μεταβλητή όπως και τα απλά υποδείγματα, όμως οι ανεξάρτητες μεταβλητές μπορούν να είναι περισσότερες από μια, και θα τις συμβολίσουμε με k, k> 1. Στην περίπτωση όπου k = 1, τότε επιστρέφουμε στην περίπτωση της απλής παλινδρόμησης, με μία μόνο ανεξάρτητη μεταβλητή X.

Τα υποδείγματα πολλαπλής γραμμικής παλινδρόμησης παίρνουν την εξής μορφή:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + u_i$$

Όπου με X_{ik} συμβολίζεται η i παρατήρηση της k ανεξάρτητης μεταβλητής και με β_i συμβολίζονται οι k συντελεστές πολλαπλής παλινδρόμησης, (Χρήστου, 2008).

Ακριβώς όπως και στην απλή παλινδρόμηση, με u_i συμβολίζονται τα κατάλοιπα του υποδείγματος.

4.2 Εξίσωση ελαχίστων τετραγώνων

4.2.1 Εισαγωγή στην Πολλαπλή Γραμμική Παλινδρόμηση με Χρήση Πινάκων

Λόγω του μεγάλου αριθμού επεξηγηματικών μεταβλητών και συντελεστών, στην πολλαπλή παλινδρόμηση κάνουμε χρήση πινάκων για να εκφράσουμε όλα αυτά τα μεγέθη και την ευθεία παλινδρόμησης, όπως παρουσιάζουμε παρακάτω.

Έστω το πολλαπλό γραμμικό μοντέλο παλινδρόμησης, το οποίο ορίσαμε στην προηγούμενη παράγραφο, ως εξής:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + u_i$$

Όπως έχουμε ήδη εξηγήσει, ο αριθμός των επεξηγηματικών μεταβλητών είναι k, όπως και ο αριθμός των συντελεστών της παλινδρόμησης, ενώ το δείγμα παίρνει τιμές i = 1,2,….,n.

Για κάθε τιμή του i, η εξίσωση παλινδρόμησης γίνεται:

$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_k X_{1k} + u_1$

$Y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_k X_{2k} + u_2$

.

.

.

$Y_{n0} = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_k X_{nk} + u_n$

Αν χρησιμοποιήσουμε τον συμβολισμό των πινάκων, μπορούμε να γράψουμε:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \dots X_{1k} \\ 1 & X_{21} & \dots X_{2k} \\ \vdots & \vdots & \vdots \\ 1 & X_{n1} & \dots X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

Όπου, ο πίνακας $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$, είναι ο πίνακας διάστασης nx1, δηλαδή ο πίνακας- στήλη με n γραμμές και 1 στήλη, που σε κάθε γραμμή περιέχεται η τιμή της εξαρτημένης μεταβλητής Y_i.

Ο πίνακας $X = \begin{bmatrix} 1 & X_{11} & \dots X_{1k} \\ 1 & X_{21} & \dots X_{2k} \\ \vdots & \vdots & \vdots \\ 1 & X_{n1} & \dots X_{nk} \end{bmatrix}$, είναι ο πίνακας διάστασης nx(k+1), με n γραμμές και k+1 στήλες, ο οποίος περιέχει τις τιμές των ερμηνευτικών μεταβλητών X_i.

Παρατηρούμε ότι τα στοιχεία της πρώτης στήλης είναι όλα ίσα με τη μονάδα. Αυτό συμβαίνει καθώς το υπόδειγμα παλινδρόμησης περιέχει και έναν σταθερό συντελεστή, ο οποίος δεν είναι συντελεστής κάποιας από τις επεξηγηματικές μεταβλητές, αλλά σταθερός όρος.

Ο πίνακας $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$, είναι ο πίνακας διάστασης (k+1)x1, δηλαδή ο πίνακας- στήλη με k+1γραμμές και 1 στήλη, που σε κάθε γραμμή περιέχεται η τιμή του κάθε συντελεστή.

Ο πίνακας $u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$, είναι ο πίνακας διάστασης nx1, δηλαδή ο πίνακας- στήλη με n γραμμές και 1 στήλη, που σε κάθε γραμμή περιέχεται η τιμή του διαταρακτικού όρου.

4.2.2 Ευθεία Παλινδρόμησης και Εκτίμηση Συντελεστών με Χρήση Πινάκων

Συνοπτικά, μπορούμε να γράψουμε την εξίσωση παλινδρόμησης ως εξής:

$$Y = X\beta + u$$

Η παραπάνω έκφραση αφορά το πολλαπλό υπόδειγμα γραμμικής παλινδρόμησης που έχουμε ήδη περιγράψει, με χρήση πινάκων.

Για να εκτιμήσουμε τους συντελεστές παλινδρόμησης β_i, χρησιμοποιούμε τον επόμενο τύπο:

$\beta_i = (X'X)^{-1}X'Y$

Ο τύπος αυτός δίνει την εκτίμηση του κάθε συντελεστή παλινδρόμησης β_i, οι οποίοι σύμφωνα με το θεώρημα Gauss–Markovείναι άριστοι, γραμμικοί και αμερόληπτοι.

Η εκτίμηση του κάθε συντελεστή επεξηγηματικής μεταβλητής, δείχνει τη μεταβολή της εξαρτημένης μεταβλητής Y, όταν η αντίστοιχη επεξηγηματική μεταβλητή αυξηθεί κατά μια μονάδα και οι υπόλοιπες επεξηγηματικές μεταβλητές παραμείνουν αμετάβλητές.

Η τιμή του σταθερού όρου δείχνει την τιμή της εξαρτημένης μεταβλητής, όταν όλες οι επεξηγηματικές μεταβλητές πάρουν την τιμή μηδέν.

4.3 Πολλαπλό Γραμμικό Υπόδειγμα

4.3.1 Παράδειγμα εκτίμησης συντελεστών εξίσωσης πολλαπλού υποδείγματος

Έστω ότι θέλουμε να εκτιμήσουμε τη σχέση που έχει η εξαρτημένη μεταβλητή των καταναλωτικών δαπανών (Υ) 20 νοικοκυριών, όταν θεωρούμε ότι σύμφωνα με την οικονομική θεωρία εξαρτάται μόνο από το διαθέσιμο εισόδημα (X_1) και τις δαπάνες στέγασης (X_2) του κάθε νοικοκυριού.

Υποθέτουμε, δηλαδή, ότι το υπόδειγμα που θέλουμε να εκτιμήσουμε είναι το εξής:

$$\Upsilon = \beta_o + \beta_1 X_1 + \beta_2 X_2 + u_i$$

Τα δεδομένα του παραδείγματος, που αφορούν τις ετήσιες παρατηρήσεις Καταναλωτικών Δαπανών (Υ), Διαθέσιμου Εισοδήματος (X_1) και Δαπανών Στέγασης (X_2), για κάθε ένα από τα 20 νοικοκυριά είναι τα εξής:

Καταναλωτικές Δαπάνες (Υ)	Διαθέσιμο Εισόδημα (X_1)	Δαπάνες Στέγασης (X_2)
5,00 €	110,00 €	30,00 €
11,00 €	120,00 €	35,00 €
18,00 €	143,00 €	40,00 €
24,00 €	150,00 €	42,00 €
29,00 €	152,00 €	45,00 €
35,00 €	162,00 €	52,00 €
37,00 €	180,00 €	57,00 €
42,00 €	182,00 €	61,00 €
43,00 €	180,00 €	61,00 €
48,00 €	175,00 €	63,00 €
49,00 €	195,00 €	65,00 €
51,00 €	196,00 €	70,00 €
53,00 €	183,00 €	75,00 €
40,00 €	190,00 €	72,00 €
34,00 €	150,00 €	65,00 €
30,00 €	150,00 €	60,00 €
52,00 €	160,00 €	73,00 €
45,00 €	170,00 €	62,00 €
46,00 €	180,00 €	57,00 €
47,00 €	185,00 €	60,00 €

Πίνακας 12. Δεδομένα Καταναλωτικών Δαπανών (Υ), Εισοδήματος (Χ1) και Δαπανών Στέγασης (Χ2) για 20 νοικοκυριά.

Με τη χρήση του στατιστικού πακέτου SPSS, είμαστε σε θέση να κάνουμε εκτίμηση των συντελεστών παλινδρόμησης και της ευθείας παλινδρόμησης.

Στον επόμενο πίνακα παρουσιάζονται οι τιμές του σταθερού όρου και των δυο συντελεστών των επεξηγηματικών μεταβλητών του διαθέσιμου εισοδήματος Χ1 και των δαπανών στέγασης Χ2.

ΣΥΝΤΕΛΕΣΤΕΣ						
		Εκτίμηση Συντελεστών		Τυποποιημένοι Συντελεστές		
	Υπόδειγμα	B	Τυπικά Σφάλματα	B	Στατιστική-t	P-τιμή Ελέγχου
1	Σταθερός Όρος	-40,604	7,632		-5,320	,000
	Εισόδημα_Χ1	,266	,074	,457	3,613	,002
	Στέγαση_Χ2	,586	,136	,543	4,295	,000

Πίνακας 13. Πίνακας συντελεστών πολλαπλής παλινδρόμησης, με εξαρτημένη τη μεταβλητή της Κατανάλωσης και ανεξάρτητες τις μεταβλητές του Εισοδήματος και των Δαπανών Στέγασης.

Όπως φαίνεται από τον Πίνακα 13, ο σταθερός όρος είναι ίσος με -40,604, ενώ ο συντελεστής του Εισοδήματος είναι 0,266 και ο συντελεστής των δαπανών στέγασης είναι ίσος με 0,586.

Από τα παραπάνω και σύμφωνα με την ερμηνεία των συντελεστών, συμπεραίνουμε ότι όταν οι επεξηγηματικές μεταβλητές πάρουν ταυτόχρονα την τιμή μηδέν, τότε η εξαρτημένη μεταβλητή Υ παίρνει την τιμή -40,604.

Όταν αυξηθεί η τιμή του εισοδήματος κατά μια μονάδα και η τιμή των δαπανών στέγασης παραμείνει σταθερή, τότε η εξαρτημένη μεταβλητή θα αυξηθεί κατά την τιμή του συντελεστή της μεταβλητής του εισοδήματος, 0,266.

Αντίστοιχα, όταν αυξηθεί η τιμή των δαπανών στέγασης κατά μια μονάδα και η τιμή του εισοδήματος παραμείνει σταθερή, τότε η εξαρτημένη μεταβλητή θα αυξηθεί κατά την τιμή του συντελεστή της μεταβλητής των δαπανών στέγασης, 0,586.

Άρα, η ευθεία παλινδρόμησης παίρνει τη μορφή:

$$Y = -40{,}604 + 0{,}266X_1 + 0{,}586X_2$$

4.3.2. Πρόβλεψη με το Πολλαπλό Γραμμικό Υπόδειγμα

Μετά την εκτίμηση της εξίσωσης πολλαπλής παλινδρόμησης, μπορούμε να τη χρησιμοποιήσουμε για να υπολογίσουμε τις τιμές των Καταναλωτικών Δαπανών (Y) που δίνονται από το υπόδειγμα, για τα δεδομένα επίπεδα του Εισοδήματος (X1) και των Στεγαστικών Δαπανών (X2). Με τον ίδιο τρόπο, μπορούμε να τη χρησιμοποιήσουμε και για την πρόβλεψη των Καταναλωτικών Δαπανών, για το κάθε επίπεδο Εισοδήματος και Στεγαστικών Δαπανών.

Στον ίδιο πίνακα υπολογίζονται και οι τιμές της διαφοράς μεταξύ παρατηρούμενων τιμών της Κατανάλωσης και των τιμών που προκύπτουν από το υπόδειγμα, οι οποίες παρουσιάζονται στη στήλη των Καταλοίπων.

Παρατηρούμενες Καταναλωτικές Δαπάνες (Y)	Διαθέσιμο Εισόδημα (X_1)	Δαπάνες Στέγασης (X_2)	Προβλεπόμενη Κατανάλωση ($\hat{Y}$)	Κατάλοιπα ($\hat{u}$)
5,00 €	110,00 €	30,00 €	6,24 €	- 1,24 €
11,00 €	120,00 €	35,00 €	11,83 €	- 0,83 €
18,00 €	143,00 €	40,00 €	20,87 €	- 2,87 €
24,00 €	150,00 €	42,00 €	23,91 €	0,09 €
29,00 €	152,00 €	45,00 €	26,20 €	2,80 €
35,00 €	162,00 €	52,00 €	32,96 €	2,04 €
37,00 €	180,00 €	57,00 €	40,68 €	- 3,68 €
42,00 €	182,00 €	61,00 €	43,55 €	- 1,55 €
43,00 €	180,00 €	61,00 €	43,02 €	- 0,02 €
48,00 €	175,00 €	63,00 €	42,86 €	5,14 €
49,00 €	195,00 €	65,00 €	49,36 €	- 0,36 €
51,00 €	196,00 €	70,00 €	52,55 €	- 1,55 €
53,00 €	183,00 €	75,00 €	52,02 €	0,98 €
40,00 €	190,00 €	72,00 €	52,13 €	- 12,13 €
34,00 €	150,00 €	65,00 €	37,39 €	- 3,39 €
30,00 €	150,00 €	60,00 €	34,46 €	- 4,46 €
52,00 €	160,00 €	73,00 €	44,73 €	7,27 €
45,00 €	170,00 €	62,00 €	40,95 €	4,05 €
46,00 €	180,00 €	57,00 €	40,68 €	5,32 €
47,00 €	185,00 €	60,00 €	43,77 €	3,23 €

Πίνακας 14. Τιμές Παρατηρούμενης Κατανάλωσης, Προβλεπόμενης Κατανάλωσης και Καταλοίπων για τις διάφορες τιμές του Εισοδήματος.

4.4 Υποθέσεις πολλαπλού γραμμικού υποδείγματος

Οι υποθέσεις του πολλαπλού γραμμικού υποδείγματος δεν διαφέρουν από αυτές του απλού γραμμικού υποδείγματος, όπως τις περιγράψαμε στο προηγούμενο κεφάλαιο, με τη μόνη διαφορά ότι στο πολλαπλό γραμμικό υπόδειγμα οι επεξηγηματικές μεταβλητές είναι περισσότερες από μια.

Έστω το πολλαπλό γραμμικό υπόδειγμα, όπως το έχουμε ορίσει, το οποίο έχει τη μορφή:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + u_i$$

Οι υποθέσεις που διέπουν το παραπάνω υπόδειγμα είναι οι εξής, (Χρήστου, 2008):

1. Η αναμενόμενη τιμή των καταλοίπων είναι μηδέν. Η υπόθεση αυτή έχει ήδη αναλυθεί ήδη, όπως ακριβώς ισχύει στην περίπτωση του απλού γραμμικού υποδείγματος.
   Συγκεκριμένα,

   $E(u_i) = 0$

2. Μια ακόμα υπόθεση που έχει αναλυθεί ήδη στην περίπτωση του απλού γραμμικού υποδείγματος είναι η υπόθεση της ομοσκεδαστικότητας, σύμφωνα με την οποία η διακύμανση των καταλοίπων παραμένει σταθερή και ίση με σ^2.
   Συμβολίζεται ως εξής:

   $\mathrm{Var}\,(u_i) = \sigma^2$.

3. Η συνδιακύμανση των καταλοίπων, για κατάλοιπα διαφορετικών χρονικών στιγμών ισούται με μηδέν. Δηλαδή, σε διαφορετικές χρονικές στιγμές, οι όροι των καταλοίπων δεν συσχετίζονται μεταξύ τους.
   Ο συμβολισμός της υπόθεσης αυτής είναι ο εξής:

   $\mathrm{Cov}\,(u_i, u_j) = 0, \forall\, i \neq j$

Για τις τιμές των καταλοίπων σε ίδιες χρονικές στιγμές, δηλαδή όταν $i = j$, τότε η συνδιακύμανση των καταλοίπων γίνεται Cov (u_i,u_i), η οποία ισούται με τη διακύμανση των καταλοίπων και η οποία, λόγω της προηγούμενης υπόθεσης πρέπει να παραμένει σταθερή και ίση με σ^2.
Ο συμβολισμός είναι ο εξής:

$\forall\ i \neq j$, Cov (u_i,u_j) =Cov (u_i,u_i) =Var $(u_i) = \sigma^2$.

4. Ο διαταρακτικός όρος (κατάλοιπα) u_i, είναι τυχαία μεταβλητή, η οποία ακολουθεί την κανονική κατανομή, με μέση τιμή μηδέν και διακύμανση σ^2.

$$u_i \sim N(0, \sigma^2)$$

5. Μια υπόθεση που δεν είναι κοινή με το απλό γραμμικό υπόδειγμα είναι ότι ο αριθμός του δείγματος (μέγεθος n) είναι μεγαλύτερος από τον αριθμό των προς εκτίμηση συντελεστών πολλαπλής παλινδρόμησης (όπως το έχουμε συμβολίσει k).

6. Οι τιμές που παίρνουν οι επεξηγηματικές μεταβλητές X_i παραμένουν σταθερές και μεταξύ τους δεν είναι όλες ίσες. Άρα, οι επεξηγηματικές μεταβλητές δεν είναι στοχαστικές, αλλά ντετερμινιστικές.

7. Τέλος, μια πολύ σημαντική υπόθεση είναι η υπόθεση ότι μεταξύ των ερμηνευτικών μεταβλητών δεν υπάρχουν ακριβείς γραμμικές σχέσεις. Υποθέτουμε, δηλαδή, ότι στο υπόδειγμα δεν υπάρχει πολυσυγγραμμικότητα.
   Όπως είναι φυσικό, αυτή η υπόθεση θα μπορούσε να υπάρχει μόνο στο πολλαπλό γραμμικό υπόδειγμα, καθώς στο απλό γραμμικό υπόδειγμα υπάρχει μόνο μια επεξηγηματική μεταβλητή.
   Η πολυσυγγραμμικότητα είναι ένα πρόβλημα το οποίο θα αναλυθεί σε επόμενο κεφάλαιο.

4.5 Κανονικές εξισώσεις - Εξίσωση ελαχίστων τετραγώνων (εξίσωση παλινδρόμησης)

Η μέθοδος ελαχίστων τετραγώνων, όπως και στην απλή παλινδρόμηση εφαρμόζεται ελαχιστοποιώντας τη συνάρτηση του αθροίσματος τετραγώνων των καταλοίπων Φ, με σκοπό την εκτίμηση των συντελεστών παλινδρόμησης $\hat{\beta}_i$.

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

$$= \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}\right)^2$$

Στη συνέχεια, ελαχιστοποιούμε την παραπάνω συνάρτηση παραγωγίζοντας k+1φορές τη συνάρτηση ως προς κάθε έναν από τους συντελεστές παλινδρόμησης $\hat{\beta}_0$ έως $\hat{\beta}_k$, και έπειτα εξισώνοντας την κάθε παράγωγο με το μηδέν.

Μετά την παραγώγιση, προκύπτουν οι παρακάτω συναρτήσεις:

$$\frac{\partial \Phi}{\partial \hat{\beta}_0} = 0 \Rightarrow -2 \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}\right) = 0$$

$$\frac{\partial \Phi}{\partial \hat{\beta}_1} = 0 \Rightarrow -2 \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}\right)(X_{i1}) = 0$$

.

.

.

$$\frac{\partial \Phi}{\partial \hat{\beta}_k} = 0 \Rightarrow -2 \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}\right)(X_{ik}) = 0$$

Λύνοντας το παραπάνω σύστημα k+1 εξισώσεων, μπορούμε να εκτιμήσουμε τους k+1 εκτιμητές παλινδρόμησης. Φαίνεται όμως ότι είναι περίπλοκη η επίλυση του συστήματος, για το λόγο αυτό χρησιμοποιήσουμε το συμβολισμό των πινάκων για την εκτίμηση των συντελεστών, που περιγράψαμε σε προηγούμενη παράγραφο.

4.6 Συντελεστής πολλαπλού προσδιορισμού - Προσαρμοσμένος (διορθωμένος) συντελεστής πολλαπλού προσδιορισμού

4.6.1 Ο συντελεστής πολλαπλού προσδιορισμού

Η έννοια του συντελεστή πολλαπλού προσδιορισμού δεν διαφέρει από το συντελεστή προσδιορισμού του απλού γραμμικού υποδείγματος. Δείχνει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής Y που ερμηνεύεται από το σύνολο των ερμηνευτικών μεταβλητών του πολλαπλού υποδείγματος.

Ορίζεται ως εξής:

$$R^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2} = 1 - \frac{\sum_{i=1}^{n} \hat{u}_i^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2} \Rightarrow$$

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$$

4.6.2 Υπολογισμός συντελεστή πολλαπλού προσδιορισμού

Για να υπολογίσουμε τα αντίστοιχα αθροίσματα τετραγώνων, αλλά και το συντελεστή πολλαπλού προσδιορισμού στο παράδειγμα των καταναλωτικών δαπανών των 20 νοικοκυριών, που εξαρτώνται από το διαθέσιμο εισόδημα και τις δαπάνες στέγασης, χρησιμοποιούμε τους παρακάτω πίνακες.

ΠΙΝΑΚΑΣ ΑΝΑΛΥΣΗΣ ΔΙΑΚΥΜΑΝΣΗΣ

	Υπόδειγμα	Άθροισμα Τετραγώνων	Βαθμοί Ελευθερίας	Μέσα Τετράγωνα	Στατιστική-F	P-τιμή Ελέγχου
1	Παλινδρόμηση	3258,311	2	1629,156	78,095	,000[b]
	Κατάλοιπα	354,639	17	20,861		
	Ολικό	3612,950	19			

Πίνακας 15. Τιμές Αθροισμάτων Τετραγώνων Παλινδρόμησης, Καταλοίπων και Συνολικό, στο παράδειγμα της απλής παλινδρόμησης.

Σύμφωνα με τον πίνακα 15, το άθροισμα τετραγώνων της παλινδρόμησης είναι 3258,311, το άθροισμα τετραγώνων των καταλοίπων 354,639, ενώ το συνολικό άθροισμα τετραγώνων ισούται με 3612,950.

Στον επόμενο πίνακα παρουσιάζεται η τιμή του συντελεστή πολλαπλής παλινδρόμησης του παραδείγματος.

Σύνοψη Υποδείγματος

Υπόδειγμα	Συντελεστής Συσχέτισης	Συντελεστής Πολλαπλού Προσδιορισμού	Προσαρμοσμένος Συντελεστής Πολλαπλού Προσδιορισμού	Τυπικό Σφάλμα της Εκτίμησης
1	,950	,902	,890	4,567

Πίνακας 16. Υπολογισμός τιμής Συντελεστή Πολλαπλού Προσδιορισμού R^2 του παραδείγματος.

Από τον πίνακα 16 προκύπτει ότι η τιμή του συντελεστή πολλαπλού προσδιορισμού του παραδείγματος ισούται με 0,902 και άρα το σύνολο των επεξηγηματικών μεταβλητών του υποδείγματος στο παράδειγμα, μπορεί να ερμηνεύσει το 90,2 % της συνολικής μεταβλητότητας της εξαρτημένης μεταβλητής των καταναλωτικών δαπανών.

4.6.3 Διορθωμένος Συντελεστής Προσδιορισμού

Ο διορθωμένος συντελεστής προσδιορισμού, ο οποίος συμβολίζεται με $\overline{R}^2$, είναι ένα μέτρο σύγκρισης μεταξύ δυο ή περισσότερων υποδειγμάτων αναφορικά με την επεξηγηματική τους ικανότητα, όταν τα υποδείγματα δεν έχουν τον ίδιο αριθμό επεξηγηματικών μεταβλητών και ίδιο μέγεθος δείγματος.

Οι παρακάτω τύποι δίνουν τον υπολογισμό του διορθωμένου συντελεστή προσδιορισμού.

$$\overline{R}^2 = 1 - \frac{\frac{\sum_{i=1}^{n} \hat{u}^2}{n-k-1}}{\frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1}} =$$

$$= 1 - \frac{\frac{SSE}{n-k-1}}{\frac{TSS}{n-1}} \Rightarrow$$

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

Στους παραπάνω τύπους, συμβολίζεται με n το μέγεθος του δείγματος, με k ο αριθμός των επεξηγηματικών μεταβλητών και με R^2 ο συντελεστής προσδιορισμού.

4.6.4 Υπολογισμός Διορθωμένου Συντελεστή Προσδιορισμού

Για τον υπολογισμό του διορθωμένου συντελεστή προσδιορισμού $\overline{R}^2$ στοπαράδειγμα, χρησιμοποιούμε τον τύπο:

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

$$= 1 - (1 - 0{,}902)\frac{20-1}{20-2-1} =$$

$$= 1 - (0{,}098)\frac{19}{17} =$$

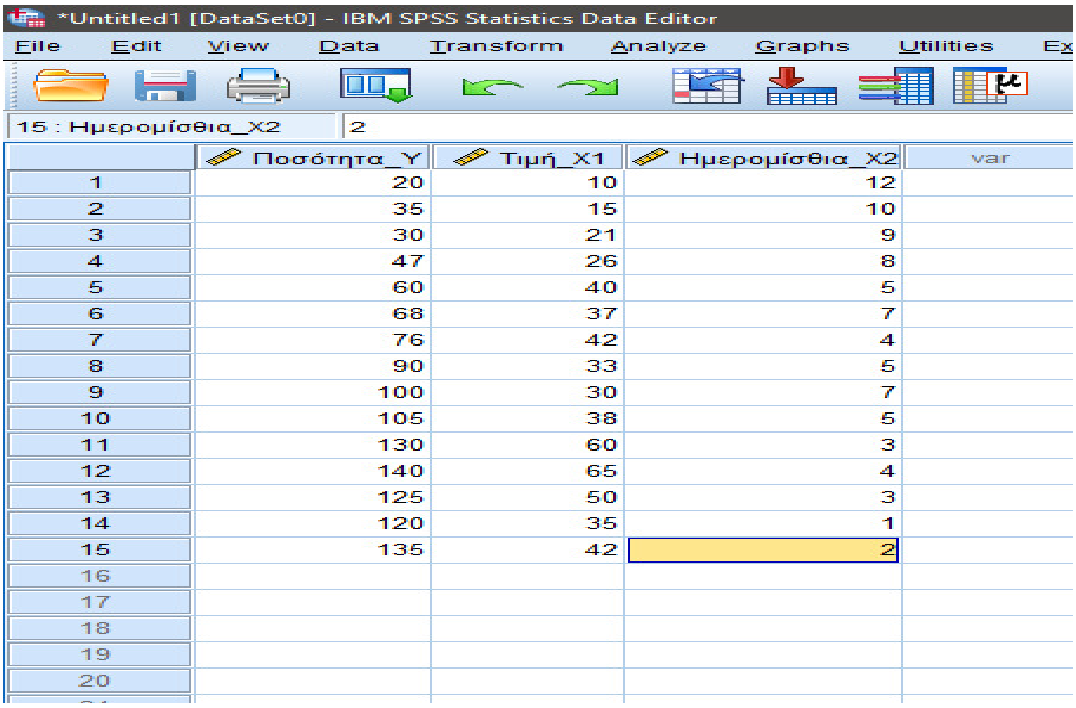
$$= 1 - 0{,}109 = 0{,}891$$

Η τιμή που υπολογίσαμε με τον τύπο που δόθηκε παραπάνω, συμπίπτει με την τιμή του διορθωμένου συντελεστή προσδιορισμού, όπως αυτή παρουσιάστηκε στον πίνακα 16.

4.7 Επίλυση Άσκησης Πολλαπλής Γραμμικής Παλινδρόμησης Με Χρήση του SPSS

Στην παρούσα υποενότητα θα χρησιμοποιήσουμε το SPSS ώστε να επιλύσουμε την άσκηση 2 της σελίδας 124 από το βιβλίο Εισαγωγή στην Οικονομετρία (Χρήστου, 2008).

Θα ξεκινήσουμε εισάγοντας τα δεδομένα σε ένα νέο αρχείο του SPSS όπως φαίνονται στην επόμενη εικόνα.



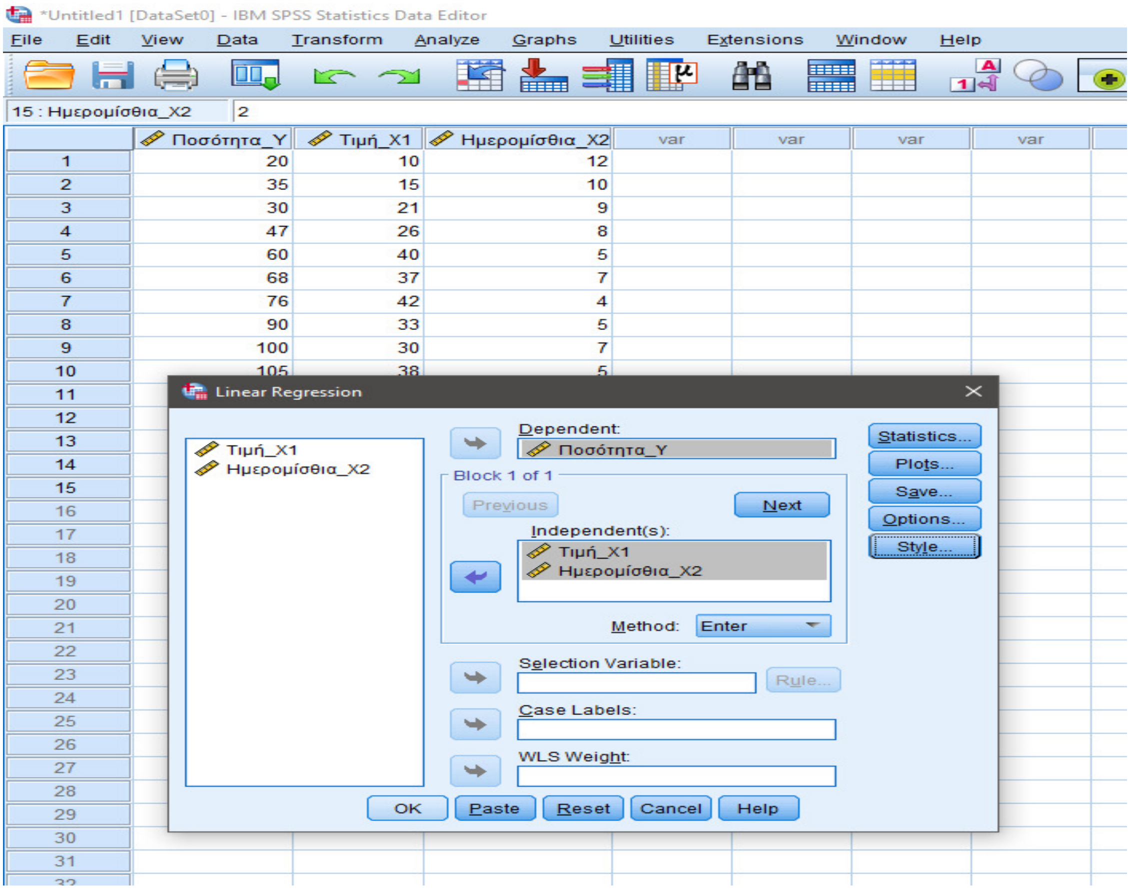
	Ποσότητα_Υ	Τιμή_Χ1	Ημερομίσθια_Χ2	var
1	20	10	12	
2	35	15	10	
3	30	21	9	
4	47	26	8	
5	60	40	5	
6	68	37	7	
7	76	42	4	
8	90	33	5	
9	100	30	7	
10	105	38	5	
11	130	60	3	
12	140	65	4	
13	125	50	3	
14	120	35	1	
15	135	42	2	
16				
17				
18				
19				
20				

Εικόνα 19. Δεδομένα της Άσκησης 2

Τα δεδομένα της Άσκησης 2 αφορούν τη σχέση της εξαρτημένης μεταβλητής Y η οποία περιγράφει την ζήτηση ενός αγαθού Ζ, ενώ οι μεταβλητές X_1 και X_2 αφορούν την τιμή του αγαθού Ζ και το ύψος των ημερομισθίων αντίστοιχα, σύμφωνα με το υπόδειγμα $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$.

Στο πρώτο ερώτημα πρέπει να εκτιμήσουμε τους συντελεστές του υποδείγματος β_0, β_1 και β_2.

Ακολουθούμε τη διαδικασία Analyze → Regression → Linear και στο παράθυρο που ανοίγει εισάγουμε στο πεδίο της εξαρτημένης μεταβλητής την Ποσότητα_Y και στο πεδίο των ανεξάρτητων μεταβλητών τις Τιμή_X1 και Ημερομίσθια_X2, όπως φαίνεται στην επόμενη εικόνα.



Εικόνα 20. Εισαγωγή των μεταβλητών του υποδείγματος στα πεδία της εξαρτημένης μεταβλητής (Y) και των ανεξάρτητων μεταβλητών (X1,X2)

Πατώντας OK προκύπτουν οι πίνακες στο SPSS ένας εκ των οποίων δίνει τις εκτιμήσεις των ζητούμενων συντελεστών, όπως φαίνονται στο παρακάτω πίνακα.

Πίνακας Συντελεστών						
		Εκτίμηση Συντελεστών		Τυποποιημένοι Συντελεστές	t-	P-τιμή
Υπόδειγμα		B	Τυπικό Σφάλμα	B	Στατιστική	Ελέγχου
1	Σταθερός Όρος	89,535	31,749		2,820	,015
	Τιμή_X1	1,053	,521	,391	2,023	,066
	Ημερομίσθια_X2	-7,470	2,529	-,570	-2,954	,012

Πίνακας 17. Πίνακας συντελεστών για το υπόδειγμα της Πολλαπλής Παλινδρόμησης

Όπως φαίνεται στον Πίνακα 18, οι εκτιμήσεις των συντελεστών είναι β_0 = 89,535, β_1 = 1,053 και β_2 = -7,47.

Στο δεύτερο ερώτημα ζητείται να υπολογίσουμε το ποσοστό της συνολικής μεταβλητικότητας της Y που ερμηνεύεται από την παλινδρόμηση, οπότε θα πρέπει να υπολογίσουμε το συντελεστή προσδιορισμού R^2.

Από την διαδικασία την οποία εφαρμόσαμε ήδη για το πρώτο ερώτημα (Analyze → Regression → Linear), προκύπτει ο πίνακας ο οποίος δίνει το συντελεστή προσδιορισμού και είναι ο παρακάτω.

Σύνοψη Υποδείγματος				
Υπόδειγμα	Συντελεστής Συσχέτισης	Συντελεστής Προσδιορισμού	Προσαρμοσμένος Συντελεστής Προσδιορισμού	Τυπικό Σφάλμα της Εκτίμησης
1	,909	,827	,798	18,302

Πίνακας 18. Πίνακας Συντελεστή Προσδιορισμού του Υποδείγματος Πολλαπλής Παλινδρόμησης

Από τον Πίνακα 18 προκύπτει ότι ο συντελεστής προσδιορισμού είναι ίσος με 0,827. Άρα το 82,7 % της μεταβλητικότητας της ζητούμενης ποσότητας του αγαθού Z ερμηνεύεται από τη μεταβλητικότητα της τιμής και του ύψους των ημερομισθίων.

Όπως το πρώτο και δεύτερο ερώτημα, έτσι και το τρίτο ερώτημα το οποίο ζητάει να ελέγξουμε τη στατιστική σημαντικότητα της συνδυασμένης επίδρασης των δυο ερμηνευτικών μεταβλητών, μπορούν να απαντηθούν από τους πίνακες που προέκυψαν από τη γνωστή διαδικασία του SPSSγια την παλινδρόμηση (Analyze → Regression → Linear).

Ο πίνακας που θα χρησιμοποιήσουμε είναι ο παρακάτω.

ΠΙΝΑΚΑΣ ΑΝΑΛΥΣΗΣ ΔΙΑΚΥΜΑΝΣΗΣ						
	Υπόδειγμα	Άθροισμα Τετραγώνων	Βαθμοί Ελευθερίας	Μέσα Τετράγωνα	Στατιστική-F	P-τιμή Ελέγχου
1	Παλινδρόμηση	19191,886	2	9595,943	28,647	,000[b]
	Κατάλοιπα	4019,714	12	334,976		
	Ολικό	23211,600	14			

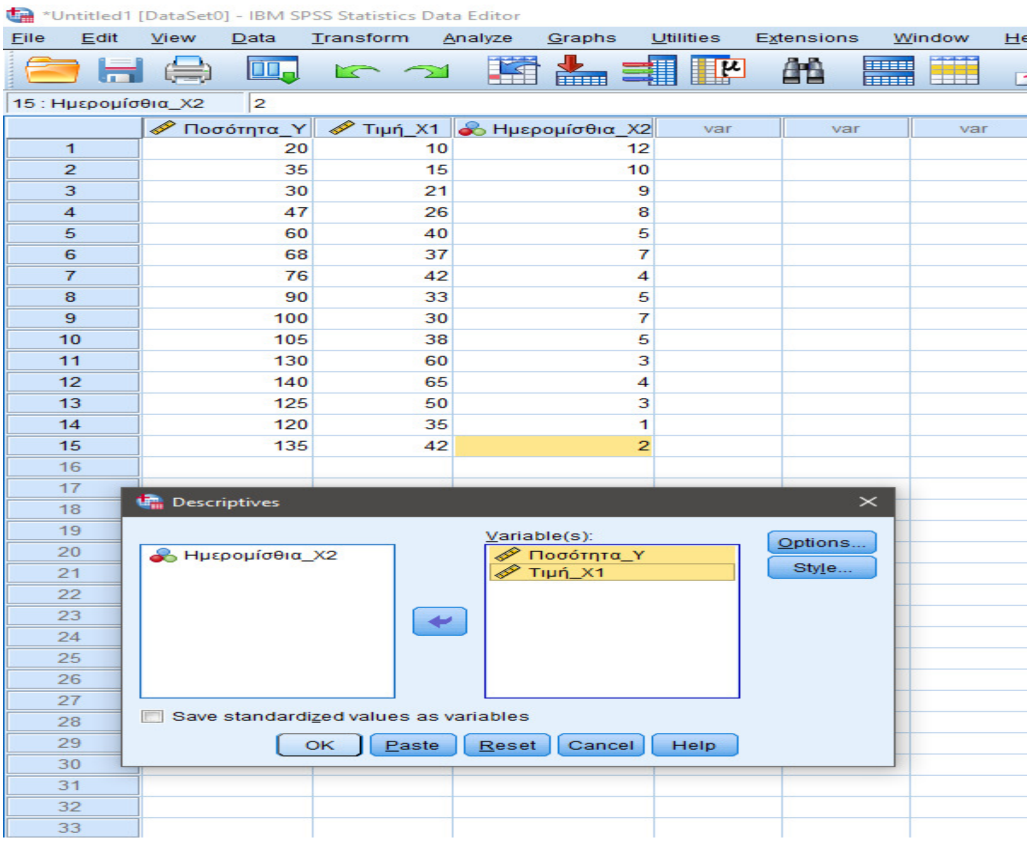
Πίνακας 19. Πίνακας Ανάλυσης Διακύμανσης του Υποδείγματος Πολλαπλής Παλινδρόμησης

Από τον πίνακα 19 φαίνεται ότι μπορούμε να απορρίψουμε τη μηδενική υπόθεση του ελέγχου η οποία αναφέρει ότι η συνδυασμένη επίδραση των ερμηνευτικών μεταβλητών είναι μηδενική και άρα μπορούμε να ισχυριστούμε ότι η συνδυασμένη επίδραση των ερμηνευτικών μεταβλητών είναι στατιστικά σημαντική.

Στο τέταρτο ερώτημα, το οποίο ζητάει να υπολογίσουμε την ελαστικότητα της ζήτησης του αγαθού Ζ ως προς την τιμή, στο σημείο των μέσων ($\bar{X}, \bar{Y}$), θα χρησιμοποιήσουμε τον εξής τύπο:

$$\varepsilon_{X1Y} = \beta_1 \frac{\bar{X}_1}{\bar{Y}}$$

Άρα θα πρέπει να υπολογίσουμε πρώτα τους μέσους των δυο μεταβλητών Y και X_1, σύμφωνα με τη διαδικασία Analyze → Descriptive Statistics → Descriptives και στο παράθυρο που ανοίγει εισάγουμε τις μεταβλητές Y και X_1 στο πεδίο των μεταβλητών που θέλουμε να υπολογίσουμε τα περιγραφικά στατιστικά, όπως φαίνεται στην παρακάτω εικόνα.



Πίνακας 20. Υπολογισμός Περιγραφικών Στατιστικών μεταβλητών Y και X1

Πατώντας OK προκύπτει ο παρακάτω πίνακας.

Περιγραφικά Στατιστικά					
	Μέγεθος Δείγματος	Ελάχιστη Τιμή	Μέγιστη Τιμή	Μέση Τιμή	Τυπική Απόκλιση
Ποσότητα_Y	15	20	140	85,40	40,718
Τιμή_X1	15	10	65	36,27	15,102
Έγκυρες Παρατηρήσεις	15				

Πίνακας 21. Περιγραφικά στατιστικά των μεταβλητών Y και X1

Άρα, ο μέσος της μεταβλητής Y είναι $\bar{Y} = 85{,}4$ και ο μέσος της μεταβλητής X_1 είναι $\bar{X}_1 = 36{,}27$.

Μπορούμε τώρα να υπολογίσουμε τη ζητούμενη ελαστικότητα ζήτησης, η οποία είναι:

$\varepsilon_{X1Y} = \beta_1 \frac{\bar{X}_1}{\bar{Y}} = 1{,}053 \frac{36{,}27}{85{,}4} = 0{,}447$.

Αφού η τιμή της ελαστικότητας ζήτησης είναι (κατ' απόλυτη τιμή) μικρότερη από τη μονάδα, μπορούμε να τη χαρακτηρίσουμε ως ανελαστική και άρα η ποσοστιαία μεταβολή της τιμής θα οδηγήσει σε μικρότερη μεταβολή της ζητούμενης ποσότητας.

Στο επόμενο ερώτημα της άσκησης ζητείται ο διορθωμένος συντελεστής προσδιορισμού.

Από τον Πίνακα 18 μπορούμε να δούμε ότι ο προσαρμοσμένος συντελεστής προσδιορισμού είναι ίσος με 0,798. Όπως έχουμε αναφέρει και στην αντίστοιχη ενότητα, ο προσαρμοσμένος συντελεστής προσδιορισμού χρησιμοποιείται όταν το υπόδειγμα παλινδρόμησης δεν έχει σταθερό όρο.

Στο τελευταίο ερώτημα της άσκησης απαιτείται να υπολογίσουμε τον συντελεστή μερικού προσδιορισμού $R^2_{y1.2}$, ο οποίος σύμφωνα με τη θεωρία θα μας δώσει τον συντελεστή προσδιορισμού μεταξύ της εξαρτημένης μεταβλητής Yκαι της ανεξάρτητης μεταβλητής X_1, όταν οι επιδράσεις της ανεξάρτητης μεταβλητής X_2 έχουν αφαιρεθεί.

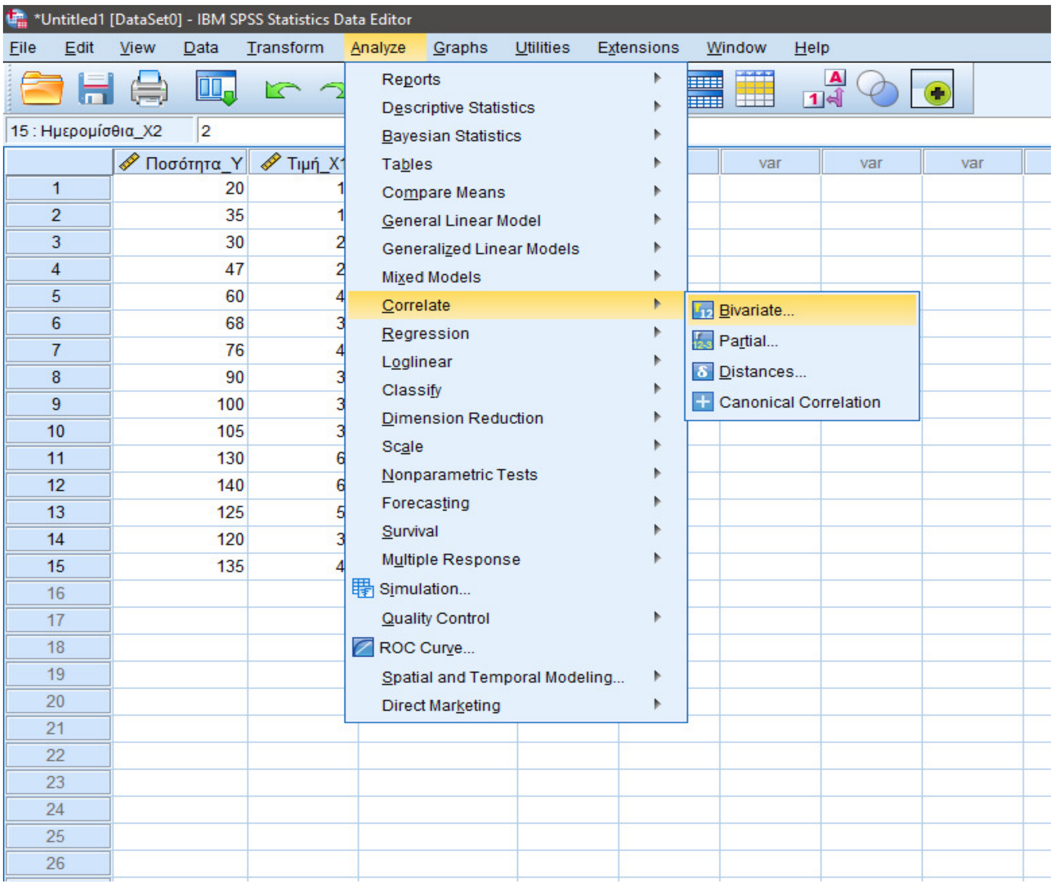
Ο τύπος που θα χρησιμοποιήσουμε είναι

$$R^2_{y1.2} = \frac{(r_{Y,X_1} - r_{Y,X_2 *} r_{X_1,X_2})^2}{[(1 - R^2_{X_1X_2})(1 - R^2_{YX_1})]}$$

Με r συμβολίζεται ο συντελεστής συσχέτισης μεταξύ των δυο μεταβλητών που υπάρχουν σαν δείκτης, ενώ με R^2 συμβολίζεται ο συντελεστής προσδιορισμού του υποδείγματος που περιέχει σαν μεταβλητές τις δυο μεταβλητές που υπάρχουν σαν δείκτης. Δηλαδή ο συμβολισμός r_{Y,X_2} αφορά το συντελεστή συσχέτισης μεταξύ των μεταβλητών Y και X_2, ενώ ο συμβολισμός $R^2_{YX_1}$ αφορά το συντελεστή προσδιορισμού που προκύπτει από το υπόδειγμα με εξαρτημένη μεταβλητή την Y και ως ανεξάρτητη μεταβλητή τη X_1.

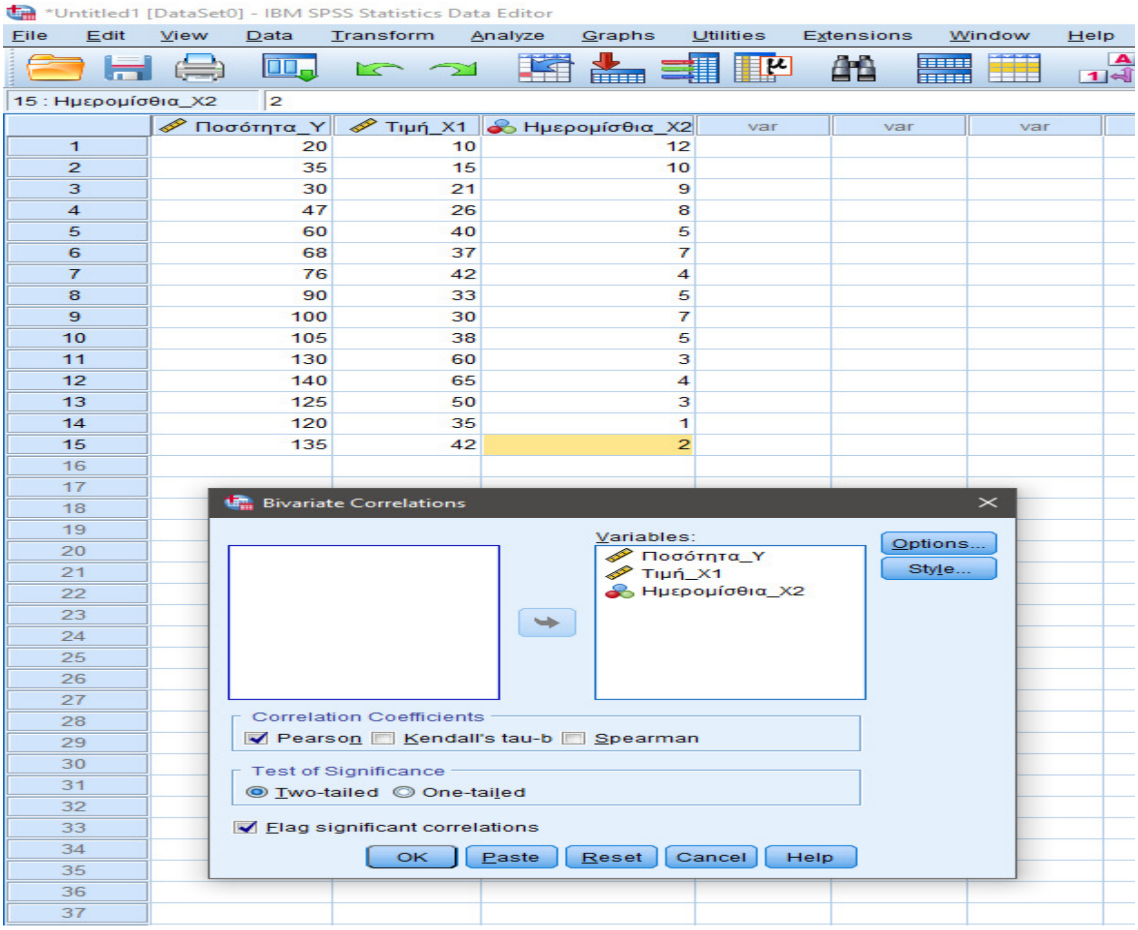
Οπότε, για να υπολογίσουμε την τιμή του συντελεστή μερικού προσδιορισμού θα πρέπει να υπολογίσουμε κάθε όρο ξεχωριστά.

Για να υπολογίσουμε τους συντελεστές συσχέτισης των τριών μεταβλητών μεταξύ τους, ακολουθούμε την εξής διαδικασία στο SPSS, Analyze → Correlate → Bivariate όπως φαίνεται στην επόμενη εικόνα.



Εικόνα 21. Διαδικασία υπολογισμού συντελεστών συσχέτισης των τριών μεταβλητών

Στο επόμενο παράθυρο που ανοίγει εισάγουμε στο πεδίο των μεταβλητών και τις τρεις μεταβλητές όπως φαίνεται στην επόμενη εικόνα.



Εικόνα 22. Εισαγωγή στο πεδίο των μεταβλητών των τριών μεταβλητών για τον υπολογισμό των συντελεστών συσχέτισης

Πατώντας το OK προκύπτει ο επόμενος πίνακας.

Πίνακας Συσχετίσεων				
		Ποσότητα_Y	Τιμή_X1	Ημερομίσθια_X2
Ποσότητα_Y	Συντελεστής Συσχέτισης Pearson	1	,837**	-,876**
	P-τιμή Ελέγχου		,000	,000
	Μέγεθος Δείγματος	15	15	15
Τιμή_X1	Συντελεστής Συσχέτισης Pearson	,837**	1	-,783**
	P-τιμή Ελέγχου	,000		,001
	Μέγεθος Δείγματος	15	15	15
Ημερομίσθια_X2	Συντελεστής Συσχέτισης Pearson	-,876**	-,783**	1
	P-τιμή Ελέγχου	,000	,001	
	Μέγεθος Δείγματος	15	15	15

Πίνακας 22. Πίνακας συντελεστών συσχέτισης μεταξύ των τριών μεταβλητών

Από τον πίνακα 22 μπορούμε να παρατηρήσουμε ότι

$r_{Y,X_1} = 0{,}837$

$r_{Y,X_2} = -0{,}876$

και

$r_{X_2,X_1} = -0{,}783$

Αυτό που απομένει να υπολογίσουμε είναι οι συντελεστές προσδιορισμού σε ένα υπόδειγμα που περιέχει τις μεταβλητές X_1 και X_2 για να υπολογίσουμε το $R^2_{X_1X_2}$ και σε ένα υπόδειγμα που περιέχει τις μεταβλητές Yκαι X_1 για να υπολογίσουμε το $R^2_{YX_1}$.

Σύμφωνα με τη γνωστή διαδικασία που περιγράφεται στις Εικόνες 7 και 8, για την απλή παλινδρόμηση, κατασκευάζουμε τα υποδείγματα, τα αποτελέσματα των οποίων παρουσιάζονται στους επόμενους πίνακες.

Σύνοψη Υποδείγματος

Υπόδειγμα	Συντελεστής Συσχέτισης	Συντελεστής Προσδιορισμού	Προσαρμοσμένος Συντελεστής Προσδιορισμού	Τυπικό Σφάλμα της Εκτίμησης
1	,783	,613	,583	9,749

Πίνακας 23. Πίνακας Συντελεστή Προσδιορισμού στο Υπόδειγμα με Εξαρτημένη Μεταβλητή την X1 και Ανεξάρτητη Μεταβλητή την X2

Σύνοψη Υποδείγματος

Υπόδειγμα	Συντελεστής Συσχέτισης	Συντελεστής Προσδιορισμού	Προσαρμοσμένος Συντελεστής Προσδιορισμού	Τυπικό Σφάλμα της Εκτίμησης
1	,837	,701	,678	23,109

Πίνακας 24. Πίνακας Συντελεστή Προσδιορισμού στο Υπόδειγμα με Εξαρτημένη Μεταβλητή την Y και Ανεξάρτητη Μεταβλητή την X1

Σύμφωνα με τους Πίνακες 23 και 24, ο συντελεστής προσδιορισμού του υποδείγματος με μεταβλητές τις X_1 και X_2 είναι

$R^2_{X_1 X_2}$ =0,613 ενώ για το υπόδειγμα με μεταβλητές τις Y και X_1 είναι

$R^2_{YX_1} =$ 0,701.

Άρα, χρησιμοποιώντας τις παραπάνω τιμές, μπορούμε να υπολογίσουμε το ζητούμενο συντελεστή μερικού προσδιορισμού.

$$R^2_{y1.2} = \frac{(r_{Y,X_1} - r_{Y,X_2 *} r_{X_1,X_2})^2}{[(1 - R^2_{X_1 X_2})(1 - R^2_{YX_1})]} = \frac{(0{,}837 - (-0{,}876) * (-0{,}783))^2}{[(1 - 0{,}613)(1 - 0{,}701)]}$$
$$= \frac{0{,}0228}{0{,}1157} = 0{,}197 = 19{,}7\%$$

Θα μπορούσαμε να πούμε ότι το 19,7 % της μεταβλητικότητας της μεταβλητής της προσφερόμενης ποσότητας του αγαθού Ζ (Y) ερμηνεύεται από τη μεταβλητικότητα της μεταβλητής της τιμής του αγαθού Ζ (X_1), έπειτα από την αφαίρεση των επιδράσεων της μεταβλητής των ημερομισθίων (X_2) πάνω στη μεταβλητή Y και επάνω στη μεταβλητή X_1.

Συμπεράσματα

Από την παρούσα εργασία, μπορούμε να εξάγουμε μερικά πολύ ενδιαφέροντα συμπεράσματα, που αφορούν κατά σειρά, στην απλή παλινδρόμηση και την πολλαπλή παλινδρόμηση όπως περιγράφονται στα κεφάλαια της εργασίας.

Αφού κάναμε μια εισαγωγή στο δεύτερο κεφάλαιο, στην απλή και πολλαπλή γραμμική παλινδρόμηση, όπως επίσης και στη μη γραμμική παλινδρόμηση, στο τρίτο κεφάλαιο χρησιμοποιήσαμε τη μεθοδολογία της απλής παλινδρόμησης για να εκτιμήσουμε τη σχέση μεταξύ του ετήσιου Εισοδήματος και της ετήσιας Κατανάλωσης 20 νοικοκυριών.

Μετά τη χρήση του διαγράμματος διασποράς των δυο αυτών μεταβλητών για να δούμε και διαγραμματικά τη σχέση που δημιουργείται μεταξύ τους, οι οποίες έχουν μια ισχυρή θετική συσχέτιση μεταξύ τους, όπως φαίνεται και στο διάγραμμα διασποράς, αλλά και στο συντελεστή συσχέτισης, ο οποίος υπολογίστηκε στο 0,968. Στη συνέχεια, και αφού διαπιστώσαμε τη γραμμική σχέση των δυο μεταβλητών, χρησιμοποιήσαμε τη μέθοδο ελαχίστων τετραγώνων ώστε να εκτιμήσουμε τους συντελεστές παλινδρόμησης, β_0 και β_1, οι οποίοι υπολογίστηκαν 0,727 και 0,587 αντίστοιχα. Τελικά, προέκυψε η εξίσωση παλινδρόμησης $Y = 0{,}727 + 0{,}587*X$, την οποία μπορούμε να χρησιμοποιήσουμε και για πρόβλεψη των τιμών της μεταβλητής της Κατανάλωσης (Y), για δεδομένη τιμή της μεταβλητής του Εισοδήματος (X).

Αναλυτικά, στον Πίνακα 4, παρουσιάζονται οι πραγματικές τιμές της Κατανάλωσης (Y) και του Εισοδήματος (X) στις δυο πρώτες στήλες, ενώ στην τρίτη στήλη παρουσιάζονται οι τιμές της Κατανάλωσης που προκύπτουν από τη χρήση του γραμμικού υποδείγματος παλινδρόμησης. Η τέταρτη στήλη είναι αυτή που παρουσιάζει ενδιαφέρον, καθώς περιέχει τη διαφορά μεταξύ των τιμών της μεταβλητής Y και της εκτίμησης της μεταβλητής που προκύπτει από το υπόδειγμα, το οποίο έχουμε ορίσει ως τα κατάλοιπα.

Στο τέλος του κεφαλαίου, ορίζουμε το μέτρο που ονομάζεται συντελεστής προσδιορισμού και συμβολίζεται με R^2, το οποίο δείχνει το ποσοστό της μεταβλητότητας της Κατανάλωσης που ερμηνεύεται από τη μεταβλητότητα του Εισοδήματος.

Στο παράδειγμα, ο συντελεστής R^2 υπολογίζεται σε 0,938, δηλαδή το ποσοστό της Κατανάλωσης που ερμηνεύεται από το Εισόδημα είναι αρκετά υψηλό και ίσο με 93,8 %.

Στο επόμενο κεφάλαιο γίνεται η επέκταση του απλού γραμμικού υποδείγματος, σε ένα υπόδειγμα με περισσότερες από μια ερμηνευτικές μεταβλητές, το οποίο ονομάζουμε πολλαπλό γραμμικό υπόδειγμα.

Σε παράδειγμα εκτίμησης του πολλαπλού υποδείγματος εκτίμησης των Καταναλωτικών Δαπανών, χρησιμοποιούμε ως ερμηνευτικές μεταβλητές το Διαθέσιμο Εισόδημα (X_1) και τις Δαπάνες Στέγασης (X_2) για 20 νοικοκυριά.

Κάνοντας χρήση της μεθόδου ελαχίστων τετραγώνων για να εκτιμήσουμε τους συντελεστές παλινδρόμησης, βρίσκουμε ότι ο σταθερός όρος της παλινδρόμησης είναι -40,604, ο συντελεστής του εισοδήματος είναι 0,266 και ο συντελεστής των στεγαστικών δαπανών είναι ίσος με 0,586. Η εξίσωση παλινδρόμησης είναι η:

$$Y = -40{,}604 + 0{,}266X_1 + 0{,}586X_2$$

Μετά την εκτίμηση του υποδείγματος, μπορούμε να το χρησιμοποιήσουμε για την πρόβλεψη των τιμών των καταναλωτικών δαπανών των νοικοκυριών, για δεδομένα επίπεδα διαθέσιμου εισοδήματος και δαπανών στέγασης, όπως φαίνονται και στον πίνακα 9.

Στο τέλος του κεφαλαίου υπολογίζουμε το συντελεστή πολλαπλού προσδιορισμού, ο οποίος δείχνει το ποσοστό των καταναλωτικών δαπανών που ερμηνεύεται από το σύνολο των ερμηνευτικών μεταβλητών.

Η τιμή του συντελεστή πολλαπλού προσδιορισμού είναι 0,902 και άρα το 90,2 % της μεταβλητότητας των καταναλωτικών δαπανών ερμηνεύεται από το διαθέσιμο εισόδημα και τις δαπάνες στέγασης.

Στο τέλος των κεφαλαίων 3 και 4, όπου ασχοληθήκαμε με την απλή και την πολλαπλή παλινδρόμηση, κάναμε χρήση των μεθόδων που αναπτύχθηκαν μέσα από το στατιστικό λογισμικό SPSS και καταφέραμε να εξάγουμε και πρακτικά μερικά από τα αποτελέσματα των μεθόδων.

Συνοψίζοντας, η εργασία αυτή παρουσιάζει αρκετά από τα ζητήματα των γραμμικών υποδειγμάτων θεωρητικά, αλλά και πρακτικά μέσα από τη χρήση του προγράμματος SPSS. Παρόλα αυτά, αρκετά ζητήματα χρήζουν μεγαλύτερης διερεύνησης στο μέλλον, όπως αυτά τα υποδείγματα τα οποία δεν είναι γραμμικά και φυσικά οι μεθοδολογίες που αυτά χρησιμοποιούν.

Βιβλιογραφία

1. Χρήστου Γ. , (2008), *Εισαγωγή στην Οικονομετρία, τόμος Α*, εκδόσεις Gutenberg

2. Αγιακλόγλου Χ., & Μπένος Θ., (2007), *Εισαγωγή στην Οικονομετρική Ανάλυση*, εκδόσεις Μπένου

3. Χαλικιάς Ι., (2010), *Στατιστική Μέθοδοι Ανάλυσης για Επιχειρηματικές Αποφάσεις*, εκδόσεις Rosili

4. Keller G., (2010), *Στατιστική για Οικονομικά και Διοίκηση Επιχειρήσεων*, εκδόσεις Επίκεντρο

5. Χαραλάμπους Χ., (2013), Περιήγηση στην Ιστορία της Στατιστικής. Ανακτήθηκε από : https://users.auth.gr/hara/courses/history_of_math/Spring2014/Presentation28_05_14.pdf