

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ  
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ  
ΕΠΙΣΤΗΜΗΣ



ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ  
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

**ΠΡΟΣΔΙΟΡΙΣΤΙΚΟΙ ΠΑΡΑΓΟΝΤΕΣ ΚΑΙ  
ΔΗΜΟΓΡΑΦΙΚΕΣ ΕΠΙΔΡΑΣΕΙΣ ΤΗΣ  
ΕΣΩΤΕΡΙΚΗΣ ΜΕΤΑΝΑΣΤΕΥΣΗΣ ΣΤΗΝ  
ΕΛΛΑΔΑ**

Τσιτσιρίγκου Α. Σταματία

Διπλωματική Εργασία

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

Πειραιάς

Μάρτιος 2014

Πανεπιστήμιο Πειραιώς

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ  
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ  
ΕΠΙΣΤΗΜΗΣ



ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ  
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

**ΠΡΟΣΔΙΟΡΙΣΤΙΚΟΙ ΠΑΡΑΓΟΝΤΕΣ ΚΑΙ  
ΔΗΜΟΓΡΑΦΙΚΕΣ ΕΠΙΔΡΑΣΕΙΣ ΤΗΣ  
ΕΣΩΤΕΡΙΚΗΣ ΜΕΤΑΝΑΣΤΕΥΣΗΣ ΣΤΗΝ  
ΕΛΛΑΔΑ**

Τσιτσιρίγκου Α. Σταματία

Διπλωματική Εργασία

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

Πειραιάς

Μάρτιος 2014

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από τη Γ.Σ.Ε.Σ. του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμό ..... συνεδρίαση του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική.

Τα μέλη της Επιτροπής ήταν :

- Καθηγητής , Τσίμπος Κλέων (Επιβλέπων)
- Επίκουρη Καθηγήτρια , Βερροπούλου Γεωργία
- Επίκουρος καθηγητής , Καλογήρου Σταμάτης

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

UNIVERSITY OF PIRAEUS

STATISTICS AND INSURANCE SCIENCE  
DEPARTMENT



MASTER PROGRAMME IN APPLIED STATISTICS

**DETERMINANTS AND DEMOGRAPHIC  
EFFECTS OF INTERNAL MIGRATION IN  
GREECE**

Tsitsirigkou L. Stamatia

Master Thesis

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

Piraeus

March 2014

Πανεπιστήμιο Πειραιώς

Πανεπιστήμιο Πειραιώς

*Στην οικογένεια μου ,*

Πανεπιστήμιο Πειραιώς



## Ευχαριστίες

Με την παρούσα διπλωματική εργασία ολοκληρώνω την φοίτηση μου στο μεταπτυχιακό πρόγραμμα σπουδών «Εφαρμοσμένη Στατιστική» του τμήματος «Στατιστικής και Ασφαλιστικής Επιστήμης» του Πανεπιστημίου Πειραιώς.

Θα ήθελα να εκφράσω τις ιδιαίτερες ευχαριστίες μου στον επιβλέποντα της διπλωματικής μου εργασίας, Καθηγητή κ. Κλέων Τσίμπο, καθώς και στα μέλη της επιτροπής, Επίκουρη Καθηγήτρια Γ. Βερροπούλου και Επίκουρο Καθηγητή Σ. Καλογήρου που με τις χρήσιμες και εποικοδομητικές τους παρατηρήσεις με καθοδήγησαν καθ' όλη την διάρκεια της συγγραφής της παρούσας εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου που με την στήριξη τους είναι δίπλα μου σε κάθε προσπάθεια.

Πανεπιστήμιο Πειραιώς

## Περίληψη

Το ζήτημα της εσωτερικής μετανάστευσης στην Ελλάδα τα τελευταία χρόνια αποτελεί ένα πολύ ενδιαφέρον αντικείμενο έρευνας . Οι αλλαγές που έχουν πραγματοποιηθεί τις τελευταίες δεκαετίες στον ελλαδικό χώρο έχουν προξενήσει σημαντικές ανακατανομές στον πληθυσμό της χώρας. Γεγονός το οποίο , εγείρει το ενδιαφέρον για τη μελέτη και τον προσδιορισμό εκείνων των παραγόντων , που επηρεάζουν την μετακίνηση των Ελλήνων από περιοχή σε περιοχή.

Σκοπός της παρούσας εργασίας είναι , με χρήση των δεδομένων από την απογραφή του 2001, να διερευνηθούν οι δημογραφικοί και κοινωνικο-οικονομικοί παράγοντες που ενδέχεται να επηρεάζουν την κινητικότητα των ατόμων μέσα στον ελλαδικό χώρο.

Το ποσοστό του αστικού πληθυσμού σε σχέση με τον συνολικό πληθυσμό της χώρας, το συνολικό ποσοστό της ανεργίας του πληθυσμού, ο μέσος αριθμός των ετών της ολοκληρωμένης εκπαιδευτικής δραστηριότητας του πληθυσμού, η συμμετοχή του εργατικού δυναμικού στον τριτογενή τομέα οικονομικής δραστηριότητας, η συμμετοχή του οικονομικά ενεργού πληθυσμού στο εργατικό δυναμικό , το κατά κεφαλήν εισόδημα , καθώς , επίσης , και η μακροχρόνια ανεργία , είναι οι παράγοντες που επιλέχθηκαν για να εξετασθούν για την επιρροή που , ίσως , ασκούν στην εσωτερική μετανάστευση.

Από την μελέτη που διεξήχθη , με χρήση του στατιστικού πακέτου SPSS , καθώς , επίσης, και με την χρήση της Πολλαπλής Γραμμικής Παλινδρόμησης (και ειδικότερα με την χρήση της μεθόδου backward elimination), βρέθηκε ότι το ποσοστό της ανεργίας και ο τριτογενής τομέας οικονομικής δραστηριότητας αποτελούν σημαντικούς παράγοντες που επηρεάζουν την εσωτερική μετανάστευση.

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ** : εσωτερική μετανάστευση , τριτογενής τομέας παραγωγής , ανεργία, πολλαπλή γραμμική παλινδρόμηση, backward elimination.

Πανεπιστήμιο Πειραιώς

## **Abstract**

The issue of internal migration in Greece in recent years is an important subject of research . Since changes introduced over the last decades in Greece have caused significant reallocation of the population of the country . This fact raises the interest in studying and identifying those factors that influence the migration of Greeks from region to region.

The purpose of this dissertation is to investigate demographic and socio - economic factors that may affect population mobility within the country , using data from the census of 2001. The percentage of urban population, the overall unemployment rate of the population , the mean number of years of completed educational level of the population , the tertiary sector of economic activity , participation in the workplace , per capita income and long-term unemployment , are the factors that were selected to examine their influence on internal migration .

The use of Multiple Linear Regression (particularly of the method of backward elimination), indicate that the rate of unemployment and the tertiary sector of economic activity are important factors influencing internal migration. The analysis was performed with SPSS.

**KEY WORDS:** internal migration, tertiary production sector, unemployment, multiple linear regression, backward elimination.

Πανεπιστήμιο Πειραιώς

## **ΠΕΡΙΕΧΟΜΕΝΑ**

ΚΕΦΑΛΑΙΟ 1 : ΜΕΤΑΝΑΣΤΕΥΣΗ .....	3
1.1 ΟΡΙΣΜΟΣ ΜΕΤΑΝΑΣΤΕΥΣΗΣ.....	3
1.2 ΚΑΤΗΓΟΡΙΕΣ ΜΕΤΑΝΑΣΤΕΥΣΗΣ.....	3
1.2.1 Εσωτερική και Εξωτερική μετανάστευση.....	3
1.2.2 Προσωρινή και Μόνιμη μετανάστευση.....	3
1.2.3 Εκούσια και Ακούσια μετανάστευση.....	3
1.2.4 Οργανωμένη και μη οργανωμένη μετανάστευση.....	4
1.2.5 Ατομική και ομαδική μετανάστευση.....	4
1.2.6 Νόμιμη και παράνομη μετανάστευση .....	4
1.2.7 Επαγγελματική και Δημογραφική μετανάστευση .....	4
ΚΕΦΑΛΑΙΟ 2 : Η ΕΣΩΤΕΡΙΚΗ ΜΕΤΑΝΑΣΤΕΥΣΗ ΣΕ ΒΑΘΟΣ ΧΡΟΝΟΥ .....	5
2.1 Η ΕΣΩΤΕΡΙΚΗ ΜΕΤΑΝΑΣΤΕΥΣΗ ΣΤΗΝ ΕΛΛΑΔΑ ΠΡΙΝ ΤΟ Β΄ ΠΑΓΚΟΣΜΙΟ ΠΟΛΕΜΟ .....	5
2.2 Η ΕΣΩΤΕΡΙΚΗ ΜΕΤΑΝΑΣΤΕΥΣΗ ΣΤΗΝ ΕΛΛΑΔΑ ΚΑΤΑ ΤΗ ΔΙΑΡΚΕΙΑ ΤΟΥ Β΄ ΠΑΓΚΟΣΜΙΟΥ ΠΟΛΕΜΟΥ .....	5
2.3 Η ΕΣΩΤΕΡΙΚΗ ΜΕΤΑΝΑΣΤΕΥΣΗ ΣΤΗΝ ΕΛΛΑΔΑ ΜΕΤΑ ΤΟ Β΄ ΠΑΓΚΟΣΜΙΟ ΠΟΛΕΜΟ .....	6
ΚΕΦΑΛΑΙΟ 3 : ΠΗΓΕΣ ΔΗΜΟΓΡΑΦΙΚΩΝ ΔΕΔΟΜΕΝΩΝ.....	11
ΚΕΦΑΛΑΙΟ 4 : ΑΝΑΔΡΟΜΗ ΣΕ ΥΠΑΡΧΟΥΣΑ ΕΡΕΥΝΑ.....	13
ΚΕΦΑΛΑΙΟ 5 : ΠΕΡΙΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ.....	15
ΚΕΦΑΛΑΙΟ 6 : ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ .....	17
6.1 ΜΕΤΡΑ ΘΕΣΗΣ.....	17
6.1.1 Δειγματική μέση τιμή (ή αριθμητικός μέσος ή μέσος όρος).....	17
6.1.2 Δειγματική διάμεσος .....	17
6.1.3 Δειγματική επικρατούσα τιμή .....	18
6.2 ΜΕΤΡΑ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ (ΔΙΑΣΠΟΡΑΣ).....	18
6.2.1 Δειγματικό εύρος.....	18
6.2.2 Δειγματική διακύμανση ή διασπορά .....	19
6.2.3 Δειγματική τυπική απόκλιση.....	19
ΚΕΦΑΛΑΙΟ 7 : ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ.....	21
7.1 ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ.....	22
7.1.1 Το μοντέλο .....	22

7.1.3 Η μεταβλητότητα του μοντέλου .....	24
7.1.4 Έλεγχος υποθέσεων για τις παραμέτρους του μοντέλου .....	26
7.2 ΠΟΛΛΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ .....	27
7.2.1 Το μοντέλο .....	27
7.2.2 Η εκτίμηση των παραμέτρων .....	27
7.2.3 Η μεταβλητότητα του μοντέλου .....	28
7.2.4 Έλεγχος υποθέσεων για τις παραμέτρους του μοντέλου .....	30
7.2.5 Επιλογή βέλτιστου γραμμικού μοντέλου .....	31
ΚΕΦΑΛΑΙΟ 8: ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ .....	35
8.1 ΠΕΡΙΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ .....	35
8.2 ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ .....	40
8.2.1 Μελέτη της κανονικότητας της εξαρτημένης μεταβλητής .....	41
8.2.2 Σχέση σύνδεσης των μεταβλητών .....	43
8.2.3 ΔΙΑΓΡΑΜΜΑΤΑ ΔΙΑΣΠΟΡΑΣ ΜΕΤΑΞΥ ΜΕΤΑΒΛΗΤΩΝ .....	45
8.2.4 Πρώτη επιλογή βέλτιστου μοντέλου .....	49
8.2.5 Backward elimination .....	53
8.2.6 Ανάλυση σφαλμάτων .....	59
ΚΕΦΑΛΑΙΟ 9 : ΣΥΜΠΕΡΑΣΜΑΤΑ .....	65
ΒΙΒΛΙΟΓΡΑΦΙΑ .....	69
ΠΑΡΑΡΤΗΜΑ Α .....	71



## ΚΕΦΑΛΑΙΟ 1 : ΜΕΤΑΝΑΣΤΕΥΣΗ

### 1.1 ΟΡΙΣΜΟΣ ΜΕΤΑΝΑΣΤΕΥΣΗΣ

Το φαινόμενο της μετανάστευσης, λόγω της πολυπλοκότητας του και των διλημάτων –πολιτικών, ιστορικών, κοινωνικών κ.ά. που θέτει, δεν είναι εύκολο να εξετασθεί κατά την εξέλιξη του.

Με τον όρο «μετανάστευση» αναφερόμαστε σε κάθε μετακίνηση ενός ατόμου από μια χώρα ή μια πόλη σε μια άλλη, με σκοπό την προσωρινή, ή ακόμα και την μόνιμη, εγκατάσταση του σε αυτή.

Η έννοια προέρχεται από το ρήμα «μεταναστεύω» και αποτελεί μια μορφή πληθυσμιακής κινητικότητας μέσα στον περιβάλλοντα χώρο.

### 1.2 ΚΑΤΗΓΟΡΙΕΣ ΜΕΤΑΝΑΣΤΕΥΣΗΣ

Η μετανάστευση μπορεί να χαρακτηριστεί από τα παρακάτω ζεύγη εννοιών:

#### 1.2.1 Εσωτερική και Εξωτερική μετανάστευση

Ως εξωτερική μετανάστευση εννοούμε το φαινόμενο της μετακίνησης του πληθυσμού από την πατρική γη σε ξένες χώρες.

Αντίστοιχα, ως εσωτερική μετανάστευση εννοούμε το φαινόμενο της μετακίνησης του πληθυσμού σε άλλο τόπο (περιοχή ή πόλη) της ίδιας χώρας.

Το ρεύμα των εκροών στην εξωτερική μετανάστευση ονομάζεται αποδημία, ενώ στην εσωτερική μετανάστευση ονομάζεται εκδημία.

Το ρεύμα των εισροών στην εξωτερική μετανάστευση αναφέρεται ως μετοικία, ενώ στην εσωτερική μετανάστευση αναφέρεται ως εισδημία.

#### 1.2.2 Προσωρινή και Μόνιμη μετανάστευση

Η μετανάστευση διακρίνεται σε προσωρινή όταν το διάστημα παραμονής του πληθυσμού στο «νέο» τόπο δεν υπερβαίνει το ένα έτος.

Αντίστοιχα, η μετανάστευση αναφέρεται ως μόνιμη, όταν το διάστημα παραμονής του πληθυσμού στο «νέο» τόπο υπερβαίνει το ένα έτος.

#### 1.2.3 Εκούσια και Ακούσια μετανάστευση

Εκούσια μετανάστευση έχουμε όταν ένα άτομο «θελήσει» να μετοικήσει σε μια άλλη περιοχή ή σε μια άλλη χώρα. Για παράδειγμα, ένας φοιτητής που σπουδάζει

στο εξωτερικό για μια 4ετία και επιστρέφει ξανά στην χώρα του, μετά το τέλος του διαστήματος αυτού επιλέγει και αποφασίζει ο ίδιος για την κίνηση αυτή.

Αντίθετα, ακούσια μετανάστευση έχουμε όταν ένα άτομο αναγκάζεται να εγκαταλείψει τον τόπο κατοικίας και εργασίας του για έναν άλλον τόπο. Οι λόγοι που τον ωθούν, συνήθως, σε τέτοιες κινήσεις είναι οι δυσμενείς συνθήκες διαβίωσης, οικολογικές καταστροφές, πόλεμοι κ.λ.π.

#### 1.2.4 Οργανωμένη και μη οργανωμένη μετανάστευση

Ανάλογα με το βαθμό οργάνωσης που παρουσιάζουν, οι μεταναστευτικές κινήσεις, διακρίνονται σε οργανωμένες και μη οργανωμένες.

#### 1.2.5 Ατομική και ομαδική μετανάστευση

Η μεταναστευτική κίνηση μπορεί να διακριθεί σε ατομική και ομαδική (μαζική), ανάλογα με το μέγεθος του μετακινούμενου πληθυσμού.

Διακρίνεται σε ατομική όταν είναι αποτέλεσμα απόφασης μεμονωμένων ατόμων και σε μαζική όταν πρόκειται για μορφή μετακίνησης κατά μάζες.

#### 1.2.6 Νόμιμη και παράνομη μετανάστευση

Ανάλογα με το καθεστώς της εισόδου και παραμονής στην χώρα υποδοχής, η μεταναστευτική κίνηση, διακρίνεται σε νόμιμη και παράνομη.

Νόμιμοι μετανάστες είναι εκείνοι που έχουν εισέλθει και παραμένουν νόμιμα στη χώρα, έχουν καταγραφεί από τις αρμόδιες αρχές και είναι κάτοχοι άδειας παραμονής και εργασίας.

Παράνομοι μετανάστες καλούνται οι αλλοδαποί που είτε εισήλθαν στην χώρα εξαρχής χωρίς νόμιμα έγγραφα (λαθρομετανάστες), είτε εισήλθαν νόμιμα (ως τουρίστες, φοιτητές κλπ) αλλά παρέμειναν μόνιμα παράνομα στη χώρα.

#### 1.2.7 Επαγγελματική και Δημογραφική μετανάστευση

Αν ένας πληθυσμός που μεταναστεύει αποτελείται από άτομα εργαζόμενα, που ζητούν καλύτερες συνθήκες εργασίας σε έναν νέο τόπο, αναφέρουμε την μετακίνηση τους αυτή ως επαγγελματική μετανάστευση. Από την άλλη, η μετακίνηση που έχει ως απώτερο σκοπό τον επαναπατρισμό των ανθρώπων και την επανασύνδεση τους με την οικογένεια τους, αναφέρεται ως δημογραφική μετανάστευση. (Παπαδάκης -Τσίμπος, 2004)

## ΚΕΦΑΛΑΙΟ 2 : Η ΕΣΩΤΕΡΙΚΗ ΜΕΤΑΝΑΣΤΕΥΣΗ ΣΕ ΒΑΘΟΣ ΧΡΟΝΟΥ

### 2.1 Η ΕΣΩΤΕΡΙΚΗ ΜΕΤΑΝΑΣΤΕΥΣΗ ΣΤΗΝ ΕΛΛΑΔΑ ΠΡΙΝ ΤΟ Β΄ ΠΑΓΚΟΣΜΙΟ ΠΟΛΕΜΟ

Ξεκινώντας την μελέτη μας από τον 19<sup>ο</sup> αιώνα και μέχρι και τον 20<sup>ο</sup> αιώνα, δεν υπάρχουν σημαντικές καταγραφές για εσωτερικές μετακινήσεις στον Ελλαδικό χώρο. Αξίζει να σημειωθεί, φυσικά, ότι τα σύνορα της Ελλάδας τότε ήταν στα Ιόνια το 1864 και στη Θεσσαλία το 1881.

Τότε δεν ήταν διαδεδομένη η αστικοποίηση, διότι είχε ιδιαίτερη ανάπτυξη η μικρή γαιοκτησία καθώς, επίσης, και η ενασχόληση με τις παραδοσιακές καλλιέργειες.

Το κέντρο της Θεσσαλονίκης μπήκε δυναμικά στο χώρο μετά τους Βαλκανικούς πολέμους το 1913. Εκείνη την εποχή, τα σύνορα της Ελλάδας επεκτείνονται, και, συνεπώς αυξάνεται και ο πληθυσμός της. Αυτό είχε ως συνέπεια να δοθεί η δυνατότητα να αναπτυχθεί ο βιομηχανικός τομέας.

Μέχρι το 1920 αναφέρεται ότι το εργατικό δυναμικό της υπαίθρου προτιμούσε την διεθνή μετανάστευση, παρά την αστικοποίηση του στην χώρα του. Το 1920, το 23% του ελληνικού πληθυσμού ζούσε σε πόλεις άνω των 10.000 κατοίκων, ενώ, μόλις το 8% ζούσε στην Αθήνα.

Πολύ σημαντικό ρόλο στις μετακινήσεις έπαιξε η Μικρασιατική Καταστροφή. Το 1922 η Ελλάδα υποδέχεται στις εκτάσεις της περίπου 1,2 εκατομμύρια πρόσφυγες. Παρατηρείται, λοιπόν, μια αύξηση της τάξεως του 30% σε αστικά κέντρα και 13% στην Αθήνα. Αυξήσεις που δεν οφείλονται σε εσωτερικές μετακινήσεις στο χώρο, αλλά στον μεγάλο αριθμό προσφύγων που εισήλθαν στη χώρα. ( Γενική Στατιστική Υπηρεσία Ελλάδος, 1935)

### 2.2 Η ΕΣΩΤΕΡΙΚΗ ΜΕΤΑΝΑΣΤΕΥΣΗ ΣΤΗΝ ΕΛΛΑΔΑ ΚΑΤΑ ΤΗ ΔΙΑΡΚΕΙΑ ΤΟΥ Β΄ ΠΑΓΚΟΣΜΙΟΥ ΠΟΛΕΜΟΥ

Ο Β΄ Παγκόσμιος πόλεμος αποτέλεσε την πιο εκτεταμένη γεωγραφικά και την πιο δαπανηρή ένοπλη σύγκρουση στην ιστορία της ανθρωπότητας. Κόστισε περίπου 55,5 εκατομμύρια ζωές. Ξεκίνησε από την Ασία το 1939, έφτασε στην Ευρώπη το 1939 και τελείωσε το 1945.

Ήταν ένα γεγονός που απέφερε πολλές ζημιές και άλλαξε τις ζωές πολλών ανθρώπων σε όλο τον κόσμο.

Όσων αφορά τις μετακινήσεις των πληθυσμών εκείνης της περιόδου, τα διαθέσιμα στοιχεία είναι πολύ περιορισμένα. Λόγω των συνθηκών της εμπόλεμης περιόδου, οι πληθυσμιακές μεταναστεύσεις δεν έχουν μελετηθεί ιδιαίτερα, άλλοτε είναι βίαιες,

άλλοτε εκούσιες, άλλοτε υποτάσσονται σε στρατιωτικές επιχειρήσεις (π.χ. εκκενώσεις χωριών) και σε μεγάλη έκταση. Υπολογίζεται ότι γύρω στο ένα δέκατο (1/10) του πληθυσμού μετακινήθηκε στην περίοδο 1940-1946 και γύρω στις 600.000 άτομα μετακινήθηκαν στην περίοδο του εμφυλίου πολέμου. (Κοτζαμάνης, 1990) Με αντίθετες κάθε φορά κατευθύνσεις – πόλη - χωριό, χωριό - πόλη - οι μετακινήσεις αυτές φέρνουν τον πληθυσμό των περιοχών της υπαίθρου, ειδικά αυτών που έγιναν θέατρο των πολεμικών και εμφυλιακών συγκρούσεων, σε γνωριμία και επαφή με τον αστικό τρόπο ζωής.

Οι άνθρωποι εκείνης της εποχής είχαν την δυνατότητα να έρθουν σε στενότερη επαφή με τα αστικά κέντρα και να γνωρίσουν έναν άλλον τρόπο ζωής. Το Υφυπουργείο Ανοικοδόμησης χρησιμοποίησε ένα σχέδιο ώστε όσοι είχαν αποχωριστεί τα χωριά τους λόγω της καταστροφής του β' παγκοσμίου πολέμου να επιστρέψουν πίσω σε αυτά. Φυσικά, αυτό είχε ως αποτέλεσμα κάποιοι να επιστρέψουν στα χωριά και στον παλιό τρόπο ζωής τους αλλά και κάποιοι να παραμείνουν σε ένα αστικό κέντρο. (Υφυπουργείον Ανοικοδόμησης, 1946)

## 2. 3 Η ΕΣΩΤΕΡΙΚΗ ΜΕΤΑΝΑΣΤΕΥΣΗ ΣΤΗΝ ΕΛΛΑΔΑ ΜΕΤΑ ΤΟ Β' ΠΑΓΚΟΣΜΙΟ ΠΟΛΕΜΟ

Όπως προαναφέρθηκε, η δεκαετία του '40 ήταν καθοριστική. Αν και είχαμε λίγα αριθμητικά στοιχεία για το τι συνέβη εκείνα τα χρόνια, η δεκαετία εκείνη έπαιξε πολύ σημαντικό ρόλο για την μετέπειτα πορεία του πληθυσμού. Από εκείνη την περίοδο και μετά η εσωτερική μετανάστευση ξεκίνησε την δυναμική της πορεία στον Ελλαδικό χώρο.

Για την μέχρι το 1950 περίοδο θα μπορούσε κανείς να πει ότι η φύση της αστικής ανάπτυξης και τα συνδεδεμένα με αυτήν σχήματα εξωτερικής μετανάστευσης, οφείλονταν περισσότερο σε εξωτερικές πολιτικές και στρατιωτικές πιέσεις και συνθήκες και παρά τις εσωτερικές κοινωνικές, οικονομικές και δημογραφικές συνθήκες.

Από την δεκαετία του '50 έως και τα μέσα του 1970 παρουσιάστηκαν μεγάλες μετακινήσεις προς την Αθήνα και την Θεσσαλονίκη. Οι δύο κεντρικές αστικές πόλεις της χώρας συγκέντρωναν το μεγαλύτερο μέρος της βιομηχανίας. Γεγονός που μετέτρεπε τις πόλεις αυτές σε πόλους έλξης για το ανθρώπινο εργατικό δυναμικό της χώρας (με βελτιώσεις στον αγροτικό τομέα όσο και με έργα οδοποιίας, καθώς και με την ενίσχυση της επαρχιακής βιομηχανίας που λόγω προσβασιμότητας συγκεντρώνεται στους νομούς γύρω από την Αθήνα). Αρχίζει, λοιπόν, η μεγάλη έξοδος προς τις πόλεις. Η Αθήνα δέχεται τον κύριο όγκο των εσωτερικών μεταναστών και αυξάνει τον πληθυσμό της κατά 34%, χωρίς αυτό να είναι αποτέλεσμα της φυσικής της αύξησης. Οι εσωτερικές μεταναστεύσεις έχουν σε μεγάλο μέρος τους ενδοαγροτικό χαρακτήρα. Πρόκειται για μεταναστεύσεις από χωριό σε άλλο χωριό και υποθέτουμε ότι αφορούν μεταναστεύσεις λόγω γάμου, εκπαίδευσης, μετεγκαταστάσεις πληθυσμών από τα ορεινά στα πεδινά κ.ά. Είναι μεταναστεύσεις οικονομικά ενεργών ατόμων πιο παραγωγικών ηλικιών που

ασχολούνται τόσο στον τομέα των υπηρεσιών, όσο και στην μεταποίηση βιομηχανία, αλλά και επειδή μεγάλο μέρος των εσωτερικών μεταναστών μετακινούνται από χωριό σε χωριό έχουν σημαντική απασχόληση στον πρωτογενή τομέα παραγωγής.

Συνεχίζοντας, με την δεκαετία του '60, παρατηρήθηκε μια σημαντική βελτίωση στον αγροτικό χώρο. Παρ' όλα αυτά οι συνθήκες αυτές δεν αποτέλεσαν σημαντικό παράγοντα για την επιστροφή στις αγροτικές περιοχές. Αντίθετα, σημειώθηκε μια αύξηση στο μέγεθος των αστικών κέντρων και συρρίκνωση του πληθυσμού της υπαίθρου.

Στο τέλος της δεκαετίας του 1970, παρουσιάζεται η πρώτη σημαντική επιστροφή στην ύπαιθρο. Τα στατιστικά στοιχεία έδειξαν μια πτώση του ρυθμού αύξησης του πληθυσμού των αστικών κέντρων Αθηνών και Θεσσαλονίκης.

Σημαντική παρατήρηση της εποχής ήταν ότι να μεν μειώθηκαν οι μετακινήσεις από την ύπαιθρο στις πόλεις αλλά μειώθηκαν και οι μετακινήσεις από και προς τις αγροτικές περιοχές. Έτσι καταγράφηκαν οι πρώτες ενδείξεις για εσωτερικές μετακινήσεις από τις αστικές περιοχές στις αγροτικές.

#### • ΠΕΡΙΟΔΟΣ 1961-1970

Πολιτική περίοδος πολιτικά ασταθής με κατάληξη την κατάλυση της δημοκρατίας από τη χούντα. Στις 21 Απριλίου 1967, μετά το πραξικόπημα των Συνταγματαρχών, επικράτησε η περίοδος της δικτατορίας στην Ελλάδα, γνωστή και με τον όρο «Χούντα» (ή αλλιώς Χούντα των Συνταγματαρχών ή Δικτατορία των Συνταγματαρχών).

Η περίοδος αυτή κράτησε επτά (7) χρόνια. Σταμάτησε, δηλαδή, στις 23 Ιουλίου 1974. Η πενταετία 1975-1981 που ακολούθησε, μετά την πρώτη μεταπολιτευτική περίοδο στην Ελλάδα, είχε και τον μεγαλύτερο όγκο εσωτερικών μεταναστών. Οικονομικά έχουμε ανάπτυξη της αγροτικής παραγωγής (εκμηχάνιση, αύξηση αγροτικού εισοδήματος, ανάπτυξη της βιομηχανίας). Η γρήγορη ανάπτυξη των παραγωγικών τομέων (πρωτογενούς και δευτερογενούς) αφενός ελευθερώνει δυναμικό από την γεωργία και τις αγροτικές περιοχές, αφετέρου δημιουργεί θέσεις εργασίας στις αστικές προκαλώντας έτσι μετακίνηση προς αυτές. Γενικά η εσωτερική μετανάστευση βρίσκεται στην πλήρη εξέλιξη της. Αρχίζει κάμψη της ενδοαγροτικής μετανάστευσης. Πάλι ο αγροτικός πληθυσμός τροφοδοτεί την Αθήνα, τη Θεσσαλονίκη και τις άλλες πόλεις. Τα λοιπά αστικά κέντρα καταγράφουν μια ανοδική τάση και παρουσιάζονται μεταναστεύσεις από αστικό σε άλλο αστικό κέντρο.

Οι εσωτερικοί μετανάστες φαίνεται ότι είναι περίπου των ίδιων χαρακτηριστικών με την προηγούμενη δεκαετία. Αλλά επειδή μειώνεται η ενδοαγροτική μετανάστευση, μειώνεται και ο αριθμός των εσωτερικών μεταναστών οι οποίοι ασχολούνται με την γεωργία – γεγονός άλλωστε που χαρακτηρίζει και την διάρθρωση της απασχόλησης στην Ελλάδα αυτή την εποχή.

• **ΠΕΡΙΟΔΟΣ 1971-1980**

Περίοδος διεθνούς οικονομικής ύφεσης. Στην Ελλάδα, επάνοδος της δημοκρατίας και ξεκίνημα για την πλήρη ένταξη στην ΕΟΚ. Μετακίνηση των βιομηχανικών μονάδων προς την περιφέρεια, επέκταση του χρηματοπιστωτικού συστήματος, μαζικός τουρισμός.

Μεταναστεύουν περισσότερα άτομα σε σύγκριση με την προηγούμενη περίοδο, αλλά, αλλάζουν οι μεταναστευτικές κατευθύνσεις. Δηλαδή, από πού προέρχονται και που εγκαθίστανται. Αρχίζει η μείωση των μεταναστεύσεων από τις αγροτικές περιοχές προς τα αστικά κέντρα, συμπεριλαμβανομένης και της Αθήνας (μείωση στο ένα τρίτο από την προηγούμενη περίοδο 1965-1971).

Διαφαίνεται μια νέα τάση για μεταναστεύσεις από τις αστικές προς τις αγροτικές περιοχές. Οι εσωτερικοί μετανάστες φαίνεται να αλλάζουν και σε κάποια χαρακτηριστικά τους (π.χ. ηλικία) τάση που παίρνει πιο δυνατή μορφή στην επόμενη δεκαετία.

• **ΠΕΡΙΟΔΟΣ 1981-1990**

Πλήρης ένταξη της χώρας στην Ε.Ε, γεγονός που διαμορφώνει ένα νέο οικονομικό και κοινωνικό πλαίσιο. Παράλληλα, αναλαμβάνει την διακυβέρνηση του κράτους σοσιαλιστική κυβέρνηση που ασκεί ιδιαίτερα ευρεία πολιτική πρόνοιας και παροχών. Ασκείται περιφερειακή πολιτική κινήτρων, γίνονται δημόσιες επενδύσεις σε τεχνικές υποδομές όσο και σε υποδομές υπηρεσιών (πχ υγεία, εκπαίδευση) σε όλη την περιφέρεια.

Παράλληλα η επέκταση των δικτύων μεταφορών επικοινωνιών, η βελτίωση του αγροτικού τομέα, η ανάπτυξη της βιομηχανίας\ βιοτεχνίας στην περιφέρεια, ο τουρισμός και η κατανάλωση υπηρεσιών αναγνυχής και πολιτισμού, η αποσύνδεση του τόπου εργασίας με τον τόπο κατοικίας, οδήγησαν στην άμβλυνση του φαινομένου της εσωτερικής μετανάστευσης στην Ελλάδα και συντέλεσαν στην σταθεροποίηση της κατανομής του πληθυσμού στον χώρο κατά την δεκαετία αυτή.

Οι αλλαγές στην χωρική κατανομή του πληθυσμού εντοπίζονται στην πτώση του ρυθμού πληθυσμιακής αύξησης των αστικών κέντρων και στην πληθυσμιακή ανάκαμψη ζωνών της υπαίθρου, κυρίως των παράκτιων και νησιώτικων περιοχών. Καταγράφεται έντονα η τάση για σταθεροποίηση της γεωγραφικής κατανομής του πληθυσμού.

Πίνακας 2.3.1

Βασικοί τύποι μεταναστευτικών ροών μεταξύ περιοχών διαφορετικού βαθμού αστικότητας

	1955-61		1965-71		1975-81		1985-91	
	(.000)	(%)**	(.000)	(%)**	(.000)	(%)**	(.000)	(%)**
Εσωτερικοί μετανάστες								
Σύνολο εσωτερικών	644,8		764,5		806,0		598,9	

μεταναστών*								
Μεταναστεύσαντες από: σε:								
Από αγροτικές περιοχές στην Αθήνα	107,8	16,7	117,0	15,3	98,5	12,2	30,4	5,1
Από αγροτικές περιοχές σε άλλες αστικές περιοχές	108,9	16,9	138,7	18,1	129,5	16,1	49,6	8,3
Από αγροτικές περιοχές σε άλλες αγροτικές περιοχές	108,6	16,8	83,8	11,0	55,9	6,9	26,6	4,4
Από άλλες αστικές περιοχές στην Αθήνα	72,8	11,3	93,8	12,3	79,4	9,9	47,2	7,9
Από αστικές περιοχές σε άλλες αστικές περιοχές	35,4	5,5	53,8	7,0	77,7	9,6	73,3	12,2
Από αστικές περιοχές σε αγροτικές περιοχές	23,4	3,6	26,5	3,5	45,6	5,7	73,9	12,3
Από την Αθήνα σε άλλες αστικές περιοχές	18,9	2,9	31,6	4,1	63,2	7,8	60,3	10,1
Από την Αθήνα σε ημιαστικές περιοχές	9,2	1,4	14,9	1,9	38,5	4,8	52,5	8,8
Από την Αθήνα σε αγροτικές περιοχές	12,3	1,9	18,7	2,4	48,6	6,0	64,4	10,8

\* Δεν συμπεριλαμβάνονται οι μετακινηθέντες εντός των Π.Σ Αθηνών και Θεσ/νίκης

\*\* Ως ποσοστό επί του συνόλου όλων των μεταναστεύσεων της περιόδου

Πηγή: ΕΣΥΕ, Αποτελέσματα απογραφών πληθυσμού 1961, 1971, 1981, 1991 (ίδιοι υπολογισμοί)

Τα σημαντικότερα χαρακτηριστικά που δίνει η απογραφή του 1991 για την εσωτερική μετανάστευση στην Ελλάδα και σε σύγκριση με τα στοιχεία της προηγούμενης απογραφής συνοψίζονται στα εξής :

- ❖ Μεγάλη μείωση του αριθμού των εσωτερικών μεταναστών.
- ❖ Ως προς τις μεταναστευτικές κατευθύνσεις παρατηρείται :
  - τάση για μικρή αύξηση του αριθμού αναχωρήσεων από τα αστικά κέντρα με κατεύθυνση τις αγροτικές και ημιαστικές περιοχές,
  - τάση για περαιτέρω μείωση τόσο της ενδοαγροτικής μετανάστευσης όσο και της αγροτικής εξόδου προς τα αστικά κέντρα γενικά,
  - μια τάση για περαιτέρω αύξηση των μεταναστεύσεων μεταξύ των πόλεων
- ❖ και τέλος μια σμίκρυνση του ρόλου της Αθήνας ως αποδέκτη μεταναστευτικών ροών.
- ❖ Ως προς τα μεταναστευτικά ισοζύγια λόγω της εσωτερικής μετανάστευσης: γενική είναι η τάση οι απώλειες ή οι αυξήσεις σε πληθυσμό όλων των περιοχών λόγω εσωτερικής μετανάστευσης να σταθεροποιούνται σε χαμηλά επίπεδα και να τρέπονται σε θετικά μεν για τις περιοχές της υπαίθρου και σε αρνητικά δε για τις αστικές περιοχές, αντιστρέφοντας τις τάσεις που επικράτησαν σε όλες τις προηγούμενες δεκαετίες.
- ❖ Ως προς τα ατομικά τους χαρακτηριστικά : οι εσωτερικοί μετανάστες αυτής της περιόδου ανήκουν, κυρίως, στην ηλικιακή κατηγορία των 20-34 ετών αλλά και αυτής των 45-64 ετών. Εξακολουθούν να είναι σε μεγαλύτερο ποσοστό οικονομικά ενεργοί από τις αντίστοιχες κατηγορίες των μη μεταναστών και στην συντριπτική τους πλειοψηφία απασχολούνται ως μισθωτοί στον κλάδο των υπηρεσιών. Υπάρχουν, όμως, έντονες διαφοροποιήσεις κατά φύλλο, ηλικία και βαθμό αστικότητας τόπου εγκατάστασης των μεταναστών οι οποίες σκιαγραφούν και το πιθανό είδος της μετανάστευσης που πραγματοποιήθηκε.



### ΚΕΦΑΛΑΙΟ 3 : ΠΗΓΕΣ ΔΗΜΟΓΡΑΦΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Οι βασικές πηγές άντλησης στατιστικών δεδομένων για τις ανάγκες της πληθυσμιακής ανάλυσης είναι οι εξής :

- ο Οι απογραφές του πληθυσμού
- ο Οι επίσημες καταγραφές δημογραφικών συμβάντων
- ο Τα μητρώα πληθυσμού
- ο Οι ειδικές δειγματοληπτικές έρευνες

Η *απογραφή* του πληθυσμού αποτελεί τη βασικότερη πηγή συλλογής εγκάρσιων δεδομένων που αφορούν το μέγεθος και την σύνθεση του πληθυσμού. Ο όρος αναφέρεται στη διαδικασία συλλογής στατιστικών στοιχείων που περιλαμβάνουν το σύνολο των μελών ενός πληθυσμού μιας συγκεκριμένης γεωγραφικής περιοχής κάποια δεδομένη χρονική στιγμή.

Πιο συγκεκριμένα, η απογραφή είναι η διαδικασία συλλογής, επεξεργασίας, αξιολόγησης, ανάλυσης και δημοσίευσης στοιχείων που αναφέρονται σε δημογραφικά, κοινωνικά και οικονομικά χαρακτηριστικά ενός πληθυσμιακού συνόλου μιας διακριτής γεωγραφικής περιοχής κάποια χρονική στιγμή. Τα στοιχεία που συλλέγονται μέσα στα πλαίσια της απογραφής δίνουν την πλήρη εικόνα του μεγέθους του πληθυσμού αλλά και της σύνθεσης του ως προς τα διάφορα δημογραφικά και κοινωνικά χαρακτηριστικά του. Τα στοιχεία που παρέχουν ακόμα τη δυνατότητα της διαχρονικής μελέτης του πληθυσμού και των μεταβολών του αλλά και τη δυνατότητα ειδικότερων προσεγγίσεων του αφορά υποσύνολα του γενικού πληθυσμού.

Οι απογραφές του Ελληνικού πληθυσμού διεξάγονται από την Εθνική Στατιστική Υπηρεσία (Ε.Σ.Υ.Ε) πάντοτε ημέρα Κυριακή (μη εργάσιμη) ώστε να διασφαλίζεται η μέγιστη δυνατή παρουσία των ατόμων στις κατοικίες τους. Επίσης, διεξάγεται την άνοιξη μέσα στο μήνα Μάρτη έτσι ώστε οι καιρικές συνθήκες να διευκολύνουν την πρόσβαση σε απομακρυσμένες γεωγραφικές περιοχές. Τέλος, διεξάγονται ανά δεκαετία. Οι τελευταίες απογραφές που διεξήχθησαν ήταν τα έτη 1951, 1961, 1971, 1981, 1991, 2001.

Η απογραφή του πληθυσμού παρέχει μία « φωτογραφική» αποτύπωση του μεγέθους και της σύνθεσης του πληθυσμού κατά το χρονικό σημείο της διεξαγωγής της.

Στοιχεία που αποτυπώνουν την διαχρονική μεταβολή του μεγέθους και της σύνθεσης του πληθυσμού σαν συνέπεια των φυσικών παραγόντων (γεννήσεις ή θάνατοι) ή της γαμηλιότητας και της συχνότητας των διαζυγίων παρέχονται κύρια από τα συστήματα επίσημων καταγραφών αυτών των γεγονότων (*ληξιαρχικές καταγραφές*).

Οι καταγραφές αυτές συγκεντρώνουν στοιχεία χρόνου και τόπου επέλευσης του συμβάντος, καθώς και τα δημογραφικά στοιχεία των ατόμων που αφορά το συμβάν

όπως όνομα, φύλο, έτος γέννησης κλπ. Συστήματα επισήμων καταγραφών των δημογραφικών συμβάντων λειτουργούν σε πολλές χώρες, κύρια στις ανεπτυγμένες αλλά και σε μέρος των υπό ανάπτυξη χωρών. Στην Ελλάδα λειτουργεί το ληξιαρχικό σύστημα καταγραφών όπου οι δηλώσεις των δημογραφικών συμβάντων είναι υποχρεωτικές. Έτσι υπάρχει σχεδόν πλήρης κάλυψη του αριθμού συμβάντων με εξαίρεση ίσως την προγεννητική θνησιμότητα όπου κάποιοι θάνατοι νεογνών που συμβαίνουν σε μικρό χρονικό διάστημα μετά τη γέννηση δε δηλώνονται. Φαινόμενο που παρατηρείται κυρίως σε επαρχιακές περιοχές.

Οι χώρες που διαθέτουν επίσημο σύστημα καταγραφών των δημογραφικών συμβάντων διατηρούν, επίσης, και στοιχεία των πολιτών τους σε ατομικό και οικογενειακό επίπεδο σε **μητρώα πληθυσμού**. Στα μητρώα αυτά υπάρχει καταχώρηση για το κάθε μέλος του πληθυσμού στην οποία αναγράφεται το πλήρες δημογραφικό ιστορικό του κάθε ατόμου. Τα στοιχεία αυτά έχουν συγκεντρωθεί για την εξυπηρέτηση διοικητικών αναγκών, παρέχουν όμως και πληροφόρηση πολύτιμη για τις ανάγκες της πληθυσμιακής ανάλυσης. Τα μητρώα αυτά αποτελούν σε κάποιο βαθμό και πηγές πληροφόρησης της μεταναστευτικής κίνησης του πληθυσμού. Τα αποτελέσματα των απογραφών του πληθυσμού παρέχουν συνήθως μετρήσεις του βαθμού κάλυψης των μητρώων. Στην Ελλάδα λειτουργούν διάφορα επίσημα μητρώα για τις ανάγκες του πολυδιάστατου μηχανισμού της δημόσιας διοίκησης τα οποία συγκεντρώνονται σε διάφορους δημόσιους φορείς για την κάλυψη των ειδικών αναγκών της. Εκτός από τα επίσημα στοιχεία απογραφών και επισήμων καταγραφών, συμπληρωματική πηγή δημογραφικών δεδομένων αποτελούν και οι **ειδικές δειγματοληπτικές έρευνες** οι οποίες διεξάγονται κατά καιρούς για διάφορους σκοπούς.

## ΚΕΦΑΛΑΙΟ 4 : ΑΝΑΔΡΟΜΗ ΣΕ ΥΠΑΡΧΟΥΣΑ ΕΡΕΥΝΑ

Σκόπιμο θεωρείται να γίνει μια ανασκόπηση σε τυχόν προηγούμενες έρευνες που έχουν διεξαχθεί και αφορούν στο φαινόμενο της εσωτερικής μετανάστευσης.

Παρ' όλο που η μελέτη του ρόλου των περιβαλλοντικών παραγόντων στην εσωτερική μετανάστευση της χώρας είναι περιορισμένη, έχει πραγματοποιηθεί μια πολύ σημαντική προσπάθεια από τους Καλογήρου Σ. και Κατσαφάδο Π. ώστε να μελετηθεί ο παράγοντας αυτός. («Ο ρόλος του κλίματος στην μεταναστευτική ροή»)

Το βασικό ερώτημα που τίθεται στο συγκεκριμένο άρθρο είναι “ η εξακρίβωση της ύπαρξης σημαντικής επίδρασης του κλίματος στην επιλογή μεταναστευτικού προορισμού και του κατά πόσο η επίδραση αυτή είναι χωρικά σταθερή στον ελλαδικό χώρο”.

Για τα αποτελέσματα χρησιμοποιήθηκαν τα στοιχεία που προέκυψαν από την απογραφή του 2001 που διενήργησε η Ελληνική Στατιστική Υπηρεσία και οι τεχνικές των Γεωγραφικών Συστημάτων Πληροφοριών.

Τα αποτελέσματα απέδειξαν ότι υπάρχει σημαντική επίδραση του τοπικού κλίματος στην επιλογή του προορισμού μετανάστευσης. Αποδείχθηκε ότι η επίδραση αυτή είναι διαφορετική για κάθε περιοχή της Ελλάδας. (Καλογήρου Σ. και Κατσαφάδος Π.)

Στην πορεία της επικείμενης εργασίας, θα χρησιμοποιηθούν άλλοι παράγοντες και άλλες τεχνικές πληροφοριών, ώστε να ελεγχθεί κατά πόσο η εσωτερική μετανάστευση επηρεάζεται και από άλλης φύσης παράγοντες ( κοινωνικοί, οικονομικοί, δημογραφικοί).

Πανεπιστήμιο Πειραιώς

## ΚΕΦΑΛΑΙΟ 5 : ΠΕΡΙΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ

Στη συνέχεια θα παρουσιαστούν οι μεταβλητές που θα χρησιμοποιηθούν , ώστε να αναλυθούν και να διεξαχθούν τα συμπεράσματα μας.

Οι παρακάτω δείκτες βασίζονται σε δεδομένα που καταγράφηκαν κατά την απογραφή που διεξήχθη το 2001 στην Ελλάδα. Τα δεδομένα παρέχονται ανά επίπεδο νομού.

- density : Αποτελεί τον πληθυσμό ανά τετραγωνικό χιλιόμετρο (  $km^2$ ) της χώρας.
- urban : Είναι η αναλογία του αστικού πληθυσμού σε σχέση με τον συνολικό πληθυσμό της χώρας.
- masculin : Αναφέρεται στο ποσοστό των ανδρών.
- p1564 : Δείχνει το ποσοστό του πληθυσμού που είναι ηλικίας 15-64 ετών.
- eduleveltotal : Δείχνει το μέσο αριθμό ετών των συνολικών ολοκληρωμένων εκπαιδευτικών δραστηριοτήτων του πληθυσμού.
- edulevelmale : Δείχνει το μέσο αριθμό ετών των ολοκληρωμένων εκπαιδευτικών δραστηριοτήτων του ανδρικού πληθυσμού.
- edulevelfemale : Δείχνει το μέσο αριθμό ετών των ολοκληρωμένων εκπαιδευτικών δραστηριοτήτων του γυναικείου πληθυσμού.
- anergiat : Είναι το συνολικό ποσοστό των ανέργων.
- anergiam : Είναι το συνολικό ποσοστό των ανέργων ανδρών.
- anergiaf : Είναι το συνολικό ποσοστό των ανέργων γυναικών.
- tomeas1 : Είναι η συμμετοχή (%) των απασχολούμενων στον πρωτογενή τομέα παραγωγής.
- tomeas2 : Είναι η συμμετοχή (%) των απασχολούμενων στον δευτερογενή τομέα παραγωγής.
- tomeas3 : Είναι η συμμετοχή (%) των απασχολούμενων στον τριτογενή τομέα παραγωγής.
- participation : Είναι η συμμετοχή (%) του οικονομικά ενεργού πληθυσμού στο εργατικό δυναμικό.
- in migration: Δείκτης εσωτερικής μετανάστευσης/εγκατάστασης. Υπολογίστηκε ως το πηλίκο των εγκατασταθέντων από άλλο δήμο προς το σύνολο των ατόμων, επί 100.

Τα παρακάτω δεδομένα αποτελούν δείκτες που υπολογίστηκαν από στοιχεία των ληξιαρχικών καταγραφών αλλά και από στοιχεία της απογραφής που διεξήχθη στην Ελλάδα το 2001.

- cbr : Αδρός δείκτης Γεννήσεων
- gfr : Γενικός δείκτης Γονιμότητας
- cwr : Λόγος παιδιού-γυναικών

Επίσης, χρησιμοποιήθηκαν οι μεταβλητές :

- incomecapita : Το κατά κεφαλήν εισόδημα της χώρας.  
(Βάση των στατιστικών στοιχείων της Τράπεζας της Ελλάδος.)
- long term unemployed : Μακροχρόνια ανεργία.  
(Βάση της Έρευνας Εργατικού Δυναμικού).

Πανεπιστήμιο Πειραιώς

## ΚΕΦΑΛΑΙΟ 6 : ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Η **Στατιστική** είναι ένας κλάδος που βοηθάει στην μελέτη και κατανόηση φαινομένων ή ιδιοτήτων πολυπληθών ομάδων. Η εφαρμογή της εκτείνεται σε πολλούς κλάδους της ανθρώπινης δραστηριότητας πχ πολιτική, οικονομία κ.λ.π. Η **Περιγραφική Στατιστική** είναι ο κλάδος της στατιστικής που ασχολείται με την οργάνωση, συλλογή και παρουσίαση ενός συνόλου δεδομένων.

Για να μπορέσουμε να παρουσιάσουμε τα δεδομένα καθαρά, σύντομα και με σαφήνεια θα χρησιμοποιήσουμε στατιστικούς πίνακες και γραφικές παραστάσεις.

Για να προσδιοριστούν τα σημαντικότερα χαρακτηριστικά των δεδομένων (όπως, για παράδειγμα, το εύρος τους, οι ακραίες τιμές τους κ.λ.π.) θα ορίσουμε ποσοτικά μεγέθη που περιγράφουν περιληπτικά τα βασικά χαρακτηριστικά της τ.μ  $X$ . Αυτά τα μεγέθη καλούνται περιγραφικά μέτρα.

Παρακάτω θα παρουσιαστούν δύο τύποι περιγραφικών μέτρων :

- Τα μέτρα θέσης, τα οποία προσδιορίζουν χαρακτηριστικές θέσεις μέσα στο εύρος των δεδομένων , και
- Τα μέτρα μεταβλητότητας, τα οποία δίνουν την διασκόρπιση και την μεταβλητότητα των δεδομένων.

### 6.1 ΜΕΤΡΑ ΘΕΣΗΣ

Εννοούνται τα μέτρα κεντρικής τάσης που προσδιορίζουν ένα κεντρικό σημείο γύρω από το οποίο τείνουν να συγκεντρώνονται τα δεδομένα. Τα κυριότερα μέτρα είναι...

#### 6.1.1 Δειγματική μέση τιμή (ή αριθμητικός μέσος ή μέσος όρος)

Αποτελεί το πιο γνωστό και εύχρηστο μέτρο του κέντρου των δεδομένων. Έστω  $x_1, x_2, \dots, x_n$  οι τιμές των παρατηρήσεων του δείγματος για μια τ.μ  $X$ . Η δειγματική μέση τιμή συμβολίζεται με  $\bar{x}$  και ορίζεται ως εξής :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i \quad (6.1.1.1)$$

Η μέση τιμή αποτελεί το κέντρο ισορροπίας των δεδομένων. Για τον υπολογισμό της χρησιμοποιούνται όλες οι τιμές του δείγματος , γι' αυτό και επηρεάζεται από μακρινές τιμές.

#### 6.1.2 Δειγματική διάμεσος

Ορίζεται ως η κεντρική τιμή όταν διατάξουμε τα δεδομένα σε αύξουσα σειρά.

Αν ο αριθμός  $n$  των δεδομένων είναι περιττός τότε η διάμεσος είναι η τιμή στη θέση  $\frac{(n+1)}{2}$ .

Αν ο αριθμός  $n$  των δεδομένων είναι άρτιος τότε η διάμεσος είναι το ημίαθροισμα των τιμών στις θέσεις  $\frac{n}{2}$  και  $\frac{n}{2} + 1$ .

Συμβολίζοντας την δειγματική διάμεσο με  $\delta$ , ορίζεται ως εξής :

$$\delta = \begin{cases} x_{\frac{(n+1)}{2}}, & n=2k+1 & (6.1.2.1.) \\ \frac{x_{n/2} + x_{\frac{n}{2}+1}}{2}, & n = 2k & (6.1.2.2) \end{cases}$$

Η διάμεσος δεν επηρεάζεται από ακραίες τιμές.

### 6.1.3 Δειγματική επικρατούσα τιμή

Χρησιμοποιείται για να δηλώσει την κεντρική τάση των δεδομένων και ορίζεται ως η τιμή που εμφανίζεται με τη μεγαλύτερη συχνότητα.

Αν υπάρχουν περισσότερες από μια τέτοιες τιμές, τότε όλες θεωρούνται επικρατούσες.

Όταν το δείγμα δεν αποτελείται από διακεκριμένες επαναλαμβανόμενες τιμές, τότε η επικρατούσα τιμή δεν έχει νόημα.

## 6.2 ΜΕΤΡΑ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ (ΔΙΑΣΠΟΡΑΣ)

Όταν τα δεδομένα είναι συγκεντρωμένα γύρω από μια κεντρική τιμή, δηλαδή η διασπορά των δεδομένων είναι μικρή, τότε η κεντρική τιμή αντιπροσωπεύει ικανοποιητικά τα δεδομένα.

Από την άλλη, όταν τα δεδομένα, είναι πολύ διασκορπισμένα, τα μέτρα κεντρικής τιμής δε δίνουν καλή περιγραφή δεδομένων.

Συνεπώς, εκτός από την κεντρική τάση μας ενδιαφέρει και η μεταβλητότητα ή διασπορά των παρατηρήσεων.

Τα κυριότερα μέτρα μεταβλητότητας είναι :

### 6.2.1 Δειγματικό εύρος

Αποτελεί τη διαφορά της ελάχιστης από τη μέγιστη τιμή του δείγματος.

Συμβολίζεται με  $R$  και ορίζεται ως εξής :

$$R = x_{max} - x_{min} \quad (6.2.1.1)$$



Είναι ένα εύκολο αλλά καθόλου αξιόπιστο μέτρο. Εξαρτάται μόνο από τις δυο ακραίες παρατηρήσεις  $x_{max}$  και  $x_{min}$  χωρίς να υπολογίζει τις υπόλοιπες παρατηρήσεις.

### 6.2.2 Δειγματική διακύμανση ή διασπορά

Μετράει τη μεταβλητότητα των παρατηρήσεων γύρω από τη μέση τιμή.

Αν ορίσουμε την απόκλιση μιας παρατήρησης  $x_i$  από τη μέση τιμή ως  $x_i - \bar{x}$ , το άθροισμα των αποκλίσεων είναι ο

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = n\bar{x} - n\bar{x} = 0 \quad (6.2.2.1)$$

Η δειγματική μέση τιμή  $\bar{x}$  έχει οριστεί έτσι ώστε οι θετικές αποκλίσεις για τιμές μεγαλύτερες του  $\bar{x}$  να είναι αθροιστικά ίδιες με τις αρνητικές αποκλίσεις για τιμές μικρότερες του  $\bar{x}$ .

Για να μετρήσουμε την μεταβλητότητα των παρατηρήσεων γύρω από τη μέση τιμή διαλέγουμε να αθροίσουμε όχι τις ίδιες τις αποκλίσεις αλλά τα τετράγωνα των αποκλίσεων. Επίσης, πρέπει να διαιρέσουμε με το πλήθος των παρατηρήσεων έτσι ώστε το μέτρο της μέσης απόκλισης να μην εξαρτάται από το πλήθος των παρατηρήσεων

Η διασπορά συμβολίζεται με  $s^2$  και ορίζεται ως εξής :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (6.2.2.2)$$

ή ισοδύναμα 
$$s^2 = \frac{1}{n-1} (\sum_{i=1}^n x_i^2 - n\bar{x}) \quad (6.2.2.3)$$

### 6.2.3 Δειγματική τυπική απόκλιση

Επειδή αντιμετωπίζονται σημαντικές δυσκολίες ερμηνείας της δειγματικής διασποράς ως φυσικό μέγεθος (λόγω του ότι προκύπτει από τα τετράγωνα των παρατηρήσεων) ορίζουμε τη δειγματική τυπική απόκλιση  $s$ , που είναι η θετική ρίζα της διασποράς  $s^2$ . Μετριέται με τη μονάδα μέτρησης της τ.μ  $X$  και εκφράζει την τυπική απόκλιση των δεδομένων από τη δειγματική μέση τιμή, δηλαδή μέχρι πόσο περιμένουμε μια τυπική τιμή της  $X$  να απέχει από τη μέση τιμή.

Πανεπιστήμιο Πειραιώς

## ΚΕΦΑΛΑΙΟ 7 : ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Σε διάφορα προβλήματα της Στατιστικής το ενδιαφέρον εστιάζεται στην ταυτόχρονη μελέτη δύο ή περισσότερων μεταβλητών, ώστε να προσδιοριστεί με ποιο τρόπο αυτές οι μεταβλητές σχετίζονται μεταξύ τους.

Για παράδειγμα, το ύψος των αποδοχών των υπαλλήλων μιας εταιρίας εξαρτάται από τα χρόνια υπηρεσίας στην εταιρία.

Ο κλάδος της Στατιστικής που εξετάζει τη διαρθρωτική σχέση μεταξύ δύο ή περισσότερων μεταβλητών με απώτερο σκοπό την πρόβλεψη μιας εξ αυτών μέσω των άλλων καλείται ανάλυση παλινδρόμησης (*regression analysis*).

Οι κατηγορίες των μεταβλητών που εμπλέκονται σε τέτοιου είδους αναλύσεις είναι οι εξής δύο :

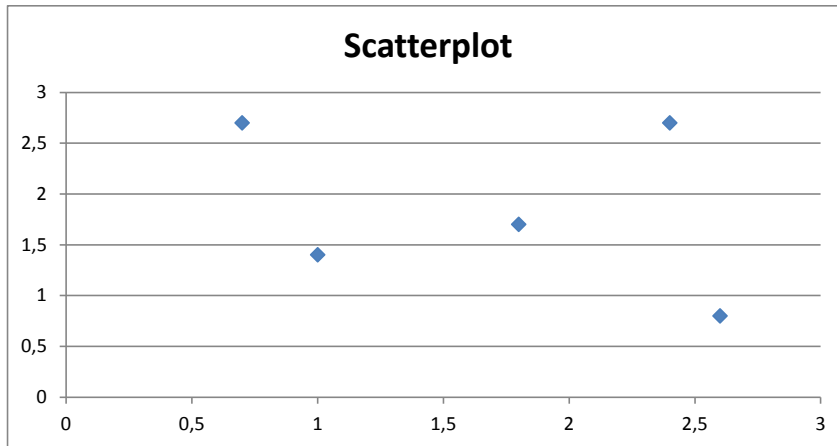
(α) η ανεξάρτητη μεταβλητή , η οποία είναι αυτή που μπορεί να ελεγχθεί, δηλαδή μπορούν οι τιμές της να καθοριστούν , και

(β) η εξαρτημένη μεταβλητή , στην οποία αντιστακτάται το αποτέλεσμα των μεταβολών των ανεξαρτήτων μεταβλητών.

Η παλινδρόμηση στην οποία υπάρχει μόνο μια ανεξάρτητη μεταβλητή καλείται απλή παλινδρόμηση. Ενώ, αν υπάρχουν περισσότερες ανεξάρτητες μεταβλητές καλείται πολλαπλή παλινδρόμηση.

Για την εύρεση του κατάλληλου μοντέλου για την περιγραφή της σχέσης μεταξύ των μεταβλητών που μας ενδιαφέρουν , ξεκινάμε κατασκευάζοντας το διάγραμμα διασποράς (*scatterplot*) στο επίπεδο των παρατηρήσεων που μας διαθέτουμε. Σ' ένα τέτοιο διάγραμμα οι τιμές της μεταβλητής X τοποθετούνται στον οριζόντιο άξονα και της μεταβλητής Y στον κάθετο άξονα.

Ένα τέτοιο διάγραμμα παρουσιάζεται στη συνέχεια.



ΔΙΑΓΡΑΜΜΑ 7.1  
ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ

## 7.1 ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

### 7.1.1 Το μοντέλο

Η απλούστερη περίπτωση παλινδρόμησης είναι η απλή γραμμική παλινδρόμηση. Ο σκοπός της είναι να περιγράψει τη σχέση μεταξύ των  $X$  και  $Y$  μ' ένα υπόδειγμα που έχει την εξής μορφή :

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (7.1.1.1)$$

Όπου ,

$Y_i$  → η τιμή της εξαρτημένης μεταβλητής

$X_i$  → η τιμή της ανεξάρτητης μεταβλητής

$\beta_0$  → το σημείο τομής του άξονα της  $Y$  από τη γραμμή παλινδρόμησης

$\beta_1$  → η κλίση της γραμμής παλινδρόμησης

$\varepsilon_i$  → το σφάλμα ή το κατάλοιπο, δηλαδή η διαφορά μεταξύ της πραγματικής τιμής της  $Y$  και της τιμής της πρόβλεψης που προκύπτει από το υπόδειγμα. Τα κατάλοιπα ακολουθούν κανονική κατανομή  $\varepsilon \sim N(0, \sigma^2)$  και είναι ανεξάρτητα μεταξύ τους.

Από την σχέση 7.1.1.1 και από το γεγονός ότι  $\varepsilon \sim N(0, \sigma^2)$  προκύπτει το εξής συμπέρασμα :

$$Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$$

με

- $E(Y_i) = E(\beta_0 + \beta_1 X_i + \varepsilon_i) = \beta_0 + \beta_1 X_i + E(\varepsilon_i) = \beta_0 + \beta_1 X_i$
- $Var(Y_i) = Var(\beta_0 + \beta_1 X_i + \varepsilon_i) = Var(\varepsilon_i) = \sigma^2$

Οι  $\tau.μ$   $Y_i$  θα είναι ανεξάρτητες εφόσον και τα σφάλματα  $\varepsilon_i$  ανεξάρτητα.

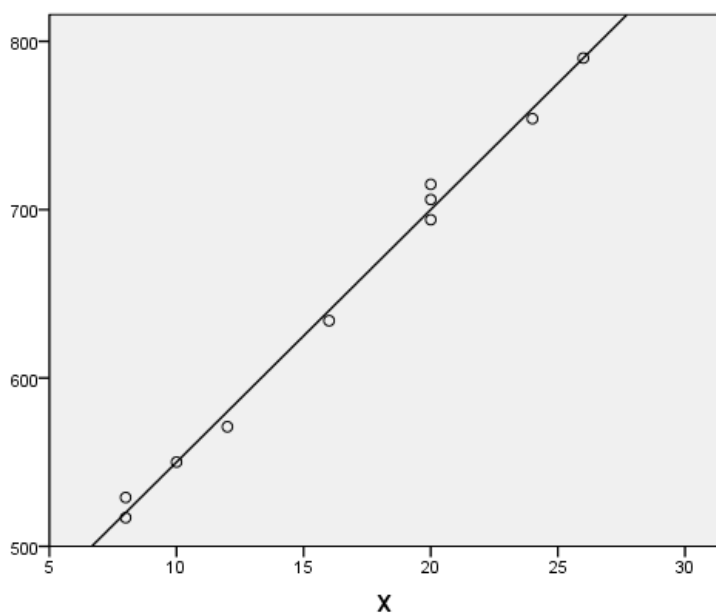
### 7.1.2 Η εκτίμηση των παραμέτρων

Μια γραμμή στο επίπεδο καθορίζεται από μια σχέση μεταξύ των δυο συντεταγμένων κάθε σημείου  $(x,y)$  του επιπέδου. Μια ευθεία απαρτίζεται από τα σημεία  $(x,y)$  του επιπέδου για τα οποία ισχύει η σχέση

$$y = \beta_0 + \beta_1 x \quad (7.1.1)$$

Για κάποια  $\beta_0, \beta_1$  τα οποία καλούνται τεταγμένη (*intercept*) και κλίση (*slope*) της ευθείας αντίστοιχα. Μια ιδιότητα της γραμμικής αυτής σχέσης μεταξύ των  $x,y$  είναι η σταθερή κλίση. Δηλαδή, για μια σταθερή αύξηση  $\Delta x$  του  $x$ , παρατηρείται σταθερή μεταβολή του  $y$  ( $\beta_1 \Delta x$ ). Αν η κλίση  $\beta_1$  είναι αρνητική, τότε θα έχουμε μείωση του  $y$ . Κάτι τέτοιο φυσικά δεν ισχύει όταν έχουμε μία μη γραμμική σχέση μεταξύ των  $x,y$  (π.χ εκθετική, λογαριθμική κ.λ.π)

Θεωρούμε το παρακάτω διάγραμμα διασποράς κάποιων δισδιάστατων παρατηρήσεων  $(x_1, y_1), \dots, (x_n, y_n)$ .



ΔΙΑΓΡΑΜΜΑ 7.1.2.1

Διάγραμμα διασποράς δισδιάστατων παρατηρήσεων

Φαίνεται ότι υπάρχει μια τάση αύξησης των  $y_i$  όταν αυξάνονται τα  $x_i$ . Η σχέση, όμως, που έχουν δεν φαίνεται να καθορίζεται από μια τέλεια ευθεία.

Για να προσδιοριστεί η ευθεία που «ταιριάζει» περισσότερο στα σημεία, θα χρησιμοποιηθεί η παρακάτω μέθοδος...

- **ΜΕΘΟΔΟΣ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ (Ordinary Least Square)**

Έστω ότι έχουμε τα παρακάτω  $n$  σημεία  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  στο επίπεδο.

Η προσεγγιστική ευθεία γραμμικής παλινδρόμησης για τα δεδομένα, είναι αδύνατον να περάσει και από τα  $n$  σημεία. Στις περισσότερες περιπτώσεις, σε κάθε σημείο  $X_i$  η ευθεία θα αντιστοιχίσει ένα εκτιμώμενο σημείο  $\hat{Y}_i = \beta_0 + \beta_1 x_i$  και είναι πολύ πιθανό να ισχύει  $\hat{Y}_i \neq Y_i$ . Όπως φαίνεται και από το παραπάνω σχήμα, ισχύει :

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (7.1.2.1)$$

Για κάθε επιλογή  $\beta_0, \beta_1$  λαμβάνουμε διαφορετική ευθεία και ένα νέο σύνολο από «κατάλοιπα» ή «υπόλοιπα» (*residuals*)  $\varepsilon_i$ .

Σύμφωνα με τη μέθοδο ελαχίστων τετραγώνων, η ευθεία που προσαρμόζεται καλύτερα στα δεδομένα είναι αυτή που ελαχιστοποιεί το άθροισμα των τετραγώνων των καταλοίπων :

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - (\beta_0 + \beta_1 x_i))^2 \quad (7.1.2.2)$$

Η εύρεση των  $\hat{\beta}_0, \hat{\beta}_1$  που ελαχιστοποιούν την παραπάνω παράσταση γίνεται με μεθόδους ελαχιστοποίησης συναρτήσεων πολλών μεταβλητών. Θέτοντας τις μερικές παραγώγους ως προς  $\beta_0$  και  $\beta_1$  ίσες με 0 και λύνοντας βρίσκουμε ότι :

- $$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{S_{XY}}{S_{XX}} \quad (7.1.2.3)$$

- $$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (7.1.2.4)$$

Η ευθεία  $y = \hat{\beta}_0 + \hat{\beta}_1 x$  καλείται ευθεία ελαχίστων τετραγώνων και οι διαφορές των προσαρμοσμένων τιμών  $\hat{Y}_i$  από τις παρατηρούμενες τιμές  $Y_i$  καλούνται κατάλοιπα και συμβολίζονται με  $\varepsilon_i = Y_i - \hat{Y}_i$ .

### 7.1.3 Η μεταβλητότητα του μοντέλου

#### *i. ΑΘΡΟΙΣΜΑΤΑ ΤΕΤΡΑΓΩΝΩΝ*

Η ολική μεταβλητότητα των παρατηρήσεων  $Y_i$  δίνεται από τον παρακάτω τύπο:

$$SSTO = \sum_{i=1}^n (Y_i - \bar{Y})^2 = SSR + SSE \quad (7.1.3.1)$$

Όπου :

$$\cdot SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \quad (7.1.3.2)$$

Είναι η μεταβλητότητα που ερμηνεύεται από την ευθεία παλινδρόμησης.

$$\cdot SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (7.1.3.3)$$

Είναι η μεταβλητότητα που παραμένει ανεξηγήτη και μας δείχνει την απόκλιση των πραγματικών δεδομένων από τα εκτιμώμενα.

#### ii. ΣΥΝΤΕΛΕΣΤΗΣ ΠΡΟΣΔΙΟΡΙΣΜΟΥ

Ο συντελεστής προσδιορισμού ορίζεται ως εξής :

$$R^2 = \frac{SSR}{SSTO} = \frac{SSTO - SSE}{SSTO} \quad (7.1.3.4)$$

και παίρνει τιμές  $0 \leq R^2 \leq 1$

Πιο συγκεκριμένα,

- ο Όταν  $SSR \approx SSTO$  τότε  $R^2 \approx 1$  και συνεπώς  $Y_i \approx \hat{Y}_i$  (Το καλύτερο μοντέλο)
- ο Όταν  $SSR \approx 0$  τότε  $R^2 \approx 0$  (Το χειρότερο μοντέλο)

- Για να μειωθεί η ολική μεταβλητότητα που δεν έχει ερμηνευθεί , ώστε να είναι πιο χρήσιμη η ευθεία παλινδρόμησης , θα πρέπει ο συντελεστής προσδιορισμού  $R^2$  να είναι όσο το δυνατόν μεγαλύτερος (πιο κοντά στην μονάδα).

#### iii. ΜΕΣΑ ΑΘΡΟΙΣΜΑΤΑ ΤΕΤΡΑΓΩΝΩΝ

Τα μέσα αθροίσματα τετραγώνων υπολογίζονται διαιρώντας το κάθε άθροισμα τετραγώνων με τους αντίστοιχους βαθμούς ελευθερίας.

Πιο συγκεκριμένα , ορίζονται ως εξής:

$$\cdot MSTO = \frac{SSTO}{n-1} \quad (7.1.3.5)$$

$$\cdot MSR = \frac{SSR}{1} = SSR \quad (7.1.3.6)$$

$$\cdot MSE = \frac{SSE}{n-2} \quad (7.1.3.7)$$

#### 7.1.4 Έλεγχος υποθέσεων για τις παραμέτρους του μοντέλου

Ο βασικός σκοπός της παλινδρόμησης είναι η μελέτη της ύπαρξης μίας γραμμικής σχέσης μεταξύ των μεταβλητών  $X$  και  $Y$ . Γι αυτό το λόγο θα πρέπει να γίνουν κάποιοι «έλεγχοι» για την σχέση αυτή.

Ο στατιστικός έλεγχος, λοιπόν, της υπόθεσης ότι υπάρχει γραμμική σχέση μεταξύ  $X$  και  $Y$  είναι :

$H_0: \beta_1=0$  (Δεν υπάρχει γραμμική σχέση μεταξύ  $X$  και  $Y$ )

$H_1: \beta_1 \neq 0$  (Υπάρχει γραμμική σχέση μεταξύ  $X$  και  $Y$ )

Ο έλεγχος υποθέσεων γίνεται με βάση την κλίση  $\beta_1$ , εφόσον αν  $\beta_1=0$  τότε οποιαδήποτε μεταβολή του  $X$  δεν επηρεάζει το  $Y$ . Πράγματι, αν  $\beta_1=0$ , τότε το γραμμικό μοντέλο  $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$  παίρνει την μορφή  $Y_i = \beta_0 + \varepsilon_i$ , όπου  $\varepsilon_i \sim N(0, \sigma^2)$ , άρα  $Y_i \sim N(\beta_0, \sigma^2)$  που σημαίνει ότι οι κατανομές των  $Y_i$ ,  $i=1,2,\dots,n$  είναι όλες ίδιες ανεξάρτητα από το επίπεδο του  $X$  και συνεπώς το  $X$  δε δίνει καμία πληροφορία για την πρόβλεψη του  $Y$ .

##### iv. Έλεγχος t-test

Για τον έλεγχο της υπόθεσης :

$H_0: \beta_1=0$

$H_1: \beta_1 \neq 0$

η στατιστική συνάρτηση που χρησιμοποιείται είναι η :

$$T = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)} \quad (7.4.1.1)$$

Ενώ ο κανόνας απόρριψης της  $H_0$  σε επίπεδο σημαντικότητας  $(1-\alpha)\%$  είναι ο εξής :

**Απορρίπτεται η  $H_0$  όταν  $|T| > t_{n-2}(\frac{\alpha}{2})$**

##### v. Έλεγχος F-test

Για τον έλεγχο της υπόθεσης :

$H_0: \beta_1=0$

$H_1: \beta_1 \neq 0$

η στατιστική συνάρτηση που χρησιμοποιείται είναι η :

$$F = \frac{MSR}{MSE} = \frac{SSR/1}{SSE/(n-2)} \quad (7.4.1.2)$$



Ενώ ο κανόνας απόρριψης της  $H_0$  σε επίπεδο σημαντικότητας  $(1-\alpha)\%$  είναι ο εξής :

**Απορρίπτεται η  $H_0$  όταν  $F > F_{1,n-2(\alpha)}$**

## 7.2 ΠΟΛΛΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Σε πολλές περιπτώσεις μπορεί η μεταβλητή  $Y$  να επηρεάζεται από περισσότερες από μια ερμηνευτικές μεταβλητές  $X_1, X_2, \dots, X_{p-1}$ . Και πάλι, στην περίπτωση αυτή, μπορούμε να χρησιμοποιήσουμε ένα γραμμικό μοντέλο προκειμένου να εξετάσουμε τη σχέση μεταξύ των μεταβλητών  $X_1, X_2, \dots, X_{p-1}$  και  $Y$ . Το μοντέλο που θα χρησιμοποιηθεί καλείται πολλαπλό γραμμικό μοντέλο.

### 7.2.1 Το μοντέλο

Το στατιστικό πολλαπλό γραμμικό μοντέλο έχει την εξής μορφή :

$$Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad (7.2.1.1)$$

Όπου  $i=1,2,\dots,n$

$Y_i$  : τυχαίες μεταβλητές

$\beta_0, \beta_1, \dots, \beta_{p-1}$ : άγνωστες παράμετροι

$X_{ij}$  : μη τυχαίες μεταβλητές ( η  $X_{i0}$  είναι μία πλασματική μεταβλητή που παίρνει συνεχώς την τιμή 1)

$\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  : τυχαία σφάλματα (τυχαίες μεταβλητές) τα οποία κατανομούνται σύμφωνα με την κανονική κατανομή  $N(0, \sigma^2)$  και είναι ασυσχέτιστα και ανεξάρτητα μεταξύ τους.

Λόγω των παραπάνω προκύπτει ότι :

$$Y_i \sim N(\beta_0 + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{i,p-1}, \sigma^2) \quad (7.2.1.2)$$

με παραμέτρους

$$\cdot E(Y_i) = \beta_0 + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{i,p-1} \quad (7.2.1.3)$$

$$\cdot Var(Y_i) = \sigma^2 \quad (7.2.1.4)$$

### 7.2.2 Η εκτίμηση των παραμέτρων

Το πολλαπλό γραμμικό μοντέλο μπορεί να παρουσιαστεί και με την εξής μορφή πινάκων:

$$Y = X\beta + \varepsilon$$

Όπου  $Y = [Y_1, Y_2, \dots, Y_n]$  ,  $\beta = [\beta_0, \beta_1, \dots, \beta_{p-1}]$  ,  $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]$  και

$$X = \begin{bmatrix} X_{10} & \dots & X_{1,p-1} \\ \dots & \dots & \dots \\ X_{n0} & \dots & X_{n,p-1} \end{bmatrix}$$

Η εύρεση του βέλτιστου επιπέδου για την προσέγγιση των διαθέσιμων σημείων γίνεται και πάλι με ελαχιστοποίηση του αθροίσματος των τετραγώνων των σφαλμάτων :

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left( Y_i - (\beta_0 X_{i0} + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{i,p-1}) \right)^2 \quad (7.2.1.5)$$

Αποδεικνύεται ότι η ελαχιστοποίηση του  $\sum_{i=1}^n \varepsilon_i^2$  είναι ισοδύναμη με το σύστημα  $P$  εξισώσεων με  $p$  αγνώστους...

$$X'X\beta = X'Y$$

Με την προϋπόθεση ότι η τάξη του πίνακα  $X$  είναι  $rank(X)=p < n$ , οι ζητούμενες εκτιμήτριες ελαχίστων τετραγώνων των συντελεστών δίνονται από τη σχέση :

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (7.2.1.6)$$

Όπου  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{p-1})$

και είναι αμερόληπτες εκτιμήτριες των αντίστοιχων παραμέτρων που εκτιμούν, δηλαδή ισχύει

$$E(\hat{\beta}) = \beta \quad (7.2.1.7)$$

### 7.2.3 Η μεταβλητότητα του μοντέλου

Αντίστοιχα με την απλή γραμμική παλινδρόμηση έχουμε,

#### vi. ΑΘΡΟΙΣΜΑΤΑ ΤΕΤΡΑΓΩΝΩΝ

Η ολική μεταβλητότητα των παρατηρήσεων  $Y_i$  δίνεται από τον παρακάτω τύπο:

$$SSTO = \sum_{i=1}^n (Y_i - \bar{Y})^2 = SSR + SSE \quad (7.2.3.1)$$

Όπου :

$$\cdot SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = \hat{Y}'\hat{Y} - n\bar{Y}^2 = \hat{\beta}'X'Y - n\bar{Y}^2 \quad (7.2.3.2)$$

$$\cdot SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = Y'Y - \hat{\beta}'X'Y \quad (7.2.3.3)$$

**vii. ΣΥΝΤΕΛΕΣΤΗΣ ΜΕΡΙΚΟΥ ΠΡΟΣΔΙΟΡΙΣΜΟΥ**

Στην πολλαπλή γραμμική παλινδρόμηση, εκτός από τα αθροίσματα των τετραγώνων και τον συντελεστή προσδιορισμού θα πρέπει να μελετήσουμε άλλες δυο έννοιες. Εκείνη του πρόσθετου αθροίσματος τετραγώνων και εκείνη του συντελεστή μερικού προσδιορισμού.

Πιο συγκεκριμένα, αν θεωρήσουμε δυο ανεξάρτητες μεταβλητές  $X_1, X_2$ , η ποσότητα

$$\begin{aligned} SSR(X_2 \setminus X_1) &= SSR(X_1, X_2) - SSR(X_1) \\ &= SSE(X_1) - SSE(X_1, X_2) \end{aligned} \quad (7.2.3.3)$$

Ονομάζεται πρόσθετο άθροισμα τετραγώνων (*extra sum of squares*). Εκφράζει το επιπλέον τμήμα της συνολικής μεταβλητότητας του  $Y$  το οποίο εξηγείται με την εισαγωγή της μεταβλητής  $X_2$  σε ένα γραμμικό μοντέλο που ήδη υπάρχει η μεταβλητή  $X_1$ . Η έννοια του πρόσθετου αθροίσματος τετραγώνων μπορεί εύκολα να επεκταθεί και στην περίπτωση που ξεκινάμε με ένα μοντέλο το οποίο περιέχει περισσότερες από μία ανεξάρτητες μεταβλητές και ενδιαφερόμαστε για το αποτέλεσμα που θα έχει η προσθήκη σε αυτό κάποιας ή κάποιων επιπλέον μεταβλητών.

Επιπλέον, η ποσότητα

$$R_{Y2,1}^2 = \frac{SSR(X_2 \setminus X_1)}{SSE(X_1)} \quad (7.2.3.4)$$

Ονομάζεται συντελεστής μερικού προσδιορισμού και εκφράζει το ποσοστό της μεταβλητότητας που έχει μείνει ανεξιχνίαστη μετά τη χρήση της ανεξάρτητης μεταβλητής  $X_1$ , την οποία εξηγεί η προσθήκη στο μοντέλο και της μεταβλητής  $X_2$ . Η μοναδική διαφορά με τον απλό συντελεστή προσδιορισμού είναι ότι τώρα χρησιμοποιούμε ως βάση ένα μοντέλο που έχει διαμορφωθεί μερικώς με χρήση κάποιας ανεξάρτητης μεταβλητής και εξετάζουμε το όφελος από την προσθήκη κάποιας άλλης.

**viii. ΜΕΣΑ ΑΘΡΟΙΣΜΑΤΑ ΤΕΤΡΑΓΩΝΩΝ**

Τα μέσα αθροίσματα τετραγώνων υπολογίζονται διαιρώντας το κάθε άθροισμα τετραγώνων με τους αντίστοιχους βαθμούς ελευθερίας.

Πιο συγκεκριμένα, ορίζονται ως εξής:

$$\bullet \quad MSTO = \frac{SSTO}{n-1} = \frac{Y'Y - n\bar{Y}^2}{n-1} \quad (7.2.3.5)$$

$$\bullet \quad MSR = \frac{SSR}{p-1} = \frac{\beta'X'Y - n\bar{Y}^2}{p-1} \quad (7.2.3.6)$$

$$\bullet \quad MSE = \frac{SSE}{n-p} = \frac{Y'Y - \beta'X'Y}{n-p} \quad (7.2.3.7)$$

#### 7.2.4 Έλεγχος υποθέσεων για τις παραμέτρους του μοντέλου

Στο γραμμικό μοντέλο πολλαπλής παλινδρόμησης  $Y=X\beta+\varepsilon$ , όπου  $\varepsilon\sim N(0, \sigma^2 I_n)$ , θεωρούμε έναν πίνακα  $A$  διάστασης  $r \times p$  και ένα διάνυσμα στήλη  $c$  διάστασης  $r \times 1$  (των οποίων τα στοιχεία είναι σταθεροί αριθμοί). Τότε ο έλεγχος της υπόθεσης

$H_0: A\beta=c$

$H_1: A\beta\neq c$

Σε επίπεδο σημαντικότητας  $(1-\alpha)\%$  γίνεται με την βοήθεια της στατιστικής συνάρτησης

$$F = \frac{n-p}{r} = \frac{(A\hat{\beta} - c)' A(X'X)^{-1} A' (A\hat{\beta} - c)}{(Y - X\hat{\beta})' (Y - X\hat{\beta})} \quad (7.2.4.1)$$

Απορρίπτεται η  $H_0$  όταν  $F > F_{r, n-p(\alpha)}$

##### ix. Έλεγχος t-test

Ο έλεγχος της υπόθεσης :

$H_0: \alpha'\beta=c$

$H_1: \alpha'\beta\neq c$

σε επίπεδο σημαντικότητας  $(1-\alpha)\%$ , όπου  $\alpha$  είναι ένα διάνυσμα  $\alpha=[\alpha_0, \alpha_1, \dots, \alpha_{p-1}]'$  γίνεται με τη βοήθεια της στατιστικής συνάρτησης

$$T = \frac{\alpha'\beta - c}{s\sqrt{\alpha'(X'X)^{-1}\alpha}} \quad (7.2.4.2)$$

Απορρίπτεται η  $H_0$  όταν  $|T| > t_{n-p(\alpha/2)}$

##### x. Έλεγχος F-test

Ο έλεγχος της υπόθεσης :

$H_0: \beta_1=\beta_2=\dots=\beta_{p-1}=0$

$H_1$ : κάποιο από τα  $\beta_i$  δεν είναι ίσο με 0

Σε επίπεδο σημαντικότητας  $(1-\alpha)\%$  γίνεται με τη βοήθεια της στατιστικής συνάρτησης

$$F^* = \frac{(n-p)}{(p-1)} \frac{(\beta'X'Y - n\bar{Y}^2)}{(Y'Y - \beta'X'Y)} = \frac{MSR}{MSE} \quad (7.2.4.3)$$

Απορρίπτεται η  $H_0$  όταν  $F^* > F_{p-1, n-p(\alpha)}$

### 7.2.5 Επιλογή βέλτιστου γραμμικού μοντέλου

Στη συνέχεια θα προσπαθήσουμε να προσδιορίσουμε το βέλτιστο γραμμικό μοντέλο για την πρόβλεψη της μεταβλητής απόκρισης  $Y$ .

Αρχικά, θα παρουσιάσουμε τη μέθοδο εξέτασης όλων των γραμμικών μοντέλων και στη συνέχεια ορισμένες επαναληπτικές μεθόδους με τις οποίες δημιουργούνται αυτόματα υποσύνολα ανεξαρτήτων μεταβλητών που προσφέρουν καλή προσαρμογή στα διαθέσιμα κάθε φορά δεδομένα. Το πρόβλημα που αντιμετωπίζουμε στην πράξη είναι ότι από τη μία μεριά, για την κατασκευή ενός γραμμικού μοντέλου με σκοπό την πρόβλεψη θέλουμε να έχουμε όσο το δυνατόν περισσότερες ανεξάρτητες μεταβλητές έτσι ώστε οι προσαρμοσμένες τιμές να είναι αξιόπιστες ενώ από την άλλη, λόγω χρόνου και κόστους θέλουμε το γραμμικό μοντέλο να περιέχει όσο το δυνατόν λιγότερες μεταβλητές. Το σημείο τομής αυτών των δυο θα ήταν η επιλογή ενός μικρού υποσυνόλου ανεξαρτήτων μεταβλητών για την δημιουργία ενός αποτελεσματικού μοντέλου πρόβλεψης.

- Στη μέθοδο της εξέτασης όλων των γραμμικών μοντέλων τα πιο συνηθισμένα κριτήρια που χρησιμοποιούνται είναι τα εξής :
  - i. Συντελεστής προσδιορισμού  $R_p^2$
  - ii. Άθροισμα τετραγώνων των υπολοίπων  $SSE_p$
  - iii. Μέσο τετραγωνικό σφάλμα  $MSE_p$
  - iv. Τροποποιημένος συντελεστής προσδιορισμού  $R_{adj}^2$
  - v. Συντελεστής  $C_p$  του *Mallows*

Ο δείκτης  $p$  σε κάθε κριτήριο δηλώνει τον αριθμό των παραμέτρων που εκτιμώνται με το αντίστοιχο μοντέλο δεδομένου ότι προσαρμόζουμε γραμμικά μοντέλα που περιέχουν διαφορετικό αριθμό μεταβλητών.

#### i. Συντελεστής προσδιορισμού $R_p^2$

Σύμφωνα με αυτό το κριτήριο, ένα γραμμικό μοντέλο θεωρείται βέλτιστο εάν δίνει στον συντελεστή προσδιορισμού  $R^2$  την μεγαλύτερη δυνατή τιμή. Επειδή, όμως, όσο αυξάνεται το πλήθος των ανεξάρτητων μεταβλητών που υπάρχουν στο μοντέλο το  $R^2$  αυξάνεται συνεχώς, θα προέκυπτε κάθε φορά ως βέλτιστο το πλήρες μοντέλο. Γεγονός που είναι πολύ χρονοβόρο και χωρίς καμία πρακτική αξία. Έτσι εφαρμόζουμε το κριτήριο  $R^2$  σε όλα τα δυνατά υποσύνολα, ανάλογα με το πλήθος των μεταβλητών δηλαδή ανά μία, δυο κ.λ.π. ανεξάρτητες μεταβλητές και τέλος στο πλήρες γραμμικό μοντέλο. Στη συνέχεια, συγκρίνουμε τις μέγιστες τιμές του  $R^2$  σε κάθε υποσύνολο και εξετάζουμε τις μεταβολές που γίνονται στην τιμή του κριτηρίου από το ένα υποσύνολο στο άλλο. Εάν η προσθήκη μιας μεταβλητής στο γραμμικό μοντέλο δεν επιφέρει σημαντική αύξηση στην τιμή τότε το βέλτιστο μοντέλο θεωρείται το μοντέλο εκείνο που αφενός δίνει τη μέγιστη τιμή στο κριτήριο  $R^2$  στο

υποσύνολο που ανήκει και αφετέρου, διατηρεί έναν μικρό αριθμό ανεξαρτήτων μεταβλητών.

**ii. Άθροισμα τετραγώνων των υπολοίπων  $SSE_p$**

Σύμφωνα με το κριτήριο αυτό, ένα γραμμικό μοντέλο θεωρείται βέλτιστο εάν δίνει την ελάχιστη τιμή στο άθροισμα των τετραγώνων των υπολοίπων  $SSE$ . Επειδή, όμως όσο αυξάνεται το πλήθος των ανεξάρτητων μεταβλητών που υπάρχουν στο μοντέλο το  $SSE$  μειώνεται συνεχώς, θα προέκυπτε και πάλι ως βέλτιστο το πλήρες γραμμικό μοντέλο. Το άθροισμα των τετραγώνων των υπολοίπων  $SSE_p$  συνδέεται με το συντελεστή προσδιορισμού  $R^2$  μέσω της σχέσης  $R_p^2 = 1 - \frac{SSE_p}{SSTO}$ . Επομένως, από τη στιγμή που το συνολικό άθροισμα τετραγώνων  $SSTO$  δεν επηρεάζεται από τις ανεξάρτητες μεταβλητές που υπάρχουν στο γραμμικό μοντέλο, η τιμή του κριτηρίου  $SSE_p$  μεταβάλλεται αντίστροφα ως προς τη μονοτονία από την αντίστοιχη τιμή του κριτηρίου  $R^2$ . Έτσι εφαρμόζουμε το κριτήριο  $SSE_p$  σε όλα τα δυνατά υποσύνολα, συγκρίνουμε τις μικρότερες τιμές του  $SSE_p$  σε κάθε υποσύνολο και εξετάζουμε τις μεταβολές που γίνονται στην τιμή του κριτηρίου από το ένα υποσύνολο στο άλλο. Εάν η προσθήκη μίας μεταβλητής στο γραμμικό μοντέλο δεν επιφέρει σημαντική μείωση στην τιμή του  $SSE_p$  τότε ως βέλτιστο θεωρείται το μοντέλο εκείνο που δίνει μικρή τιμή για το  $SSE_p$ , ενώ παράλληλα διατηρεί ένα μικρό αριθμό μεταβλητών.

**iii. Μέσο τετραγωνικό σφάλμα  $MSE_p$**

Σύμφωνα με το κριτήριο αυτό, ένα γραμμικό μοντέλο θεωρείται βέλτιστο εάν δίνει την ελάχιστη τιμή στο μέσο τετραγωνικό υπόλοιπο  $MSE_p$ . Το κριτήριο  $MSE_p$  συνδέεται με το κριτήριο  $SSE_p$  με τη σχέση  $MSE_p = \frac{SSE_p}{n-p}$ . Το  $MSE_p$  επηρεάζεται από την προσθήκη νέων ανεξάρτητων μεταβλητών στο μοντέλο, για το λόγο αυτόν αναζητούμε εκείνο το γραμμικό μοντέλο που δίνει τη μικρότερη τιμή για το  $MSE_p$  ενώ παράλληλα διατηρεί ένα μικρό αριθμό ανεξάρτητων μεταβλητών συγκρίνοντας τις μικρότερες τιμές του  $MSE_p$  σε κάθε υποσύνολο και εξετάζοντας σε ποιο σημείο αρχίζει να αυξάνεται η τιμή του κριτηρίου μετά τις συνεχόμενες μειώσεις που δέχεται από την προσθήκη μεταβλητών στο μοντέλο.

**iv. Τροποποιημένος συντελεστής προσδιορισμού  $R_{adj}^2$**

Σύμφωνα με το κριτήριο αυτό, ένα γραμμικό μοντέλο θεωρείται βέλτιστο εάν δίνει τη μέγιστη τιμή στον τροποποιημένο συντελεστή προσδιορισμού

$$R_{adj}^2 = 1 - \frac{SSE_p \backslash n - p}{SSTO \backslash n - 1} = 1 - \frac{n - 1}{SSTO} MSE_p \quad (7.2.5.1)$$

Για τον καθορισμό του βέλτιστου συνόλου ανεξάρτητων μεταβλητών εφαρμόζουμε το κριτήριο  $R_{adj}^2$  σε όλα τα δυνατά υποσύνολα, συγκρίνουμε τις μέγιστες τιμές του σε κάθε υποσύνολο και εξετάζουμε σε ποιο σημείο αρχίζει να μειώνεται η τιμή του

κριτηρίου μετά τις συνεχόμενες αυξήσεις που δέχεται από την προσθήκη μεταβλητών στο μοντέλο.

#### ν. Συντελεστής $C_p$ του Mallows

Ο συντελεστής δίνεται από τον τύπο

$$C_p = \frac{SSE_p}{MSE} - (n - 2p) \quad (7.2.5.2)$$

Για την επιλογή του βέλτιστου μοντέλου, υπολογίζουμε την τιμή του συντελεστή για όλα τα δυνατά υποσύνολα ανάλογα με το πλήθος των μεταβλητών, εντοπίζουμε την μικρότερη τιμή του συντελεστή σε κάθε υποσύνολο και επιλέγουμε εκείνο το υποσύνολο μεταβλητών για το οποίο η τιμή του  $C_p$  είναι μικρή και βρίσκεται κοντά στην τιμή του  $p$ .

- Οι πιο συνηθισμένες μέθοδοι επαναληπτικών διαδικασιών που χρησιμοποιούνται για τη δημιουργία ενός βέλτιστου μοντέλου ανεξάρτητων μεταβλητών είναι

#### ι. Η μέθοδος της κατά βήμα παλινδρόμησης (stepwise regression)

Προσαρμόζουμε όλα τα απλά γραμμικά μοντέλα

$$Y_i = \beta_0 + \beta_k X_{i,k} + \varepsilon_i, \quad k = 1, 2, \dots, p - 1$$

Και για κάθε ένα από αυτά υπολογίζουμε την ποσότητα

$$F_k^* = \frac{MSR(X_k)}{MSE(X_k)}$$

Από όλες τις μεταβλητές για τις οποίες ισχύει  $F_1^* > F_{1,n-p}(a)$  επιλέγουμε τη μεταβλητή εκείνη που μεγιστοποιεί την  $F_k^*$  (έστω  $X_1$ ) και την εισάγουμε στο μοντέλο. Στη συνέχεια, προσαρμόζουμε όλα τα μοντέλα

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_k X_{i,k} + \varepsilon_i, \quad k = 1, 2, \dots, p - 1$$

και επιλέγουμε να εισάγουμε στο μοντέλο τη μεταβλητή εκείνη (έστω  $X_2$ ) που μεγιστοποιεί την ποσότητα

$$F_{k,1}^* = \frac{MSR(X_k \setminus X_1)}{MSE(X_k)} > F$$

Οπότε προκύπτει το μοντέλο

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_k X_{i,k} + \varepsilon_i, \quad k = 1, 2, \dots, p - 1$$

Στη συνέχεια, γίνεται έλεγχος για τυχόν απομάκρυνση από το μοντέλο μεταβλητών που κρίνονται άχρηστες μετά την εισαγωγή της τελευταίας που μπήκε. Πιο συγκεκριμένα, επιλέγουμε τις μεταβλητές  $X_k$  για τις οποίες ισχύει

$$F_{k,2}^* = \frac{MSR(X_k \setminus X_2)}{MSE(X_2, X_k)} < F$$

Αν δεν υπάρχουν τέτοιες μεταβλητές, τότε δεν εξάγουμε καμία μεταβλητή από το μοντέλο. Διαφορετικά, επιλέγουμε να εξάγουμε από το μοντέλο τη μεταβλητή που ελαχιστοποιεί το  $F_{k,2}^*$  και εφαρμόζουμε ξανά τα δυο τελευταία βήματα.

**ii. Η μέθοδος της προς τα εμπρός επιλογής (forward selection)**

Μια μέθοδος παρόμοια με την προηγούμενη. Η μόνη διαφορά βρίσκεται στο γεγονός ότι τώρα δεν μπορεί να απομακρυνθεί από το μοντέλο κάποια μεταβλητή που κρίθηκε μη σημαντική μετά την εισαγωγή κάποιας άλλης μεταβλητής στο μοντέλο με την οποία έχει υψηλή συσχέτιση.

**iii. Η μέθοδος της προς τα πίσω απαλοιφής (backward elimination)**

Ξεκινώντας από το πλήρες γραμμικό μοντέλο

$$Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

και για τις μεταβλητές  $X_k$  για τις οποίες ισχύει

$$F_k^{**} = \frac{\text{MSR}(X_k \setminus X_1, \dots, X_{p-1})}{\text{MSE}(X_1, \dots, X_{p-1})} < F$$

Απορρίπτουμε από το μοντέλο τη μεταβλητή  $X_k$  που ελαχιστοποιεί την  $F_k^{**}$ . Η διαδικασία επαναλαμβάνεται με επόμενα βήματα μέχρι να μη γίνεται να απορρίψουμε καμία μεταβλητή.



## ΚΕΦΑΛΑΙΟ 8: ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

### 8.1 ΠΕΡΙΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ

Για την διεξαγωγή των αποτελεσμάτων θα χρησιμοποιηθεί το στατιστικό πακέτο *IMB SPSS 19.0*.

Στον πίνακα 8.1 παρουσιάζονται τα περιγραφικά μέτρα των μεταβλητών των δεδομένων. Πιο συγκεκριμένα έχουν υπολογιστεί τα εξής :

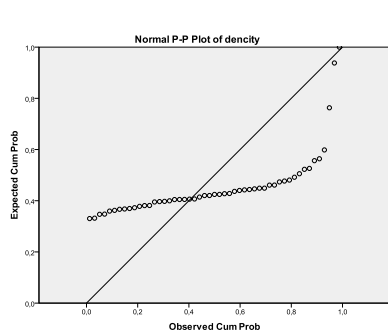
- Το εύρος *range*. Ενδεικτικά, η διαφορά της ελάχιστης τιμής από την μέγιστη τιμή της μεταβλητής που δείχνει το ποσοστό του αστικού πληθυσμού του εκάστοτε νομού είναι 70,3.
- Η ελάχιστη τιμή *minimum*. Ενδεικτικά, ο ελάχιστος μέσος αριθμός ετών των συνολικών επιτυχημένων εκπληρωμένων εκπαιδευτικών δραστηριοτήτων του πληθυσμού είναι περίπου 6 έτη.
- Η μέγιστη τιμή *maximum*. Ενδεικτικά, ο μέγιστος μέσος αριθμός ετών των συνολικών επιτυχημένων εκπληρωμένων εκπαιδευτικών δραστηριοτήτων του πληθυσμού είναι περίπου 9 έτη.
- Η μέση τιμή *mean*. Ενδεικτικά, η μέση τιμή του ποσοστού συμμετοχής στον εργασιακό χώρο είναι 44,4.
- Η τυπική απόκλιση *st.deviation*. Ενδεικτικά, η τυπική απόκλιση της μακροχρόνιας ανεργίας είναι 14,9.

Ας σημειωθεί, επίσης, ότι από τον πίνακα φαίνεται ότι δεν υπάρχουν *missing values* στο αρχείο των δεδομένων για καμία από τις μεταβλητές.

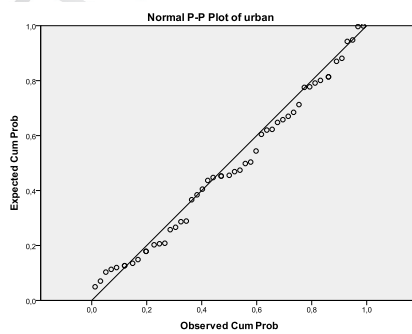
Πίνακας 8.1  
Περιγραφικά μέτρα δεδομένων

Descriptive Statistics						
	N	Range	Minimum	Maximum	Mean	Std. Deviation
density	51	971,3068	16,5635	987,8703	76,655105	136,8442237
urban	51	70,300000	28,700000	99,000000	53,16277927	14,843425487
natural	51	12,1380	-7,5827	4,5552	-1,775885	2,7775425
masculin	51	4,2136	48,1673	52,3809	50,504355	,9930224
tomeas1	51	48,2775	1,3259	49,6034	25,312981	10,4125659
tomeas2	51	23,0906	11,8118	34,9024	20,166911	4,7891207
tomeas3	51	37,7761	31,7268	69,5029	49,712644	8,0729728
anergiat	51	16,4452	7,2044	23,6496	12,179215	2,9163135
anergiam	51	12,8374	6,9499	19,7874	10,731203	2,5089131
anergiaf	51	22,7350	7,6363	30,3713	14,809673	3,8827171
incomecapita	51	5203,2934	1751,4001	6954,6934	3697,962784	783,6310582
long_term_unemployed	51	59,0862	13,1353	72,2216	52,606171	14,9805855
eduleveltotal	51	3,0098	6,8794	9,8892	7,949543	,5718227
edulevelmales	51	2,8421	7,3472	10,1893	8,334579	,5203440
edulevelfemales	51	3,2774	6,3321	9,6094	7,559012	,6532110
participationt	51	13,3538	36,6249	49,9787	44,476290	3,2779067
Valid N (listwise)	51					

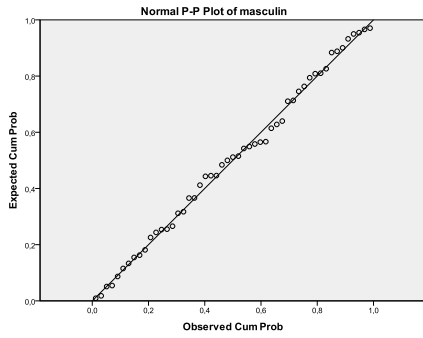
Στη συνέχεια παρουσιάζονται κάποια διαγράμματα *P-P plots* των μεταβλητών αυτών.



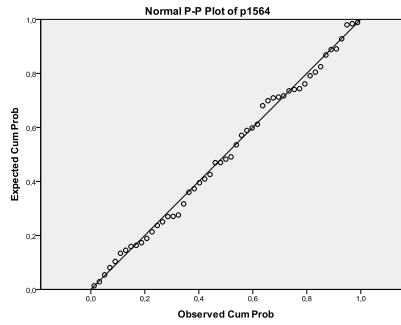
Διάγραμμα 8.1  
*P-P plot (density)*



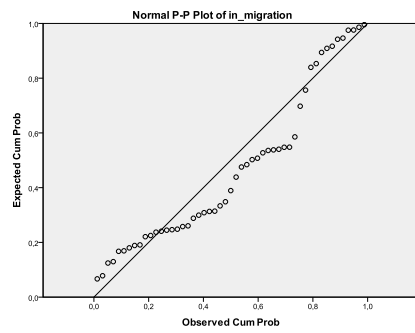
Διάγραμμα 8.2  
*P-P plot (urban)*



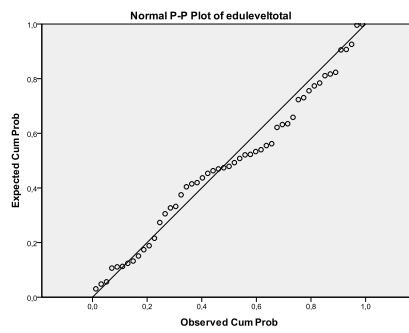
Διάγραμμα 8.3  
P-P plot (masculin)



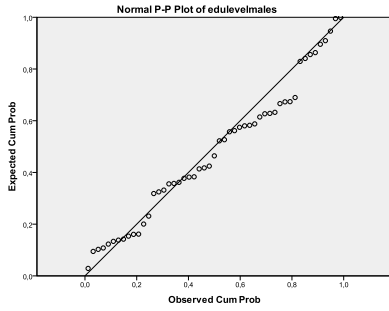
Διάγραμμα 8.4  
P-P plot (p1564)



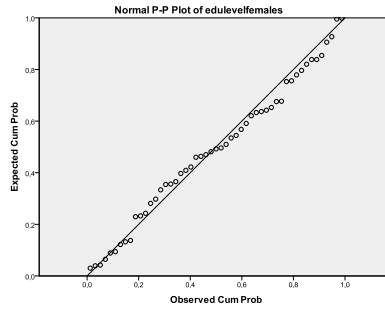
Διάγραμμα 8.5  
P-P plot (in\_migration)



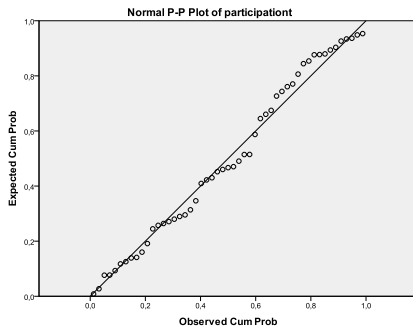
Διάγραμμα 8.6  
P-P plot (eduleveltotal)



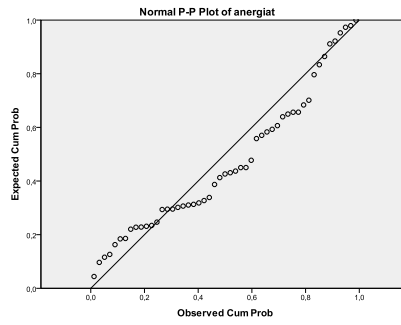
Διάγραμμα 8.7  
*P-P plot (edulevelfemales)*



Διάγραμμα 8.8  
*P-P plot (edulevelmales)*

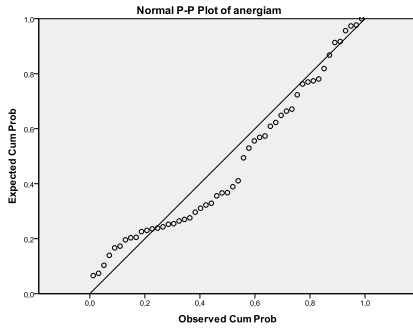


Διάγραμμα 8.9  
*P-P plot (participation)*

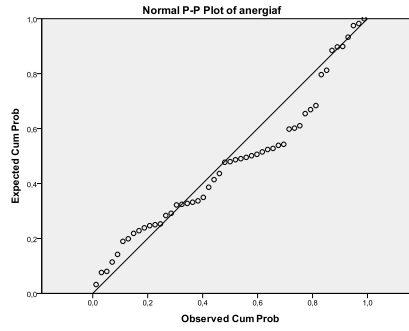


Διάγραμμα 8.10  
*P-P plot (anergiat)*

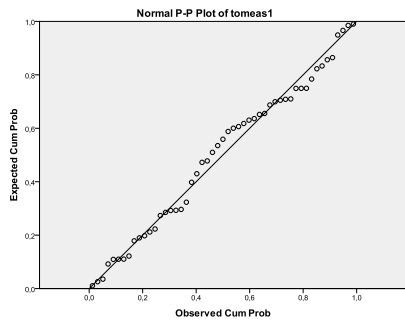
ΠΑΝΕΠΙΣΤΗΚ



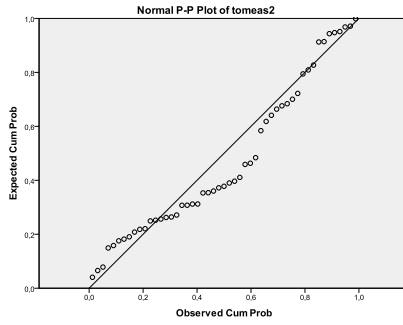
Διάγραμμα 8.11  
*P-P plot (anergiam)*



Διάγραμμα 8.12  
*P-P plot (anergiaf)*

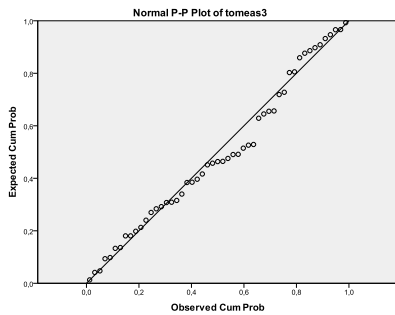


Διάγραμμα 8.13  
*P-P plot (tomeas1)*

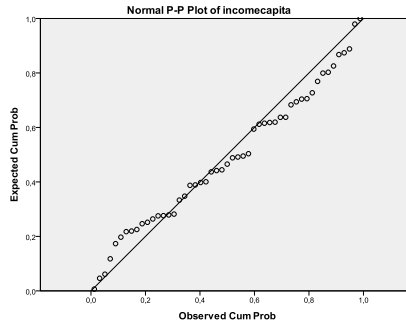


Διάγραμμα 8.14  
*P-P plot (tomeas2)*

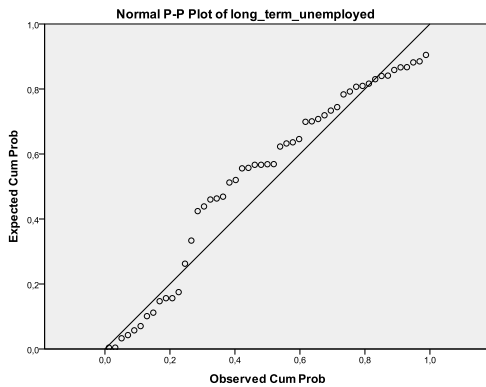
ΠΑΝΕΠΙΣΤΗΚ



Διάγραμμα 8.15  
P-P plot (tomeas3)



Διάγραμμα 8.16  
P-P plot (incomcapita)



Διάγραμμα 8.15  
P-P plot (long\_term\_unemployed)

## 8.2 ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Εφαρμόζουμε την πολλαπλή γραμμική παλινδρόμηση στα δεδομένα μας. Αρχικά, θα πρέπει να καθορίσουμε ποιες από τις μεταβλητές θα συμπεριληφθούν στην παλινδρομική εξίσωση πριν την ανάλυση και στη συνέχεια θα καθορίσουμε τις μεταβλητές που θα στην παλινδρόμηση στην διάρκεια της εκτέλεσης της.

Στον σχεδιασμό της εργασίας συμπεριλαμβάνεται η επιλογή των μεταβλητών που θα καθορίσουν και το μοντέλο που τελικά θα αναλύσουμε.

Ως εξαρτημένη μεταβλητή ορίζουμε την *in\_migration*. Η μεταβλητή αυτή είναι υπολογισμένη ως εξής: Διαιρέσαμε τους «εγκατασταθέντες από άλλο Δήμο» προς το «σύνολο των ατόμων του Δήμου» και το πολλαπλασιάσαμε επί 100. Μας δίνει, λοιπόν, ένα μέτρο εσωτερικής μετανάστευσης/εγκατάστασης το οποίο θα έχει πάντα θετικό πρόσημο. Ως εκ τούτου είναι πιο εύκολο στην ερμηνεία (ως μέγεθος έλξης

Formatted: Greek

Formatted: Greek

εσωτερικών μεταναστών) από την μεταβλητή  $net$  που αποτελεί τον δείκτη καθαρής μετανάστευσης.

Ως ανεξάρτητες μεταβλητές ορίζουμε τις παρακάτω :

- $urban$
- $eduleveltotal$
- $participation$
- $anergiat$
- $tomeas3$
- $incomecapita$
- $long\_term\_unemployed$

Πριν καταλήξουμε στο βέλτιστο μοντέλο μας θα πρέπει να εφαρμόσουμε ορισμένους διαγνωστικούς ελέγχους.

Συνοπτικά αναφέρονται ως εξής:

- Έλεγχος της κανονικότητας της εξαρτημένης μεταβλητής. Θα ελέγξουμε εάν η εξαρτημένη μας μεταβλητή ακολουθεί την κανονική κατανομή. Εάν όχι, θα προβούμε σε κάποιους μετασχηματισμούς (π.χ. θα μπορούσαμε να υπολογίσουμε και να χρησιμοποιήσουμε τον λογάριθμο της εξαρτημένης μεταβλητής).
- Συντελεστές συσχέτισης των μεταβλητών. Θα ελέγξουμε την ύπαρξη σχέσης μεταξύ της εξαρτημένης και των ανεξάρτητων μεταβλητών αλλά και την ύπαρξη σχέσης ανάμεσα στις ανεξάρτητες μεταβλητές.
- Διαγράμματα διασποράς. Θα σχεδιάσουμε τα κατάλληλα διαγράμματα διασποράς ώστε να ελέγξουμε την ετεροσκεδαστικότητα.
- Πρώτη επιλογή βέλτιστου μοντέλου. Χρησιμοποιώντας τους συντελεστές  $R^2$ ,  $R^2_{adj}$ ,  $MSE_p$  και  $C_p$ , θα λάβουμε μια πρώτη εικόνα για το ποιο είναι το βέλτιστο μοντέλο.
- Backward Elimination. Με την μέθοδο αυτή θα αποφασίσουμε, τελικά, ποιο μοντέλο περιγράφει καλύτερα την συσχέτιση μεταξύ των μεταβλητών μας.
- Ανάλυση Σφαλμάτων. Τέλος, θα ελέγξουμε αν ισχύουν οι κατάλληλες προϋποθέσεις για τα σφάλματα του μοντέλου που επιλέξαμε ως βέλτιστο.

Formatted: List Paragraph, Bulleted + Level: 1 + Aligned at: 0,63 cm + Indent at: 1,27 cm

Formatted: Font: Italic, Greek

Formatted: Font: Italic, Greek

Formatted: Font: Italic, Greek

Formatted: Greek

Formatted: Greek

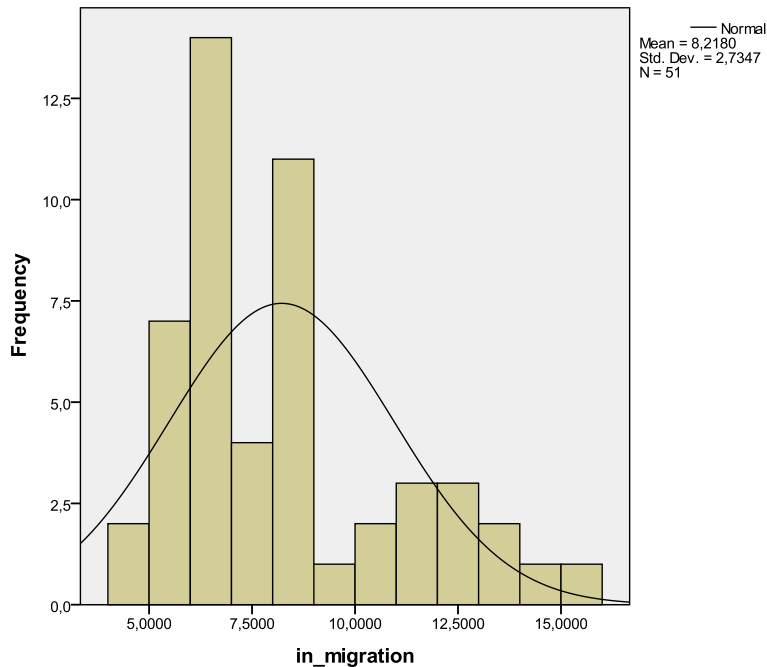
Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Underline, Font color: Auto

### 8.2.1 Μελέτη της κανονικότητας της εξαρτημένης μεταβλητής

Μετά την αρχική κατανόηση των δεδομένων, θα πρέπει να ελέγξουμε την κατανομή της εξαρτημένης μεταβλητής ( $in\_migration$ ).

Η κανονικότητα του εξεταζόμενου μεγέθους μπορεί να απεικονιστεί γραφικά μέσω του Ιστογράμματος ( $histogram$ ) και να επιβεβαιωθεί στατιστικά μέσω του μη παραμετρικού ελέγχου  $Kolmogorov-Smirnov$ .



Διάγραμμα 8.16  
Ιστόγραμμα εξαρτημένης μεταβλητής

Από το ιστόγραμμα παίρνουμε μια πρώτη, αλλά αρκετά σαφή εικόνα, για την κατανομή των δεδομένων. Στο ιστόγραμμα των στερεών καταλοίπων παρατηρούμε ότι η κατανομή της εξαρτημένης μεταβλητής τείνει να είναι κανονική.

Το παραπάνω συμπέρασμα θα το επιβεβαιώσουμε στατιστικά μέσω του μη παραμετρικού ελέγχου *Kolmogorov – Smirnov*. Βάσει του ελέγχου αυτού μπορούμε να δεχθούμε ή να απορρίψουμε την αρχική μας υπόθεση ότι η κατανομή είναι κανονική. Παρατηρώντας την τιμή του *sig* στον πίνακα που ακολουθεί απορρίπτουμε ή όχι την εξής υπόθεση:

$H_0$ : η εξαρτημένη μεταβλητή ακολουθεί κανονική κατανομή  
 $H_1$ : η εξαρτημένη μεταβλητή δεν ακολουθεί κανονική κατανομή

Αν η τιμή αυτή είναι μεγαλύτερη από το επίπεδο σημαντικότητας αποδεχόμαστε την  $H_0$ . Αντίστοιχα, αν η τιμή είναι μικρότερη την απορρίπτουμε.

Χρησιμοποιώντας και πάλι το στατιστικό πακέτο *SPSS*, έχουμε :

Formatted: Centered

Formatted: Font: Italic

Formatted: Centered

Formatted: Font: Italic



**Πίνακας 8.2**  
**Έλεγχος Κανονικότητας εξαρτημένης μεταβλητής**

One-Sample Kolmogorov-Smirnov Test		
		in_migration
N		51
Normal Parameters <sup>a,b</sup>	Mean	8,218041
	Std. Deviation	2,7347329
Most Extreme Differences	Absolute	,178
	Positive	,178
	Negative	-,089
Kolmogorov-Smirnov Z		1,268
Asymp. Sig. (2-tailed)		<b>,080</b>

a. Test distribution is Normal.

b. Calculated from data.

~~Πίνακας 8.2 Έλεγχος κανονικότητας εξαρτημένης μεταβλητής~~

Όπως φαίνεται από τον πίνακα 8.2, η τιμή του *sig* είναι 0,080. Συνεπώς, σε επίπεδο σημαντικότητας  $\alpha=5\%$ , ισχύει ότι  $0,080 > 0,050$ . Άρα, αποδεχόμαστε την αρχική μας υπόθεση και η εξαρτημένη μεταβλητή ακολουθεί κανονική κατανομή. Συνεπώς, δεν υπάρχει λόγος να προβούμε σε κάποιον μετασχηματισμό. Συνεχίζουμε την ανάλυση με την εξαρτημένη μεταβλητή ως έχει.

### 8.2.2 Σχέση σύνδεσης των μεταβλητών

Σε ένα επόμενο βήμα πρέπει να ελέγξουμε αν οι ανεξάρτητες μεταβλητές που θα χρησιμοποιηθούν για την επεξήγηση της συμπεριφοράς της εξαρτημένης μεταβλητής συνδέονται γραμμικά με το εξαρτημένο μέγεθος. Την ύπαρξη ή όχι μιας τέτοιας σχέσης θα παρατηρήσουμε γραφικά σε διαγράμματα και θα επιβεβαιώσουμε στατιστικά μέσω των πινάκων συσχετίσεων.

Όσο μεγαλύτερη είναι η συσχέτιση τόσο μεγαλύτερο μέρος πληροφοριών της εξαρτημένης μεταβλητής εξηγεί η ανεξάρτητη μεταβλητή.

Ωστόσο, όταν οι ανεξάρτητες μεταβλητές είναι περισσότερες από μία είναι σκόπιμο, οι ανεξάρτητες μεταβλητές, μεταξύ τους να μην είναι ισχυρά συσχετισμένες.

Όταν οι ανεξάρτητες μεταβλητές είναι συσχετισμένες επεξηγούν το ίδιο μέρος της διακύμανσης της εξαρτημένης μεταβλητής.

Η συμμετοχή δύο ανεξάρτητων συσχετισμένων μεταβλητών αυξάνει στην ουσία το λάθος του μοντέλου, λόγω ύπαρξης πολυσυγγραμικότητας.

Για το λόγο αυτό, είναι εύλογο, πριν την εφαρμογή της γραμμικής παλινδρόμησης να εξετάζονται οι συσχετίσεις των μεταβλητών μεταξύ τους.

Formatted: Centered

Formatted Table

Formatted: Right: 0 cm, Line spacing: At least 20 pt

Formatted: Font: Italic

Formatted: Underline, Font color: Auto

Η μελέτη των συσχετίσεων γίνεται μέσω των πινάκων συσχετίσεων. Στους πίνακες αυτούς απεικονίζεται η τιμή του δείκτη που περιγράφει την συσχέτιση καθώς και την τιμή του ενός ελέγχου σημαντικότητας για την τιμή του συντελεστή αυτού. Ένας τέτοιος πίνακας είναι και ο πίνακας 8.3.

Πίνακας 8.3

Συσχετίσεις μεταβλητών

		Correlations							
		in_migration	urban	eduleveltotal	participationt	anergiat	tomeas3	incomecapita	long_term_unemployed
in_migration	Pearson Correlation	1	,171	,469**	-,129	-,197	,550**	,361**	-,278*
	Sig. (2-tailed)		,230	,001	,365	,166	,000	,009	,049
	N	51	51	51	51	51	51	51	51
urban	Pearson Correlation	,171	1	,618**	,403**	-,010	,333*	,685**	,052
	Sig. (2-tailed)	,230		,000	,003	,944	,017	,000	,718
	N	51	51	51	51	51	51	51	51
eduleveltotal	Pearson Correlation	,469**	,618**	1	,214	-,072	,722**	,812**	-,130
	Sig. (2-tailed)	,001	,000		,131	,613	,000	,000	,363
	N	51	51	51	51	51	51	51	51
participationt	Pearson Correlation	-,129	,403**	,214	1	-,042	-,086	,295*	-,318*
	Sig. (2-tailed)	,365	,003	,131		,769	,549	,035	,023
	N	51	51	51	51	51	51	51	51
anergiat	Pearson Correlation	-,197	-,010	-,072	-,042	1	,123	-,071	,083
	Sig. (2-tailed)	,166	,944	,613	,769		,390	,618	,564
	N	51	51	51	51	51	51	51	51
tomeas3	Pearson Correlation	,550**	,333*	,722**	-,086	,123	1	,639**	-,301*
	Sig. (2-tailed)	,000	,017	,000	,549	,390		,000	,032
	N	51	51	51	51	51	51	51	51
incomecapita	Pearson Correlation	,361**	,685**	,812**	,295*	-,071	,639**	1	-,178
	Sig. (2-tailed)	,009	,000	,000	,035	,618	,000		,212
	N	51	51	51	51	51	51	51	51
long_term_unemployed	Pearson Correlation	-,278*	,052	-,130	-,318*	,083	-,301*	-,178	1
	Sig. (2-tailed)	,049	,718	,363	,023	,564	,032	,212	
	N	51	51	51	51	51	51	51	51

\*\* Correlation is significant at the 0.01 level (2-tailed).  
\* Correlation is significant at the 0.05 level (2-tailed).

Πίνακας 8.3 Συσχετίσεις μεταβλητών

Από τον πίνακα 8.3 παρατηρούμε ότι από τις ανεξάρτητες μεταβλητές η μεταβλητή *eduleveltotal*, η *tomeas3*, η *incomecapita* και η *longtermunemployed* είναι οι ισχυρότερα συσχετισμένες με την εξαρτημένη μεταβλητή *in\_migration*. Η μεγαλύτερη κατά απόλυτη τιμή του συντελεστή συσχέτισης Pearson που περιγράφει τη σχέση της εξαρτημένης μεταβλητής και των ανεξάρτητων είναι η τιμή 0,55. Παρουσιάζεται ανάμεσα στην εξαρτημένη μεταβλητή και τη μεταβλητή του τριτογενή τομέα οικονομικής δραστηριότητας. Επομένως, διαφαίνεται να είναι ισχυρή η πρόβλεψη της *in\_migration* από την *tomeas3*.

Από τις επτά ανεξάρτητες μεταβλητές, οι τρεις έχουν αρνητική συσχέτιση με την εξαρτημένη μεταβλητή. Αυτές είναι οι *participation*, *anergiat*, *long\_term\_unemployed*. Αυτό φαίνεται να είναι λογικό, μιας και οι τρεις μεταβλητές αυτές αφορούν στην συμμετοχή του πληθυσμού στο εργατικό δυναμικό, στην ανεργία κ.λ.π.

Στην συνέχεια, παρατηρούμε ότι η μεγαλύτερη κατά απόλυτη τιμή του συντελεστή Pearson παρουσιάζεται μεταξύ των μεταβλητών *incomecapita* και *eduleveltotal* και είναι ίση με 0,812. Επίσης, παρατηρούμε ότι οι σχέσεις μεταξύ των μεταβλητών *eduleveltotal-urban*, *eduleveltotal-tomeas3*, *eduleveltotal-incomecapita* και *incomecapita-tomeas3* είναι πολύ ισχυρά συσχετισμένες μεταξύ τους (πολυσυγγραμκότητα). Εάν είχαμε στην ευχέρεια μας μόνο τον συγκεκριμένο τρόπο

Formatted: Centered

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic, Greek

Formatted: Font: Italic, No underline

Formatted: Greek

Formatted: Greek

Formatted: Greek

Formatted: Greek

Formatted: Greek

Formatted: Greek

Formatted: Greek

Formatted: Greek

Formatted: Greek

Formatted: Greek

Formatted: Greek

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

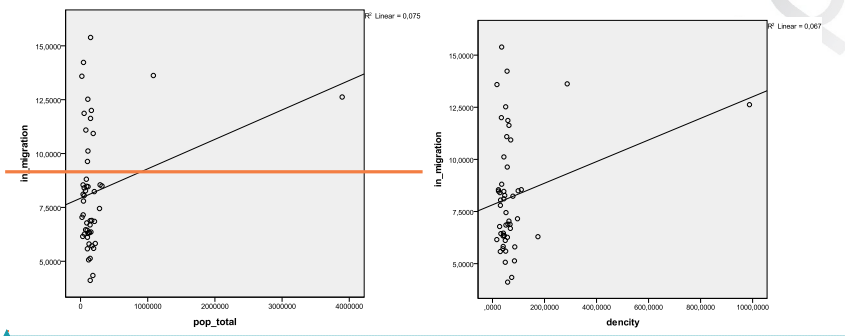
Formatted: Font: Italic

επιλογής των κατάλληλων ανεξάρτητων μεταβλητών θα επιλέγαμε τις παραπάνω λόγω της ισχυρής τους συσχέτισης με την εξαρτημένη. Φυσικά, λόγω των ισχυρών συσχετίσεων της μεταβλητής *eduleveltotal* με αρκετές από τις υπόλοιπες μεταβλητές, δεν θα την επιλέγαμε για το τελικό (βέλτιστο) μας μοντέλο. Παρ' όλα αυτά, για να επιβεβαιώσουμε την επιλογή μας θα συνεχίσουμε εφαρμόζοντας και άλλες μεθόδους πριν την κατάληξη μας.

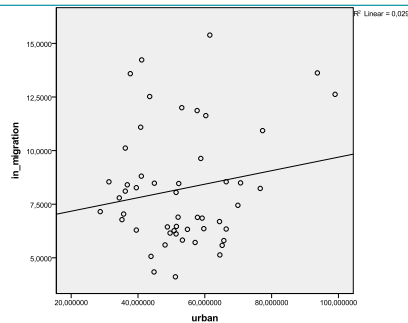
Formatted: Font: Italic

### 8.2.3 ΔΙΑΓΡΑΜΜΑΤΑ ΔΙΑΣΠΟΡΑΣ ΜΕΤΑΞΥ ΜΕΤΑΒΛΗΤΩΝ

Σχεδιάζοντας τα κατάλληλα διαγράμματα διασποράς (scatterplots) θα δούμε την γραφική απεικόνιση των σχέσεων των ανεξάρτητων μεταβλητών με την εξαρτημένη μεταβλητή.

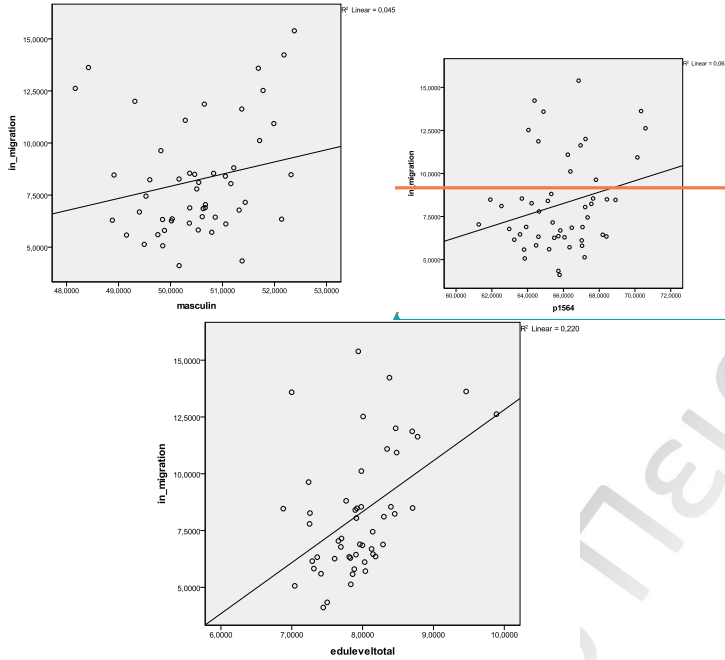


Formatted: Font: (Default) Times New Roman, 12 pt



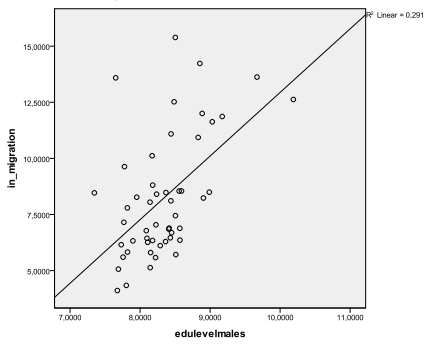
Διάγραμμα 8.17  
Scatterplot (*in\_migration*-*density*)

Διάγραμμα 8.18  
Scatterplot (*in\_migration*-*urban*)

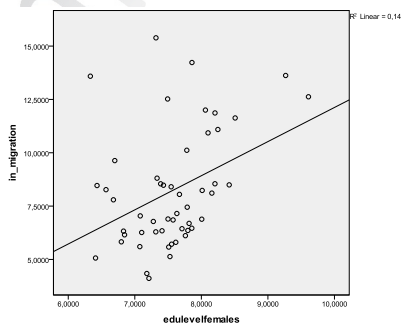


Διάγραμμα 8.19  
Scatterplot (in\_migration-masculin)

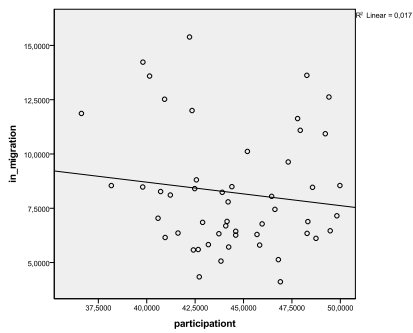
Διάγραμμα 8.20  
Scatterplot (in\_migration-



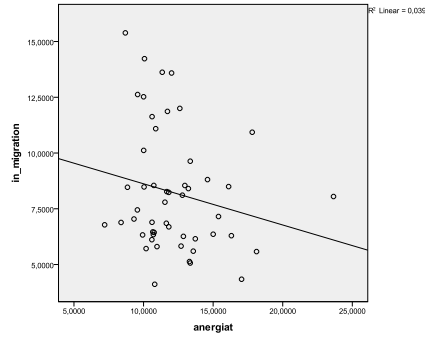
Διάγραμμα 8.21  
Scatterplot(in\_migration-edulevelfemales)



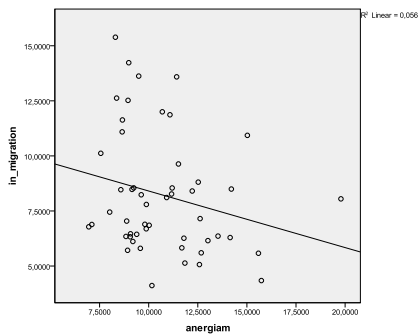
Διάγραμμα 8.22  
Scatterplot(in\_migration-edulevelmales)



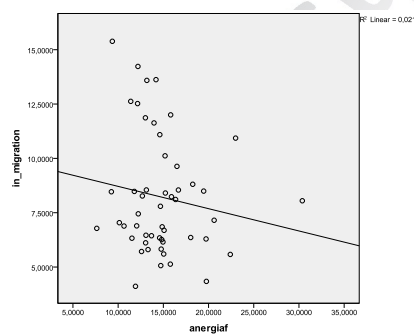
Διάγραμμα 8.23  
Scatterplot (in\_migration-participation)



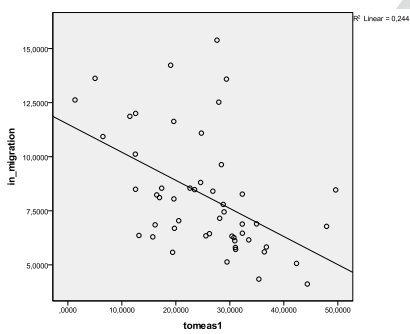
Διάγραμμα 8.24  
Scatterplot (in\_migration-anergia1)



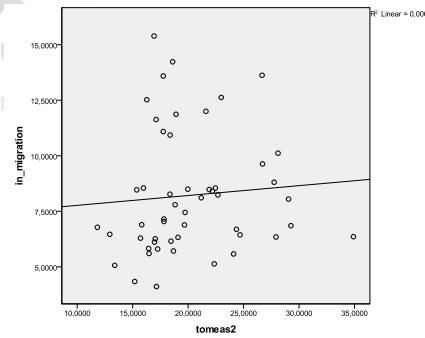
Διάγραμμα 8.25  
Scatterplot (in\_migration-anergia2)



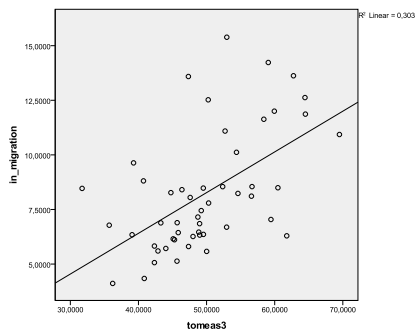
Διάγραμμα 8.26  
Scatterplot (in\_migration-anergia3)



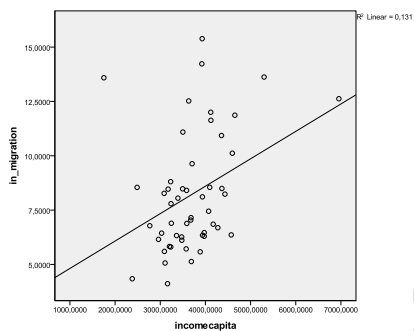
Διάγραμμα 8.27  
Scatterplot (in\_migration-tomes1)



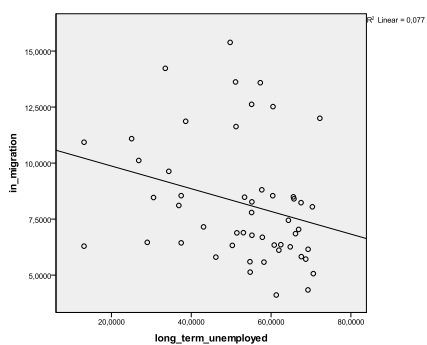
Διάγραμμα 8.28  
Scatterplot (in\_migration-tomes2)



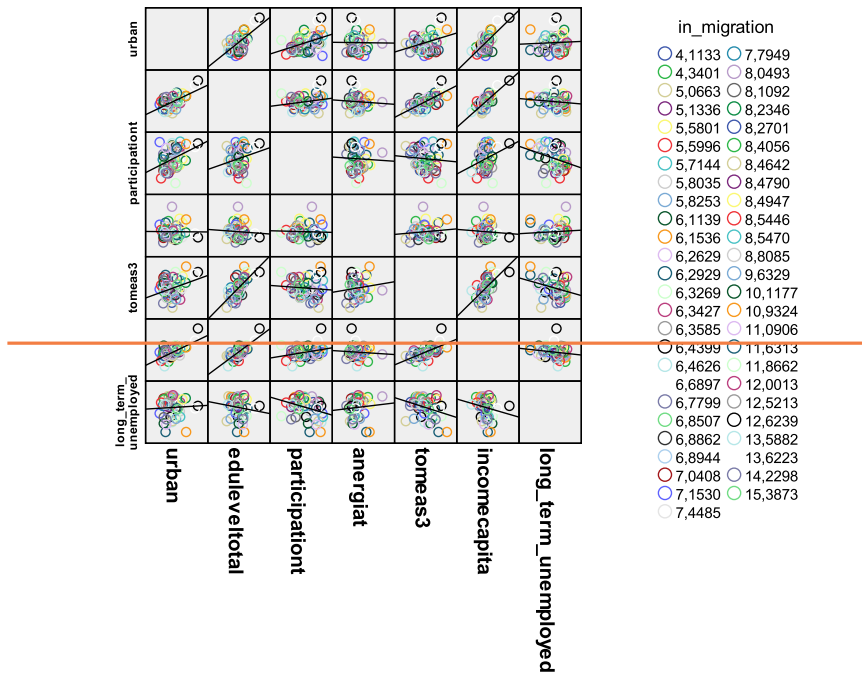
Διάγραμμα 8.29  
Scatterplot (in\_migration-tomeas3)



Διάγραμμα 8.30  
Scatterplot (in\_migration-incomecapita)



Διάγραμμα 8.31  
Scatterplot (in\_migration-long\_term\_unemployed)



Μελετώντας τα παραπάνω διαγράμματα συσχετίσεων παρατηρούμε ότι, ίσως, σημαντικότερο ρόλο στην εσωτερική μετανάστευση παίζουν οι εξής παράγοντες :

1. Η μεταβλητή *urban* (παρατηρείται μία θετική συσχέτιση, στατιστικά μη σημαντική εφόσον  $sig=0,23$ )
2. Η μεταβλητή *eduleveltotal* (παρατηρείται μία θετική συσχέτιση, στατιστικά σημαντική σε ε.σ. 1% , εφόσον  $sig=0,01$ )
3. Η μεταβλητή *anergiati* (παρατηρείται μία αρνητική συσχέτιση, στατιστικά μη σημαντική εφόσον  $sig=0,166$ )
4. Η μεταβλητή *tomeas3* (παρατηρείται μια θετική συσχέτιση, στατιστικά σημαντική σε ε.σ. 1% εφόσον  $sig=0,003$ )
5. Η μεταβλητή *incomecapita* (παρατηρείται μια θετική συσχέτιση, στατιστικά σημαντική σε ε.σ. 1% εφόσον  $sig=0,009$ )
6. Η μεταβλητή *longtermunemployed* (παρατηρείται μια αρνητική συσχέτιση στατιστικά σημαντική σε ε.σ. 5% εφόσον  $sig=0,049$ )

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Greek

Formatted: Font: Italic

#### 8.2.4 Πρώτη επιλογή βέλτιστου μοντέλου

Formatted: Underline, Font color: Auto

Για να πάρουμε μία πρώτη εικόνα για το αν οι παραπάνω παράγοντες επηρεάζουν την εσωτερική μετανάστευση θα υπολογίσουμε τον συντελεστή προσδιορισμού  $R^2$  για όλα τα δυνατά μοντέλα.

Formatted: Underline

Το πλήρες μοντέλο είναι το εξής :

$$\begin{aligned}
 Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 \\
 &= \beta_0 + \beta_1 \text{urban} + \beta_2 \text{edulevelttotal} + \beta_3 \text{anergiat} \\
 &\quad + \beta_4 \text{tomeas3} + \beta_5 \text{incomecapita} + \beta_6 \text{longtermunemployed} \\
 &\quad + \varepsilon_i \quad (8.2.4.1)
 \end{aligned}$$

Formatted: Greek

Προκειμένου να μειωθούν οι συνδυασμοί που θα πρέπει να εκτελεστούν ώστε να καταλήξουμε στο βέλτιστο μοντέλο θα χρησιμοποιήσουμε το *t-directed search* (διορθωτική παρέμβαση *t*).

Προκύπτει, λοιπόν, ότι :

Πίνακας 8.4  
Πλήρες μοντέλο

Formatted: Centered

Formatted: Greek

Formatted Table

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2,334	6,441		-,362	,719
	urban	-,003	,032	-,019	-,107	,916
	edulevelttotal	1,025	1,145	,214	,895	,376
	anergiat	-,230	,117	-,245	-1,972	,055
	tomeas3	,169	,066	,499	2,577	,013
	incomecapita	-,001	,001	-,155	-,673	,504
	long_term_unemployed	-,019	,023	-,106	-,828	,412

Formatted: Font color: Red

Formatted: Font color: Red

a. Dependent Variable: in\_migration

Πίνακας 8.4 Πλήρες μοντέλο

Formatted: Indent: Left: 0,11 cm, Right: 0,11 cm, Line spacing: At least 16 pt



Ο πίνακας 8.4 αποτελεί το αποτέλεσμα που λήφθηκε εκτιμώντας παλινδρόμηση με το πλήρες μοντέλο. Βλέπουμε ότι το μικρότερο  $\text{sig}$  παρατηρείται στην μεταβλητή  $\text{tomeas3}$  ( $\text{sig}=0,013$ ) και  $\text{anergiat}$  ( $\text{sig}=0,055$ ). Οι μεταβλητές αυτές είναι στατιστικά σημαντικές σε επίπεδο σημαντικότητας  $\alpha=5\%$ .

Επομένως, το βέλτιστο μοντέλο περιλαμβάνει τουλάχιστον τις μεταβλητές  $\text{tomeas3}$  και  $\text{anergiat}$ .

Εκτιμώντας τις αντίστοιχες παλινδρομήσεις για τους παρακάτω συνδυασμούς λαμβάνουμε τις εξής τιμές για τον συντελεστή  $R^2$ , τον  $R^2_{\text{ADJ}}$ , το  $\text{MSE}_p$  και τον  $C_p$ ...

Πίνακας 8.5  
Υπολογισμοί συντελεστών

	$R^2$	$R^2_{\text{ADJ}}$	$\text{MSE}_p$	$C_p$
$X_3, X_4$	0,374	<b>0,348</b>	<b>4,87</b>	<b>3,072</b>
$X_3, X_4, X_1$	0,375	0,335	4,97	4,001
$X_3, X_4, X_2$	0,376	0,336	4,96	4
$X_3, X_4, X_5$	0,373	0,336	4,96	4
$X_3, X_4, X_6$	0,381	0,342	4,92	3,99
$X_3, X_4, X_1, X_2$	0,380	0,326	5,042	4,99
$X_3, X_4, X_5, X_6$	0,382	0,329	5,021	5
$X_3, X_4, X_1, X_5$	0,376	0,321	5,076	5
$X_3, X_4, X_1, X_6$	0,381	0,327	5,030	5

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Centered

Formatted Table

Formatted: Font: Italic

Formatted: Font: Bold, Italic, Font color: Accent 4

Formatted: Font: Italic

Formatted: Font: Bold, Italic, Font color: Accent 4

Formatted: Font: Italic

Formatted: Font: Bold, Italic, Font color: Accent 4

Formatted: Font: Italic

Formatted: Font: Bold, Italic, Font color: Accent 4

Formatted: Font: Italic

Formatted: Font: Bold, Italic, Font color: Accent 4

Formatted: Font: Italic

Formatted: Font: Bold, Italic, Font color: Accent 4

Formatted: Font: Italic

Formatted: Font: Bold, Italic, Font color: Accent 4

Formatted: Font: Italic

Formatted: Font: Bold, Italic, Font color: Accent 4

Formatted: Font: Italic

Formatted: Font: Bold, Italic, Font color: Accent 4

Formatted: Font: Italic

Formatted: Font: Bold, Italic, Font color: Accent 4

Formatted: Font: Italic

Formatted: Font: Bold, Italic, Font color: Accent 4

Formatted: Font: Italic

Formatted: Font: Bold, Italic, Font color: Accent 4

Formatted: Font: Italic

Formatted: Font: Bold, Italic, Font color: Accent 4

$X_3, X_4, X_2, X_5$	0,383	0,330	5,013	4,99
$X_3, X_4, X_2, X_6$	0,384	0,331	5,005	5
$X_3, X_4, X_1, X_2, X_5$	0,384	0,316	5,118	5,99
$X_3, X_4, X_1, X_2, X_6$	0,387	0,319	5,091	6
$X_3, X_4, X_1, X_5, X_6$	0,383	0,314	5,131	5,99
$X_3, X_4, X_2, X_5, X_6$	0,393	0,326	5,040	6
$X_3, X_4, X_1, X_2, X_5, X_6$	0,394	0,311	5,154	6,99

Πίνακας 8.5 Υπολογισμοί συντελεστών

Παρατηρώντας τον πίνακα 8.5, βγάζουμε τα εξής συμπεράσματα:

- I. Με βάση το κριτήριο  $R^2$ , επιλέγουμε ως βέλτιστο μοντέλο αυτό με το μεγαλύτερο συντελεστή προσδιορισμού. Από τους υπολογισμούς φαίνεται ότι το πλήρες μοντέλο έχει τον μεγαλύτερο συντελεστή προσδιορισμού. Άρα,

$$\begin{aligned}
 Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 \\
 &= \beta_0 + \beta_1 \text{urban} + \beta_2 \text{eduleveltotal} + \beta_3 \text{anergiat} \\
 &\quad + \beta_4 \text{tomeas3} + \beta_5 \text{incomecapita} \\
 &\quad + \beta_6 \text{longtermunemployed} + \varepsilon_i \quad (8.2.4.2)
 \end{aligned}$$

- II. Με βάση το κριτήριο  $R^2_{\text{adj}}$ , επιλέγουμε ως βέλτιστο μοντέλο αυτό με το μεγαλύτερο  $R^2_{\text{adj}}$ . Άρα,

$$\begin{aligned}
 Y &= \beta_0 + \beta_3 X_3 + \beta_4 X_4 \\
 &= \beta_0 + \beta_3 \text{anergiat} + \beta_4 \text{tomeas3} + \varepsilon_i \quad (8.2.4.3)
 \end{aligned}$$

- III. Με βάση το κριτήριο  $MSE_p$ , επιλέγουμε ως βέλτιστο μοντέλο εκείνο με την μικρότερη τιμή  $MSE_p$ . Άρα,

$$\begin{aligned}
 Y &= \beta_0 + \beta_3 X_3 + \beta_4 X_4 \\
 &= \beta_0 + \beta_3 \text{anergiat} + \beta_4 \text{tomeas3} + \varepsilon_i \quad (8.2.4.4)
 \end{aligned}$$

- IV. Για να υπολογίσουμε τον συντελεστή Mallows θα χρησιμοποιήσουμε τον τύπο που δίνεται με τα αντίστοιχα  $SSE_p$ ,  $p$ ,  $n$  για το κάθε μοντέλο, καθώς, επίσης και χρησιμοποιώντας το  $MSE$  του πλήρους μοντέλου. Επιλέγουμε, λοιπόν, το μοντέλο στο οποίο πραγματοποιούνται οι εξής προϋποθέσεις:

- $C_p \approx p$
- $p$  μικρό

Άρα,

$$\begin{aligned}
 Y &= \beta_0 + \beta_3 X_3 + \beta_4 X_4 \\
 &= \beta_0 + \beta_3 \text{anergiat} + \beta_4 \text{tomeas3} + \varepsilon_i \quad (8.2.4.5)
 \end{aligned}$$

Τελικό Συμπέρασμα :

Formatted: Font: Italic  
Formatted: Font: Bold, Italic, Font color: Accent 4  
Formatted: Font: Italic  
Formatted: Font: Bold, Italic, Font color: Accent 4  
Formatted: Font: Italic  
Formatted: Font: Bold, Italic, Font color: Accent 4  
Formatted: Font: Italic  
Formatted: Font: Bold, Italic, Font color: Accent 4  
Formatted: Font: Italic  
Formatted: Font: Bold, Italic, Font color: Accent 4  
Formatted: Font: Italic  
Formatted: Font: Bold, Italic, Font color: Accent 4  
Formatted: Font: Italic  
Formatted: Font: Bold, Italic, Font color: Accent 4  
Formatted: Font: Italic  
Formatted: Font: Bold, Italic, Font color: Accent 4

Formatted: English (United States)

Formatted: Font: Italic  
Formatted: Font: Italic  
Formatted: Font: Italic  
Formatted: Font: Italic  
Formatted: Font: Italic  
Formatted: Font: Italic  
Formatted: Font: Italic  
Formatted: Font: Italic

Λαμβάνοντας υπόψιν μας τα παραπάνω κριτήρια καταλήγουμε στο συμπέρασμα ότι το βέλτιστο μοντέλο παλινδρόμησης περιλαμβάνει τις μεταβλητές  $X_3$  και  $X_4$ . Δηλαδή, είναι το

$$Y = \beta_0 + \beta_3 X_3 + \beta_4 X_4 \\ = \beta_0 + \beta_3 \text{anergiat} + \beta_4 \text{tomeas} + \varepsilon_i \quad (8.2.4.6)$$

Στο παράρτημα Α, παρουσιάζονται ενδεικτικά τα αποτελέσματα που πήραμε από το *spss* και από με τα οποία συμπληρώθηκε ο πίνακας 8.5.

### 8.2.5 Backward elimination

Για πιο έγκυρα συμπεράσματα και για να καταλήξουμε πιο σίγουρα στο βέλτιστο μοντέλο μας, ώστε να αποφασίσουμε ποιος παράγοντας επηρεάζει περισσότερο την εσωτερική μετανάστευση, θα χρησιμοποιήσουμε την μέθοδο που είναι γνωστή ως *backward elimination*...

Ξεκινώντας από το πλήρες μοντέλο και διαλέγοντας την κατάλληλη μέθοδο στο *SPSS* παίρνουμε τα εξής αποτελέσματα :

Πίνακας 8.12  
Variables-Entered/Removed

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: English (United States)

Formatted: Font: Italic

Formatted: Underline, Font color: Auto

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Centered

Formatted: English (United States)

Variables Entered/Removed<sup>b</sup>

Model	Variables Entered	Variables Removed	Method
1	long_term_unem ployed, urban, anergiat, tomeas3, incomecapita, eduleveltotal	.	Enter
2	.	urban	Backward (criterion: Probability of F- to-remove $\geq$ ,100).
3	.	incomecapita	Backward (criterion: Probability of F- to-remove $\geq$ ,100).
4	.	eduleveltotal	Backward (criterion: Probability of F- to-remove $\geq$ ,100).
5	.	long_term_unem ployed	Backward (criterion: Probability of F- to-remove $\geq$ ,100).

a. All requested variables entered.

b. Dependent Variable: in\_migration

**Πίνακας 8.12**

Στον πίνακα 8.12 παρουσιάζεται συνοπτικά η διαδικασία εισαγωγής-εξαγωγής των μεταβλητών στο μοντέλο. Ξεκινώντας με το πλήρες μοντέλο και τις έξι μεταβλητές μας, βλέπουμε σταδιακά σε κάθε βήμα να απορρίπτεται και μια μεταβλητή καταλήγοντας σε ένα μοντέλο μόνο με δύο μεταβλητές.

**Πίνακας 8.13**

Formatted Table

Formatted: Indent: Left: 0,11 cm, Right: 0,11 cm, Line spacing: At least 16 pt

Formatted: Greek

Formatted: Centered

Model summary

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,627 <sup>a</sup>	,394	,311	2,2701668
2	,627 <sup>b</sup>	,393	,326	2,2450904
3	,620 <sup>c</sup>	,384	,331	2,2372074
4	,617 <sup>d</sup>	,381	,342	2,2191512
5	,611 <sup>e</sup>	,374	,348	2,2084841

- a. Predictors: (Constant), long\_term\_unemployed, urban, anergiat, tomeas3, incomecapita, eduleveltotal
- b. Predictors: (Constant), long\_term\_unemployed, anergiat, tomeas3, incomecapita, eduleveltotal
- c. Predictors: (Constant), long\_term\_unemployed, anergiat, tomeas3, eduleveltotal
- d. Predictors: (Constant), long\_term\_unemployed, anergiat, tomeas3
- e. Predictors: (Constant), anergiat, tomeas3

Πίνακας 8.13

Στον πίνακα 8.13 απεικονίζονται μερικοί βασικοί δείκτες καλής προσαρμογής. Ο συντελεστής προσδιορισμού  $R^2$  με όλες τις μεταβλητές στο μοντέλο είναι 0,394. Παρατηρούμε ότι όσο αφαιρούνται μεταβλητές από το μοντέλο, μειώνεται και ο συντελεστής προσδιορισμού. Αυτό αποτελεί ένα ανησυχητικό στοιχείο δεδομένου ότι οι υψηλές τιμές των συντελεστών δείχνουν ότι το μοντέλο ερμηνεύει πολύ καλά τις μεταβολές της εξαρτημένης μεταβλητής με την βοήθεια των ανεξαρτήτων. Παρατηρούμε, όμως, ότι η διαφορά του  $R^2$  του πλήρους μοντέλου με το  $R^2$  του βέλτιστου είναι πολύ μικρή.

Το προσαρμοσμένο  $R^2_{adj}$  επιχειρεί να διορθώσει τον συντελεστή προσδιορισμού ώστε να αντανάκλα πιστότερα τη καλή προσαρμογή του μοντέλου στον πληθυσμό. Επομένως, από την τιμή του προσαρμοσμένου  $R^2$ , που παίρνει την μεγαλύτερη τιμή του στο τελικό μοντέλο ( $R^2_{adj}=0,348$ ) του παραπάνω πίνακα καταλαβαίνουμε ότι αυτό είναι και το καταλληλότερο, δηλαδή το μοντέλο με τις μεταβλητές *tomeas3* και *anergiat*.

Τέλος, ο  $R^2_{adj}$  δηλώνει ότι το 34,8% της μεταβλητότητας της εξαρτημένης μεταβλητής που υπάρχει ερμηνεύεται με το μοντέλο. Ποσοστό μάλλον χαμηλό.

Πίνακας 8.14  
ANOVA

Formatted: English (United States)

Formatted Table

Formatted: Centered

Formatted: Centered, Tab stops: Not at 7,32 cm + 14,65 cm

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Font: Italic

Formatted: Font: Italic, Greek

Formatted: Greek

Formatted: Font: Italic

Formatted: Font: Italic, Greek

Formatted: Font: Italic

Formatted: Font: Italic, Greek

Formatted: Greek

Formatted: Font: Italic

Formatted: Font: Italic, Greek

Formatted: Font: Italic

Formatted: Greek

Formatted: Font: Italic

Formatted: Font: Italic, Greek

Formatted: Greek

Formatted: Font: Italic

Formatted: Font: Italic, Greek

Formatted: Font: Italic

Formatted: Greek

Formatted: Font: Italic

Formatted: Font: Italic, Greek

Formatted: Font: Italic

Formatted: Centered

Formatted: Font: Italic

Formatted: Font: Italic, English (United States)

ANOVA<sup>f</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	147,177	6	24,530	4,760	,001 <sup>a</sup>
	Residual	226,761	44	5,154		
	Total	373,938	50			
2	Regression	147,119	5	29,424	5,838	,000 <sup>b</sup>
	Residual	226,819	45	5,040		
	Total	373,938	50			
3	Regression	143,704	4	35,926	7,178	,000 <sup>c</sup>
	Residual	230,234	46	5,005		
	Total	373,938	50			
4	Regression	142,480	3	47,493	9,644	,000 <sup>d</sup>
	Residual	231,458	47	4,925		
	Total	373,938	50			
5	Regression	139,823	2	69,911	14,334	,000 <sup>e</sup>
	Residual	234,115	48	4,877		
	Total	373,938	50			

a. Predictors: (Constant), long\_term\_unemployed, urban, anergiat, tomeas3, incomecapita, edulevelttotal

b. Predictors: (Constant), long\_term\_unemployed, anergiat, tomeas3, incomecapita, edulevelttotal

c. Predictors: (Constant), long\_term\_unemployed, anergiat, tomeas3, edulevelttotal

d. Predictors: (Constant), long\_term\_unemployed, anergiat, tomeas3

e. Predictors: (Constant), anergiat, tomeas3

f. Dependent Variable: in\_migration

#### Πίνακας 8.14

Η βασιμότητα ή όχι της υπόθεσης της γραμμικής παλινδρόμησης μπορεί να ελεγχθεί από τον πίνακα διακύμανσης (ANOVA). Η συνολική μεταβλητότητα των δεδομένων

χωρίζεται και αποδίδεται εν μέρει στο μοντέλο και εν μέρει στα λάθη (κατάλοιπα). Στον πίνακα 8.14 παρατηρούμε ότι το άθροισμα των τετραγώνων από την παλινδρόμηση είναι μικρότερο από αυτό των καταλοίπων. Το σύνολο της διακύμανσης του δείγματος  $Total=377,938$  είναι άθροισμα της διακύμανσης της παλινδρόμησης  $Regression=139,823$  και της διακύμανσης του λάθους  $Residual=234,115$ . Όσο καλύτερο είναι το μοντέλο της παλινδρόμησης τόσο μεγαλύτερο το μέρος της συνολικής διακύμανσης του δείγματος που επεξηγεί.

Στη συνέχεια παρουσιάζονται οι βαθμοί ελευθερίας ( $df$ ) και τα μέσα τετραγωνικά αθροίσματα ( $Mean Squares$ ). Το άθροισμα αυτό υπολογίζεται διαιρώντας το άθροισμα των τετραγώνων με τους αντίστοιχους βαθμούς ελευθερίας.

Στον πίνακα 8.14, απεικονίζεται ένας συνολικός έλεγχος για την σημαντικότητα του μοντέλου της παλινδρόμησης. Ο έλεγχος βασίζεται στη στατιστική συνάρτηση  $F$  και

Formatted Table

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Greek

Formatted: Font: Italic

Formatted: Greek

ελέγχει την υπόθεση ότι οι συντελεστές των ανεξαρτήτων μεταβλητών που συμμετέχουν στο μοντέλο είναι ταυτόχρονα μηδέν. Παρατηρούμε ότι σε όλα τα μοντέλα η τιμή του *sig* είναι πολύ μικρή (*sig*=0.000). Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% , απορρίπτουμε την αρχική μας υπόθεση. Δηλαδή, συμπεραίνουμε ότι υπάρχει στατιστικά σημαντική σχέση (πρόβλεψη) της εξαρτημένης μεταβλητής με τις ανεξάρτητες (και στα πέντε μοντέλα που δημιουργήθηκαν κατά τη διάρκεια της μεθόδου).

Πίνακας 8.15

*Συντελεστές Παλινδρόμησης και έλεγχος σημαντικότητας των μεταβλητών*

Formatted: Font: Italic, Greek

Coefficients <sup>a</sup>		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
Model		B	Std. Error	Beta		
1	(Constant)	-2,334	6,441		-,362	,719
	urban	-,003	,032	-,019	-,107	,916
	eduleveltotal	1,025	1,145	,214	,895	,376
	anergiat	-,230	,117	-,245	-1,972	,055
	tomeas3	,169	,066	,499	2,577	,013
	inmecapita	-,001	,001	-,155	-,673	,504
	long_term_unemployed	-,019	,023	-,106	-,828	,412
2	(Constant)	-2,175	6,196		-,351	,727
	eduleveltotal	,994	1,095	,208	,907	,369
	anergiat	-,231	,115	-,247	-2,021	,049
	tomeas3	,171	,063	,504	2,724	,009
	inmecapita	-,001	,001	-,166	-,823	,415
	long_term_unemployed	-,020	,023	-,108	-,863	,392
3	(Constant)	,447	5,296		,084	,933
	eduleveltotal	,413	,835	,086	,494	,623
	anergiat	-,227	,114	-,242	-1,995	,052
	tomeas3	,165	,062	,488	2,661	,011
	long_term_unemployed	-,018	,023	-,099	-,801	,427
4	(Constant)	2,670	2,775		,962	,341
	anergiat	-,242	,109	-,258	-2,212	,032
	tomeas3	,188	,041	,555	4,557	,000
	long_term_unemployed	-,016	,022	-,089	-,735	,466
5	(Constant)	1,464	2,226		,658	,514
	anergiat	-,252	,108	-,269	-2,335	,024

tomeas3	.198	.039	.583	5.069	.000
---------	------	------	------	-------	------

a. Dependent Variable: in\_migration

Στον πίνακα 8.15 αρχικά παρουσιάζονται οι συντελεστές για τις ανεξάρτητες μεταβλητές για όλα τα μοντέλα διαδοχικά. Για το βέλτιστο μοντέλο έχουμε:

- ❖  $\beta_0 = 1,464$
- ❖  $\beta_3 = -0,252$
- ❖  $\beta_4 = 0,198$

Συνεπώς, χρησιμοποιώντας αυτούς τους συντελεστές μπορούμε να γράψουμε την εκτιμώμενη εξίσωση παλινδρόμησης :

$$\begin{aligned} \text{in migration} &= 1,464 - 0,252 * \text{anergiat} + 0,198 \\ &* \text{tomeas3} \quad (8.2.5.2) \end{aligned}$$

Η ερμηνεία των συντελεστών αποτυπώνεται ως εξής :

- ✚ Για κάθε αύξηση στον τριτογενή τομέα οικονομικής δραστηριότητας κατά 1%, ο δείκτης της εσωτερικής μετανάστευσης αυξάνει κατά 0,198.
- ✚ Για κάθε αύξηση της συνολικής ανεργίας κατά 1%, ο δείκτης εσωτερικής μετανάστευσης μειώνεται κατά 0,252.

Στον πίνακα 8.15 παρουσιάζονται οι σταθμισμένοι ή αλλιώς τυποποιημένοι (*standardized coefficients*).

Οι συντελεστές αυτοί δείχνουν την ιδιαίτερη επίδραση που ασκεί κάθε ανεξάρτητη μεταβλητή στην εξαρτημένη. Παρατηρούμε, λοιπόν, ότι κατά απόλυτη τιμή ο τριτογενής τομέας οικονομικής δραστηριότητας (*tomeas3*) ασκεί σημαντικότερη επίδραση στην εσωτερική μετανάστευση (*in\_migration*) από την συνολική ανεργία (*anergiat*).

Επίσης, παρουσιάζονται και τα αποτελέσματα του t ελέγχου για τους συντελεστές παλινδρόμησης.

Τέλος, από τις τιμές του sig λαμβάνουμε την τιμή σημαντικότητας βάση της οποίας επιβεβαιώνεται ή όχι η διατήρηση της μεταβλητής στο μοντέλο. Ο έλεγχος που κάνουμε έχει τις εξής υποθέσεις :

$H_0$ : οι συντελεστές  $\beta$  είναι μηδενικοί  
 $H_1$ : οι συντελεστές  $\beta$  δεν είναι μηδενικοί.

- ✚ Για την μεταβλητή *anergiat* φαίνεται ότι  $sig = 0,024 < 0,05$ . Άρα, σε επίπεδο σημαντικότητας 5% απορρίπτουμε την αρχική υπόθεση. Αυτό σημαίνει ότι οι συντελεστές  $\beta$  δεν είναι μηδενικοί και επομένως, η μεταβλητή είναι στατιστικά σημαντική. (Υπάρχει, δηλαδή, μια στατιστικά σημαντική πρόβλεψη της εσωτερικής μετανάστευσης από την συνολική ανεργία).
- ✚ Για την μεταβλητή *tomeas3* φαίνεται ότι  $sig = 0,000 < 0,05$ . Άρα σε επίπεδο σημαντικότητας 5% απορρίπτουμε την αρχική υπόθεση. Δηλαδή, υπάρχει μια στατιστικά σημαντική πρόβλεψη της εσωτερικής μετανάστευσης από τον τριτογενή τομέα οικονομικής δραστηριότητας.

Πίνακας 8.16



*Excluded variables*

**Excluded Variables<sup>e</sup>**

Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
					Tolerance
2 urban	-,019 <sup>a</sup>	-,107	,916	-,016	,456
3 urban	-,072 <sup>b</sup>	-,473	,638	-,070	,580
incomecapita	-,166 <sup>b</sup>	-,823	,415	-,122	,332
4 urban	-,014 <sup>c</sup>	-,110	,913	-,016	,858
incomecapita	-,048 <sup>c</sup>	-,312	,756	-,046	,568
eduleveltotal	,086 <sup>c</sup>	,494	,623	,073	,440
5 Urban	-,029 <sup>d</sup>	-,239	,812	-,035	,886
incomecapita	-,053 <sup>d</sup>	-,347	,730	-,051	,569
eduleveltotal	,063 <sup>d</sup>	,366	,716	,053	,452
long_term_unemployed	-,089 <sup>d</sup>	-,735	,466	-,107	,895

a. Predictors in the Model: (Constant), long\_term\_unemployed, anergiat, tomeas3, incomecapita, eduleveltotal

b. Predictors in the Model: (Constant), long\_term\_unemployed, anergiat, tomeas3, eduleveltotal

c. Predictors in the Model: (Constant), long\_term\_unemployed, anergiat, tomeas3

d. Predictors in the Model: (Constant), anergiat, tomeas3

e. Dependent Variable: in\_migration

Στον πίνακα 8.16 παρουσιάζονται οι αντίστοιχες εκτιμήσεις για τις μεταβλητές που αποκλείστηκαν από το μοντέλο.

Βασική παρατήρηση που μπορούμε να κάνουμε είναι ότι οι τιμές των επιπέδων σημαντικότητας που παρουσιάζονται είναι πολύ υψηλές σε όλες τις μεταβλητές. Αυτό σημαίνει ότι σε ε.σ. 5% , δεχόμαστε την αρχική υπόθεση και, συνεπώς, οι μεταβλητές αυτές δεν είναι στατιστικά σημαντικές.

**Τελικό Συμπέρασμα :**

Από τον πίνακα 8.15 καταλήγουμε στο συμπέρασμα ότι το βέλτιστο μοντέλο παλινδρόμησης περιλαμβάνει τις μεταβλητές  $X_3$  και  $X_4$ . Δηλαδή , είναι το

$$Y = \beta_0 + \beta_3 X_3 + \beta_4 X_4 = \beta_0 + \beta_3 \text{anergiat} + \beta_4 \text{tomeas3}$$

**8.2.6 Ανάλυση σφαλμάτων**

Όταν προσαρμόζουμε μοντέλα σε δεδομένα , τα κατάλοιπα παίζουν πολύ σημαντικό ρόλο. Εξετάζοντας την κατανομή των υπολειμματικών τιμών και τη σχέση τους με τις άλλες μεταβλητές, μπορούμε να εντοπίσουμε παρεκκλίσεις από τις παραδοχές της παλινδρόμησης.

Μια υπολειμματική τιμή (ή κατάλοιπο) είναι αυτό που απομένει μετά την προσαρμογή του μοντέλου. Είναι η διαφορά μεταξύ της παρατηρούμενης τιμής της εξαρτημένης μεταβλητής και της τιμής που προβλέπεται από την ευθεία παλινδρόμησης. Για να ικανοποιούνται οι απαιτούμενες παραδοχές για μια ανάλυση παλινδρόμησης, θα πρέπει τα κατάλοιπα να έχουν τα παρακάτω χαρακτηριστικά :

- ✓ Να είναι κατά προσέγγιση κανονικά κατανομημένα.
- ✓ Η διακύμανση τους να είναι ίδια για όλες τις τιμές της ανεξάρτητης μεταβλητής.
- ✓ Να μην παρουσιάζουν κανένα μοτίβο όταν σχεδιάζονται συναρτήσει των προβλεπόμενων τιμών.
- ✓ Διαδοχικές υπολειμματικές τιμές να είναι κατά προσέγγιση ανεξάρτητες.

Πρώτα , πρέπει να εκτιμήσουμε τα σφάλματα μας. Στον πίνακα 8.17 παρουσιάζονται κάποια βασικά στοιχεία για τα σφάλματα. (π.χ. *minimum* τιμή)

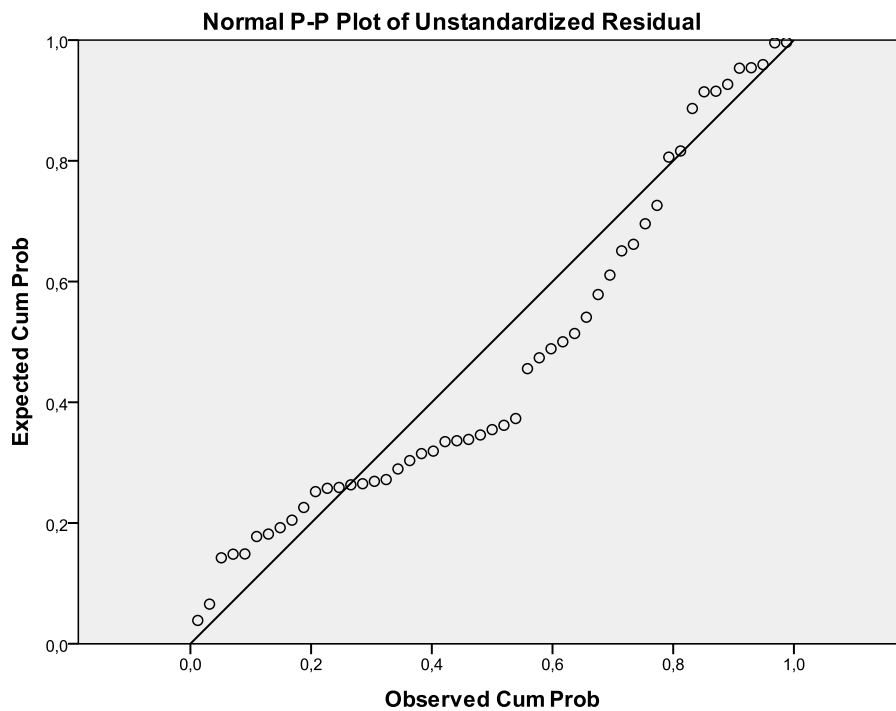
Πίνακας 8.17  
*Residuals statistics*  
**Residuals Statistics<sup>a</sup>**

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	4,909192	11,785313	8,218041	1,6722613	51
Residual	-3,8236744	5,7993622	,0000000	2,1638636	51
Std. Predicted Value	-1,979	2,133	,000	1,000	51
Std. Residual	-1,731	2,626	,000	,980	51

a. Dependent Variable: in\_migration

✓ **ΚΑΝΟΝΙΚΟΤΗΤΑ ΣΦΑΛΜΑΤΩΝ**

Η οπτική επιβεβαίωση της τυχαίας κίνησης των σφαλμάτων γίνεται μέσω ενός διαγράμματος *P-P plot* . Το διάγραμμα αυτό παρουσιάζεται στη συνέχεια



Διάγραμμα 8.32  
P-P plot σφαλμάτων

«Όσο πιο κοντά τα σημεία στην ευθεία τόσο πιο πολύ οδηγούμαστε σε κανονικότητα»

Στο διάγραμμα 8.32 παρατηρείται η τάση να τείνει η κατανομή των σφαλμάτων προς την κανονική χωρίς όμως να είμαστε απόλυτα σίγουροι.

Ο κατάλληλος έλεγχος που γίνεται για την κανονικότητα των σφαλμάτων είναι ο έλεγχος Kolmogorov-Smirnov.

Η υπόθεση που ικανοποιεί τον έλεγχο αυτόν είναι η εξής :

$H_0$ : ύπαρξη κανονικότητας σφαλμάτων

$H_1$ : μη ύπαρξη κανονικότητας σφαλμάτων

Πίνακας 8.18  
Kolmogorov-Smirnov test

One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
N		51
Normal Parameters <sup>a,b</sup>	Mean	,0000000
	Std. Deviation	2,16386364
Most Extreme Differences	Absolute	,176
	Positive	,176
	Negative	-,103
Kolmogorov-Smirnov Z		1,256
Asymp. Sig. (2-tailed)		<b>,085</b>

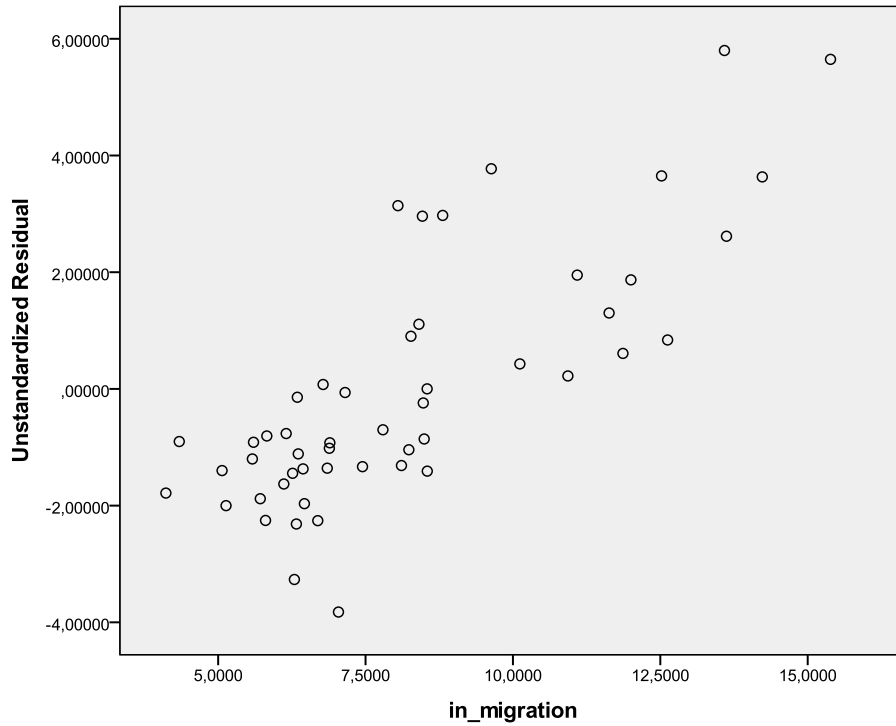
a. Test distribution is Normal.

b. Calculated from data.

Στον πίνακα 8.18 παρουσιάζεται ο έλεγχος Kolmogorov-Smirnov .  
Μας δίνεται ότι  $sig = 0,085 > 0,05$ . Συνεπώς , σε επίπεδο σημαντικότητας 5% , αποδεχόμαστε την αρχική μας υπόθεση περί κανονικότητας των σφαλμάτων.

✓ **ΑΝΕΞΑΡΤΗΣΙΑ ΣΦΑΛΜΑΤΩΝ**

Για να απεικονίσουμε την ανεξαρτησία ή μη των καταλοίπων , θα σχεδιάσουμε ένα διάγραμμα διασποράς (*scatterplot*). Αυτό που μας ενδιαφέρει είναι να μην υπάρχει συστηματική διάταξη των στοιχείων. Αυτό διαφαίνεται στο Διάγραμμα 8.33



Διάγραμμα 8.33  
Scatterplot καταλοίπων

Ο κατάλληλος έλεγχος για την ανεξαρτησία των σφαλμάτων είναι το τεστ Ρούν (Runs test).

Ο έλεγχος εξετάζει την υπόθεση ότι:

$H_0$ : ύπαρξη ανεξαρτησίας σφαλμάτων  
 $H_1$  μη ύπαρξη ανεξαρτησίας σφαλμάτων

Πίνακας 8.19

*Runs test*

**Runs Test**

	Unstandardized Residual
Test Value <sup>a</sup>	-.80629
Cases < Test Value	25
Cases >= Test Value	26
Total Cases	51
Number of Runs	25
Z	-.422
Asymp. Sig. (2-tailed)	.673

a. Median

Από τον πίνακα 8.19, λαμβάνουμε ότι  $\text{sig} = 0.673 > 0.05$ . Συνεπώς, σε ε.σ. 5% αποδεχόμαστε την αρχική υπόθεση περί ανεξαρτησίας των σφαλμάτων.

## ΚΕΦΑΛΑΙΟ 9 : ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα εργασία , διερευνώνται οι παράγοντες εκείνοι , οι οποίοι επηρεάζουν την εσωτερική μετανάστευση στην Ελλάδα κατά τη διάρκεια της πενταετίας 1996-2001. Το δείγμα της έρευνας προέρχεται από την απογραφή που διεξήχθη το 2001 στην Ελλάδα. Τα δεδομένα που χρησιμοποιήθηκαν είναι ομαδοποιημένα ανά νομό. Ορισμένες μετρήσεις προήλθαν από στοιχεία των ληξιαρχικών καταγραφών της χώρας, και βοήθησαν στον υπολογισμό κάποιων δημογραφικών μεταναστευτικών δεικτών.

Η αναλογία του αστικού πληθυσμού σε σχέση με τον συνολικό πληθυσμό της χώρας, το συνολικό ποσοστό των ανέργων σε ολόκληρη την Ελλάδα , το συνολικό επίπεδο εκπαίδευσης , ο τριτογενής τομέας οικονομικής δραστηριότητας , το κατά κεφαλήν εισόδημα , η συμμετοχή στην αγορά εργασίας και η μακροχρόνια ανεργία αποτελούν τους παράγοντες εκείνους που κρίθηκε σκόπιμο να μελετηθούν σε βάθος , στηριζόμενοι πάντα στην σημαντικότητα τους , όπως αυτή προκύπτει από την βιβλιογραφική ανασκόπηση. Ας σημειωθεί ότι οι παραπάνω παράγοντες αναφέρονται στο σύνολο. Δεν έχουν ομαδοποιηθεί ανά φύλο. Αυτό συνέβη διότι το μέτρο της εσωτερικής μετανάστευσης (μεταβλητή *in\_migration*) είναι υπολογισμένο συνολικά και όχι διαχωρισμένο ανά φύλο. Έτσι , φυσικά, τα συμπεράσματα που θα εξαχθούν θα είναι πιο σωστά και πιο αξιόπιστα. Το μέτρο της εσωτερικής μετανάστευσης αποτελεί το πηλίκο των εγκατασταθέντων από άλλο δήμο προς το σύνολο των ατόμων του Δήμου(επί 100).

Ξεκινώντας , την παρούσα έρευνα, κρίθηκε απαραίτητο να βρεθούν κάποιες συσχετίσεις , που ενδεχομένως υπάρχουν, μεταξύ των μεταβλητών.

Αναφορικά με την μεταβλητή *urban* , η οποία αναφέρεται στο ποσοστό του αστικού πληθυσμού, παρατηρήθηκε ότι παρουσιάζει μια θετική συσχέτιση με την εσωτερική μετανάστευση . Γεγονός, απόλυτα φυσιολογικό , εφόσον τα τελευταία έτη παρατηρούμε μια τάση του πληθυσμού της χώρας να συγκεντρώνεται στα αστικά κέντρα. Παρ' όλα αυτά, η συσχέτιση αυτή δεν ήταν στατιστικά σημαντική. Επίσης, το ποσοστό του αστικού πληθυσμού παρουσίασε μια αρνητική συσχέτιση με το ποσοστό της ανεργίας. Αυτό δείχνει ότι, στους νομούς όπου υπάρχουν αστικά κέντρα, η ανεργία είναι σε χαμηλότερα επίπεδα από ότι στους υπόλοιπους νομούς όπου υπάρχουν περισσότερες επαρχιακές περιοχές.

Παρατηρώντας στη συνέχεια, το μέσο αριθμό των ετών των συνολικών επιτυχημένων εκπληρωμένων εκπαιδευτικών δραστηριοτήτων του πληθυσμού της χώρας, αντιλαμβανόμαστε την ισχυρή συσχέτιση που παρουσιάζει με την εσωτερική μετανάστευση. Η συσχέτιση αυτή είναι στατιστικά σημαντική και επομένως, η εκπαίδευση θα θεωρούνταν ένας πολύ σημαντικός παράγοντας για την μετακίνηση του πληθυσμού στον ελλαδικό χώρο. Έτσι θα ήταν απόλυτα φυσιολογικό να συμπεριληφθεί στην έρευνα , ώστε να εξετασθεί διεξοδικά. Παρ' όλα αυτά, η εκπαίδευση δεν θα συμπεριληφθεί στην παρούσα έρευνα. Το γεγονός αυτό, κρίθηκε σκόπιμο διότι , εκτός από την ισχυρή συσχέτιση που παρατηρήθηκε με την μετανάστευση, εντύπωση προκάλεσαν και οι ισχυρές συσχετίσεις που παρουσίασε

και άλλους παράγοντες. Πιο συγκεκριμένα, φάνηκε ότι η εκπαίδευση συσχετίζεται πολύ έντονα και με τον αστικό πληθυσμό και με τον τριτογενή τομέα οικονομικής δραστηριότητας, αλλά και με το κατά κεφαλήν εισόδημα. Όλα αυτά, βέβαια, είναι αναμενόμενα και απολύτως φυσιολογικά, εφόσον όσα περισσότερα χρόνια έχει εκπαιδευτεί ένα άτομο, τόσο μεγαλύτερο εισόδημα ίσως λαμβάνει αλλά και τόσο περισσότερο θα απορροφάται από τον τριτογενή τομέα οικονομικής δραστηριότητας και τα αστικά κέντρα. Επειδή, όμως, αυτή η ισχυρή επίδραση του συγκεκριμένου παράγοντα, ενδεχομένως να επηρεάσει τα αποτελέσματα της έρευνας και να μην προκύψει σαφή εικόνα, κρίθηκε σκόπιμο να εξαιρεθεί στην παρούσα φάση.

Ένας ακόμη παράγοντας που παρουσιάζει μια θετική, πολύ ισχυρή συσχέτιση με την εσωτερική μετανάστευση είναι ο τριτογενής τομέας οικονομικής δραστηριότητας, ή αλλιώς βιομηχανικός τομέας. Ο τομέας αυτός περιλαμβάνει τις διάφορες υπηρεσίες, ιδιωτικές και κρατικές, που χρησιμοποιούνται από τους τελικούς καταναλωτές ή βοηθούν στην παραγωγή των προϊόντων των δύο άλλων τομέων της οικονομίας. Σε αυτό τον τομέα οικονομικής δραστηριότητας ανήκουν οι μεταφορές, το εμπόριο, οι τράπεζες, οι ασφαλιστικές επιχειρήσεις, οι ελεύθεροι επαγγελματίες (δικηγόροι, γιατροί, μηχανικοί), οι απασχολούμενοι στον κλάδο αναψυχής (ηθοποιοί, τραγουδιστές) και στον τουρισμό. Κατ' απόλυτη τιμή, ο παράγοντας αυτός έχει τον μεγαλύτερο συντελεστή Pearson άρα και την μεγαλύτερη εξάρτηση με την μετανάστευση. Για αυτό τον λόγο κρίθηκε απαραίτητο, ο παράγοντας αυτός να μελετηθεί διεξοδικά στην συνέχεια της παρούσας έρευνας.

Τέλος, ολοκληρώνοντας με τους παράγοντες που παρουσίασαν μια θετική επίδραση στην εσωτερική μετανάστευση, θα πρέπει να συμπεριλάβουμε και το κατά κεφαλήν εισόδημα. Ο παράγοντας αυτός παρουσίασε σημαντική επίδραση και με τον τριτογενή τομέα, αλλά, όπως αναφέρθηκε και προηγουμένως είναι λογικό, αφού ο βιομηχανικός τομέας αποτελεί μια ανάπτυξη στον τομέα παραγωγής και μια μεγάλη ευκαιρία για περισσότερες θέσεις εργασίας.

Με την πρώτη ανάλυση που έγινε για τις επιδράσεις μεταξύ των παραγόντων εντοπίστηκαν και κάποιοι οι οποίοι επιδρούν αρνητικά στην εσωτερική μετανάστευση. Αυτοί οι παράγοντες έχουν ένα κοινό χαρακτηριστικό. Και οι τρεις αναφέρονται κατά έναν τρόπο στην ανεργία. Ο ένας αναφέρεται καθαρά στο συνολικό ποσοστό της ανεργίας σε κάθε νομό της Ελλάδος (*anergiat*) και οι άλλοι δυο αφορούν στο ποσοστό συμμετοχής των ατόμων του κάθε νομού στην αγορά εργασίας (*participation*) αλλά και στην μακροχρόνια ανεργία (*long\_term\_unemployed*). Εφόσον και οι τρεις μεταβλητές αναφέρονται ουσιαστικά σε έναν κοινό παράγοντα, θα συνεχίσουμε την έρευνα με τις δυο που παρουσιάζουν την μεγαλύτερη (κατ' απόλυτη τιμή) συσχέτιση με την εσωτερική μετανάστευση. Αυτές είναι οι *anergiat* και *long\_term\_unemployed*.

Παίρνοντας με τα παραπάνω μία πρώτη εικόνα για το ποιους παράγοντες επηρεάζουν την εσωτερική μετανάστευση, δημιουργήθηκε το πλήρες μοντέλο. Τρέχοντας το μοντέλο αυτό, παρατηρήθηκε ότι μόνο δυο παράγοντες είναι στατιστικά σημαντικοί. Παρατηρήθηκε ότι, ο τριτογενής τομέας οικονομικής δραστηριότητας και το



συνολικό ποσοστό της ανεργίας αποτελούν δυο μεταβλητές στατιστικά σημαντικές. Σε αντίθεση με τις υπόλοιπες, οι οποίες κρίθηκαν στατιστικά μη σημαντικές.

Για να αποφασιστεί ποιο μοντέλο είναι το βέλτιστο, χρησιμοποιήθηκε η μέθοδος «backward elimination». Τα αποτελέσματα της μεθόδου αυτής επιβεβαίωσαν την αρχική υπόθεση, επιλέγοντας ως βέλτιστο μοντέλο εκείνο που περιελάμβανε τους δυο παράγοντες που αναφέρθηκαν προηγουμένως (ανεργία και τριτογενής τομέας οικονομικής δραστηριότητας).

Στη συνέχεια για να πιστοποιηθεί ότι οι παραδοχές της πολλαπλής παλινδρόμησης δεν παραβιάζονται, πραγματοποιήθηκε μια ανάλυση των σφαλμάτων, η οποία επιβεβαίωσε ότι όλες οι παραδοχές τηρούνται.

Το γεγονός που παρατηρήθηκε και προβληματίσε ιδιαίτερα ήταν ότι ο συντελεστής προσδιορισμού  $R^2$  του τελικού μοντέλου ήταν αρκετά χαμηλός. Γνωρίζοντας ότι, οι υψηλές τιμές του συντελεστή δείχνουν ότι το μοντέλο ερμηνεύει πολύ καλά τις μεταβολές της εξαρτημένης μεταβλητής με την βοήθεια των ανεξάρτητων, μπορούμε να θεωρηθεί ότι το τελικό μοντέλο ήταν το βέλτιστο στην έρευνα που διεξήχθη αλλά δεν μπορεί να θεωρηθεί και το καλύτερο για την ερμηνεία της εσωτερικής μετανάστευσης. Το γεγονός αυτό μπορεί να οφείλεται σε ενδεχόμενα σφάλματα στα δεδομένα τα οποία δόθηκαν και δεν μπορούν να εντοπιστούν από τον ερευνητή, ή ακόμα μπορεί και να οφείλεται στο ότι θα μπορούσε να χρησιμοποιηθεί κάποιο άλλο μοντέλο πέραν του γραμμικού.

Καταλήγοντας, θα πρέπει να σημειωθεί ότι η παρούσα εργασία πραγματοποιήθηκε με σκοπό την διερεύνηση των δημογραφικών παραγόντων που επηρεάζουν την εσωτερική μετανάστευση στην Ελλάδα. Στην προσπάθεια αυτή, κρίθηκε σκόπιμο, να μελετηθούν ορισμένοι παράγοντες. Ωστόσο, υπάρχουν παράγοντες οι οποίοι δεν αναλύθηκαν στην παρούσα μελέτη αλλά κρίνεται σκόπιμο να ερευνηθούν σε μελλοντικές έρευνες, αφού μπορούν να προσφέρουν μια προσέγγιση, από μια άλλη οπτική γωνία, που ίσως δώσει μια καλύτερη εικόνα της εσωτερικής μετανάστευσης στον ελλαδικό χώρο.

Πανεπιστήμιο Πειραιώς

## ΒΙΒΛΙΟΓΡΑΦΙΑ

### ΕΛΛΗΝΙΚΗ

**Βαλαώρας, Β.** (1973). *Το δημογραφικό πρόβλημα της Ελλάδος. Ο ανθρώπινος παράγον*, Αθήνα

**Ευαγγελάρας, Χ.** (2008), Πανεπιστημιακές σημειώσεις στο μάθημα *Ανάλυση δεδομένων με την χρήση Στατιστικών πακέτων* του ΠΜΣ Εφαρμοσμένη Στατιστική, Πανεπιστήμιο Πειραιώς, Αθήνα.

**Ευαγγελάρας, Χ. και Κούτρας, Μ.** (2010), *Ανάλυση Παλινδρόμησης*, εκδόσεις Σταμούλη, Αθήνα.

**Καλογήρου Σ. και Κατσαφάδος Π.,** *Ο ρόλος του κλίματος στην μεταναστευτική απόφαση*, Αθήνα.

**Κοτζαμάνης, Β.** (2010), *Δημογραφικά νέα, Νο11*, Περιοδική έκδοση του εργαστηρίου δημογραφικών και κοινωνικών αναλύσεων του Πανεπιστημίου Θεσσαλίας, Θεσσαλία.

**Κοτζαμάνης, Β.** (1990), *Η κινητικότητα του αγροτικού πληθυσμού στην δεκαετία 1940-1950 και η αναδιάρθρωση του κοινωνικού-δημογραφικού χάρτη της μεταπολεμικής Ελλάδας.*

**Κούτρας, Μ. και Δαμιανού, Χ.** (2003), *Εισαγωγή στη Στατιστική*, Εκδόσεις Συμμετρία, Αθήνα.

**Κούτρας, Μ.** (2011), Πανεπιστημιακές σημειώσεις στο μάθημα *Ανάλυση παλινδρόμησης και Διακρίμανσης* του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς, Αθήνα.

**Κυριαζή – Αλλισον, Ε.** (1998), *Εσωτερική Μετανάστευση στην Ελλάδα το 1990: Τάσεις - προβληματισμοί – προοπτικές*, Επιθεώρηση Κοινωνικών Ερευνών.

**Μπούτσικας, Μ.** (2007), Πανεπιστημιακές σημειώσεις στο μάθημα *Στατιστικά Προγράμματα* του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς, Αθήνα.

**Τριανταφυλλίδου, Α.** (2005), *Ελληνική Μεταναστευτική Πολιτική. Προβλήματα και Κατευθύνσεις*, Ελληνικό ίδρυμα εσωτερικής και εξωτερικής Πολιτικής, Αθήνα.

**Τσίμπος, Κ. και Γεωργιάδης, Φ.** (1999), *Περιγραφική και Διερευνητική Στατιστική ανάλυση δεδομένων*, Εκδόσεις Σταμούλη, Αθήνα.

**Τσίμπος, Κ. και Παπαδάκης, Μ** (2004), *Δημογραφική Ανάλυση*, Εκδόσεις Σταμούλη, Αθήνα.

**Τσίμπος, Κ.** (2008), *Εισαγωγή στην Πληθυσμιακή Γεωγραφία*, Εκδόσεις Σταμούλη, Αθήνα.

**Φράγκος, Χ.** (2004), *Μεθοδολογία έρευνας Αγοράς και ανάλυση δεδομένων με τη χρήση του στατιστικού πακέτου SPSS for Windows*, Interbooks, Αθήνα.

### ΞΕΝΟΓΛΩΣΣΗ

**Emke-Poulopoulou, I.** (1991), *Immigrants and Refugees in Greece. Statistical Evaluation and Causes of Entry in the Country.*

**Fakiolas, R.** (2000), *Migration and unregistered labour in the Greek economy*, in King, R. and Lazaridis, G. and Tsardanidis, Ch. eds., *Eldorado or Fortress, Migration in Southern Europe*, London: Macmillan.

**Kostaki, A. and Kotzamanis and Stathakis, D.** *The spatial clustering of immigrant population in Greece*, Athens.

**Lazaridis, G.** (1996), *Immigration to Greece: a critical evaluation of Greek policy*, New Community.

**Lianos, T.** (2001), *Illegal Immigrants in Greece and their choice of destination*, International Migration.

### ΙΣΤΟΤΟΠΟΙ

[www.demography-lab.prd.uth](http://www.demography-lab.prd.uth)

[www.statistics.gr](http://www.statistics.gr)

<http://www.iom.int/cms/en/sites/iom/home/where-we-work/europa/european-economic-area/greece.html>

<http://el.wikipedia.org/wiki/%CE%9C%CE%B5%CF%84%CE%B1%CE%BD%CE%AC%CF%83%CF%84%CE%B5%CF%85%CF%83%CE%B7>

[http://geogr.eduportal.gr/senaria/population\\_gr/info\\_meta\\_gr.htm](http://geogr.eduportal.gr/senaria/population_gr/info_meta_gr.htm)

## ΠΑΡΑΡΤΗΜΑ Α

Στη συνέχεια αναφέρονται, ενδεικτικά ορισμένοι πίνακες, από τους οποίους λάβαμε τα αποτελέσματα του πίνακα 8.5.

- ο Παλινδρόμηση με τις μεταβλητές  $X_3$  και  $X_4$ .

**Πίνακας 10.1**  
Μοντέλο παλινδρόμησης με τις  $X_3$  και  $X_4$

Model	R	R Square	Adjusted Square	R	Std. Error of the Estimate
1	.611 <sup>a</sup>	.374	.348		2,2084841

a. Predictors: (Constant), tomas3, anergiat

**Πίνακας 8.6** Μοντέλο παλινδρόμησης με τις  $X_3$  και  $X_4$

**Πίνακας 10.2**  
ANOVA για το μοντέλο με τις  $X_3$  και  $X_4$

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	139,823	2	69,911	14,334	,000 <sup>a</sup>
	Residual	234,115	48	4,877		
	Total	373,938	50			

a. Predictors: (Constant), tomas3, anergiat

b. Dependent Variable: in\_migration

**Πίνακας 8.7** ANOVA για το μοντέλο με τις  $X_3$  και  $X_4$

- ο Παλινδρόμηση με τις μεταβλητές  $X_1$ ,  $X_3$  και  $X_4$ .

**Πίνακας 10.3**  
Μοντέλο παλινδρόμησης με τις  $X_1$ ,  $X_3$  και  $X_4$

Model	R	R Square	Adjusted Square	R	Std. Error of the Estimate

Formatted: Centered

Formatted: Font: Italic

Formatted: Font: Italic

Formatted Table

Formatted: Indent: Left: 0 cm, Right: 0 cm, Line spacing: At least 20 pt

Formatted: Centered

Formatted: Font: Italic

Formatted: Greek

Formatted: Font: Italic

Formatted: Font: Italic

Formatted Table

Formatted: Centered

Formatted: Centered

Formatted: Centered, Tab stops: Not at 7,32 cm + 14,65 cm

Formatted: Centered

Formatted: Left

Formatted: Left, Indent: Left: 0 cm, Right: 0 cm, Line spacing: At least 20 pt

Formatted: Indent: Left: 1,27 cm, No bullets or numbering

Formatted: Centered

Formatted Table

1	,612 <sup>a</sup>	,375	,335	2,2304976
---	-------------------	------	------	-----------

a. Predictors: (Constant), urban, anergiat, tomeas3

Πίνακας 8.8 Μοντέλο παλινδρόμησης με τις X<sub>1</sub>, X<sub>3</sub> και X<sub>4</sub>

Πίνακας 10.4  
ANOVA για το μοντέλο με τις X<sub>1</sub>, X<sub>3</sub> και X<sub>4</sub>

ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	140,108	3	46,703	9,387	,000 <sup>a</sup>
	Residual	233,831	47	4,975		
	Total	373,938	50			

a. Predictors: (Constant), urban, anergiat, tomeas3

b. Dependent Variable: in\_migration

Πίνακας 8.9 ANOVA για το μοντέλο με τις X<sub>1</sub>, X<sub>3</sub> και X<sub>4</sub>

- ο Παλινδρόμηση για το πλήρες μοντέλο

Πίνακας 10.5  
Πλήρες μοντέλο παλινδρόμησης

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,627 <sup>a</sup>	,394	,311	2,2701668

a. Predictors: (Constant), long\_term\_unemployed, urban, anergiat, tomeas3, incomecapita, eduleveltotal

Πίνακας 8.10 Πλήρες Μοντέλο παλινδρόμησης

Πίνακας 10.6

ANOVA πλήρους μοντέλου

ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	147,177	6	24,530	4,760	,001 <sup>a</sup>
	Residual	226,761	44	5,154		
	Total	373,938	50			

a. Predictors: (Constant), long\_term\_unemployed, urban, anergiat, tomeas3, incomecapita, eduleveltotal

b. Dependent Variable: in\_migration

Πίνακας 8.11 ANOVA πλήρους μοντέλου

Formatted: Centered

Formatted: Greek

Formatted Table

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Left

Formatted: Centered

Πανεπιστήμιο Πειραιώς