



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ ΚΑΙ ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ ΥΠΟΔΟΜΗΣ

Παράγοντες που επηρεάζουν το χρόνο άφιξης λεωφορείου στη στάση

Διπλωματική Εργασία



Πουλιάση Αλίκη

Επιβλέπων: Ματθαίος Γ. Καρλαύτης, Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Μάιος 2014

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον κ. Ματθαίο Γ. Καρλαύτη, Αναπληρωτή Καθηγητή της Σχολής Πολιτικών Μηχανικών του Ε.Μ.Π., για την ανάθεση της παρούσας διπλωματικής εργασίας, καθώς και για τις πολύτιμες συμβουλές και κατευθύνσεις κατά τη διάρκεια της εκπόνησής της. Έπειτα, θα ήθελα να ευχαριστήσω βαθύτατα την κα. Βλαχογιάννη Ελένη για την πολύτιμη βοήθεια και υποστήριξή της, καθώς και για τη συνεργασία που είχαμε. Επιπρόσθετα, ιδιαίτερες ευχαριστίες οφείλω να δώσω στον κ. Σπανό Γεώργιο, Διευθυντή Κίνησης του ΟΑΣΘ, και την κα. Χρυσοστόμου Αικατερίνη για την παραχώρηση της βάσης δεδομένων πάνω στην οποία βασίστηκε η παρούσα διπλωματική εργασία. Τέλος, οφείλω να ευχαριστήσω την οικογένειά μου για τη συνεχή υποστήριξη που μου προσέφερε καθ' όλη τη διάρκεια των σπουδών μου.

ΣΥΝΟΨΗ

Σκοπός της παρούσας διπλωματικής εργασίας είναι η εκτίμηση των βασικών παραγόντων που επιδρούν στους χρόνους διαδρομής και στις καθυστερήσεις της Λεωφορειακής Γραμμής 3 των Αστικών Συγκοινωνιών στην πόλη της Θεσσαλονίκης. Τη βάση δεδομένων αποτέλεσαν μετρήσεις από τον Οργανισμό Αστικών Συγκοινωνιών Θεσσαλονίκης, οι οποίες προήλθαν από δελτία κίνησης οχήματος και μετρήσεις κυκλοφοριακών στοιχείων. Οι στατιστικοί έλεγχοι των χρόνων διαδρομής πραγματοποιήθηκαν με χρήση γραμμικών προτύπων παλινδρόμησης και προέκυψαν οι στατιστικά σημαντικότερες μεταβλητές, ενώ οι καθυστερήσεις μελετήθηκαν με λογιστικά πρότυπα. Και στις δυο παραπάνω περιπτώσεις ερευνήθηκαν ξεχωριστά η συνολική διαδρομή της λεωφορειακής γραμμής και η διαδρομή εντός του κέντρου. Τα αποτελέσματα για τους χρόνους άφιξης και τις καθυστερήσεις στην κάθε στάση έδειξαν ότι παράγοντες όπως η μέση ταχύτητα, ο καιρός, ο φόρτος, ο αριθμός των διασταυρώσεων και η απόσταση μεταξύ των στάσεων σχετίζονται άμεσα με τους χρόνους διαδρομής και την εμφάνιση καθυστερήσεων.

Λέξεις κλειδιά: χρόνος διαδρομής, καθυστέρηση, λεωφορείο, γραμμική παλινδρόμηση, λογιστική παλινδρόμηση

ABSTRACT

The purpose of this study is to estimate the main factors affecting on bus travel time and delay, for the Bus Route 3 in the town of Thessaloniki. The data base consists of real time data and traffic data that were collected by the Organization of Urban Transportation of Thessaloniki. There were conducted statistical tests, using linear regression, to predict all the statistically significant parameters that impact on bus travel time, while delays were approached with logistic regression models to conclude in the statistically significant parameters that cause delays. In both cases the whole route and the route into the center of the town were researched separately. The results of the models show that factors such speed, volume, number of intersections and distance between bus stops are related to the bus travel time and the delay.

Key Words: bus travel time, delay, bus, linear regression, logistic regression

ΠΕΡΙΛΗΨΗ

Σκοπός της παρούσας διπλωματικής εργασίας είναι η εκτίμηση των παραγόντων που επηρεάζουν τους χρόνους άφιξης και τις καθυστερήσεις στην κάθε στάση για την λεωφορειακή γραμμή 3 στην πόλη της Θεσσαλονίκης, με χρήση γραμμικών και μη γραμμικών προτύπων. Συγκεκριμένα, με τη χρήση γραμμικών προτύπων καταβάλλεται προσπάθεια επιλογής των παραγόντων που επηρεάζουν σε μεγαλύτερο βαθμό τους χρόνους άφιξης. Στη συνέχεια, με τη χρήση μη γραμμικών στατιστικών προτύπων αναλύεται ποσοτικά η εμφάνιση ή μη καθυστερήσεων. Και στις δύο παραπάνω περιπτώσεις η διερεύνηση αφορά τους χρόνους άφιξης και τις καθυστερήσεις της συνολικής διαδρομής και της διαδρομής εντός του κέντρου της πόλης. Αυτός ο διαχωρισμός πραγματοποιήθηκε για να μπορεί ο αναγνώστης εκτός από τη συνολική εικόνα του δρομολογίου, να αποκτήσει μια εικόνα για την κατάσταση που επικρατεί στο κέντρο, όπου οι κυκλοφοριακές συνθήκες είναι διαφορετικές.

Μετά τον καθορισμό του στόχου πραγματοποιήθηκε βιβλιογραφική ανασκόπηση ερευνών με συναφές αντικείμενο με τη μελέτη των χρόνων άφιξης σε λεωφορειακές γραμμές σε αστικά κέντρα. Ειδικότερα, παρατίθενται τα στοιχεία και η μεθοδολογία με την οποία διερευνούνται και προτυποποιούνται τα μεγέθη αυτά, καθώς και τα αποτελέσματα και συμπεράσματα που προκύπτουν από τις έρευνες αυτές. Μέσω αυτής της διαδικασίας επιλέγεται ο τρόπος με τον οποίο θα μελετηθεί το δείγμα. Τέλος από τη βιβλιογραφική ανασκόπηση που πραγματοποιήθηκε στα πλαίσια της παρούσας διπλωματικής εργασίας γίνεται εμφανής η ανάγκη για περαιτέρω έρευνα πάνω στους χρόνους άφιξης και τις καθυστερήσεις των λεωφορείων στον ελλαδικό χώρο.

Στη συνέχεια πραγματοποιήθηκε η συλλογή των στοιχείων. Τη βάση δεδομένων αποτέλεσαν μετρήσεις για επτά συνολικά ημέρες του Ιανουαρίου και Φεβρουαρίου 2012 στην εν λόγω λεωφορειακή γραμμή. Τα στοιχεία προήλθαν από μετρήσεις του Οργανισμού Αστικών Συγκοινωνιών Θεσσαλονίκης με τη μορφή δελτίων κίνησης οχήματος και μετρήσεων κυκλοφοριακών στοιχείων από φωρατές κυκλοφορίας. Μετά τον έλεγχο και την κωδικοποίηση των στοιχείων εισήχθησαν στο πρόγραμμα EXCEL, όπου έγινε η πρώτη επεξεργασία τους. Αρχικά αφαιρέθηκαν και βελτιστοποιήθηκαν κάποιες ακραίες τιμές και έπειτα υπολογίστηκαν τα περιγραφικά στατιστικά χαρακτηριστικά της βάσης δεδομένων.

Στο θεωρητικό υπόβαθρο παρουσιάζεται η μεθοδολογία με την οποία πραγματοποιήθηκε η στατιστική επεξεργασία των στοιχείων της παρούσας διπλωματικής εργασίας. Αναλυτικότερα, γίνεται αναφορά στη Γραμμική Παλινδρόμηση και στην Λογαριθμοκανονική Παλινδρόμηση (Lognormal Regression), που τελικά χρησιμοποιείται σε αυτή την έρευνα για την μελέτη των χρόνων. Στη συνέχεια παρατίθεται η θεωρία της Λογιστικής Παλινδρόμησης (Logistic Regression), με την

οποία πραγματοποιείται η στατιστική επεξεργασία των καθυστερήσεων. Τέλος, περιγράφεται ο τρόπος διεξαγωγής των παραπάνω μεθόδων βήμα προς βήμα μέσα από το λογισμικό SPSS.

Αναπτύχθηκαν, λοιπόν, μαθηματικά πρότυπα σχετικά με τους χρόνους άφιξης και τις καθυστερήσεις στην κάθε στάση για τη συνολική διαδρομή και για τη διαδρομή του κέντρου. Έτσι, εντοπίστηκαν οι παράμετροι που επηρεάζουν τις δύο παραπάνω μεταβλητές και επιλέχθηκαν τα βέλτιστα μοντέλα.

Στο τελευταίο στάδιο της εργασίας περιλαμβάνεται η παρουσίαση των αποτελεσμάτων της. Τα κυριότερα συμπεράσματα που προκύπτουν είναι τα εξής:

Για τους χρόνους άφιξης της συνολικής διαδρομής, κύριος παράγοντας που επηρεάζει την αύξηση των χρόνων άφιξης του λεωφορείου στη στάση είναι η περίοδος της ημέρας. Ειδικότερα κατά τη δεύτερη και τρίτη περίοδο της ημέρας οι χρόνοι μεταξύ των στάσεων αυξάνονται. Αντίθετα, η μεταβολή του καιρού από βροχή σε ηλιοφάνεια και η αύξηση της μέσης ταχύτητας επιδρούν σημαντικά στη μείωση των χρόνων άφιξης.

Για τους χρόνους άφιξης εντός του κέντρου, σημαντική επίδραση στην αύξηση τους έχει η περίοδος της ημέρας, καθώς κατά τη δεύτερη και την τρίτη περίοδο οι χρόνοι εμφανίζουν τη μεγαλύτερη αύξηση. Επιπλέον, η αύξηση των διασταυρώσεων μεταξύ των στάσεων επηρεάζει τους χρόνους διαδρομής. Αντίθετα για το συγκεκριμένο πρότυπο, κατά τη διάρκεια της Δευτέρας, Τρίτης και Τετάρτης οι χρόνοι εμφανίζουν τη μεγαλύτερη μείωση.

Για τις καθυστερήσεις της συνολικής διαδρομής, κατά τη διάρκεια της Τετάρτης αυξάνεται η πιθανότητα εμφάνισης καθυστερήσεων σε σχέση με το αναμενόμενο, ενώ ο παράγοντας που παρατηρείται ότι μειώνει περισσότερο την πιθανότητα εμφάνισης καθυστερήσεων είναι η ταχύτητα. Επιπλέον, η πιθανότητα εμφάνισης καθυστερήσεων μειώνεται και κατά τη δεύτερη και την τρίτη περίοδο.

Για τις καθυστερήσεις εντός του κέντρου, όλες οι μεταβλητές αυξάνουν την πιθανότητα εμφάνισης καθυστερήσεων εκτός της σταθεράς που είναι η μόνη μεταβλητή του προτύπου με αρνητικό πρόσημο. Ειδικότερα κατά τις ημέρες Δευτέρα και Τρίτη αυξάνεται ιδιαίτερα η πιθανότητα εμφάνισης καθυστερήσεων.

Σε αυτό το τελευταίο στάδιο της παρούσας διπλωματικής εργασίας αναφέρονται επίσης συμπεράσματα από την βιβλιογραφική ανασκόπηση και προτάσεις για περαιτέρω ανάλυση και έρευνα όπως αυτές προκύπτουν μέσα από τη διπλωματική εργασία.

ΠΕΡΙΕΧΟΜΕΝΑ

1. ΕΙΣΑΓΩΓΗ.....	9
1.1 Γενική Ανασκόπηση.....	9
1.2 Σκοπός της Διπλωματικής Εργασίας.....	11
1.3 Δομή της Διπλωματικής Εργασίας.....	11
2.ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ.....	14
2.1 Εισαγωγή.....	14
2.2 Έρευνες από τη διεθνή βιβλιογραφία.....	14
2.3 Συμπεράσματα.....	20
3. ΠΕΡΙΓΡΑΦΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ.....	24
3.1 Περιγραφή βάσης δεδομένων.....	24
3.1.1 Εισαγωγή.....	24
3.1.2 Στοιχεία μετρήσεων σε πραγματικό χρόνο.....	26
3.1.3 Κυκλοφοριακά Στοιχεία.....	28
3.2 Επεξεργασία Βάσης Δεδομένων.....	30
3.2.1 Επεξεργασία στοιχείων μετρήσεων σε πραγματικό χρόνο.....	31
3.2.2 Επεξεργασία Κυκλοφοριακών Στοιχείων.....	32
3.3 Περιγραφικά στατιστικά χαρακτηριστικά βάσης δεδομένων.....	33
3.3.1 Μελέτη συνολικής διαδρομής.....	35
3.3.2 Μελέτη διαδρομής εντός του κέντρου.....	40
3.4 Μελέτη Κατανομής βάσης δεδομένων.....	45
4. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ.....	49
4.1 Εισαγωγή.....	49
4.2 Μέθοδοι Στατιστικής Ανάλυσης.....	50
4.2.1 Γραμμική Παλινδρόμηση – Λογαριθμοκανονική Παλινδρόμηση.....	50
4.2.2 Λογιστική Παλινδρόμηση.....	53
4.3 Βήματα Κατασκευής Μοντέλου Γραμμικής Παλινδρόμησης και Λογιστικής Παλινδρόμησης – Κριτήρια Αποδοχής.....	56
4.4 Λειτουργία του Λογισμικού SPSS.....	57
4.4.1 Λειτουργία SPSS για τη Γραμμική Παλινδρόμηση.....	57
4.4.2 Λειτουργία SPSS για τη Λογιστική Παλινδρόμηση.....	63

4.5 Κριτήρια Αποδοχής Αποτελεσμάτων	69
4.5.1 Κριτήρια Αποδοχής Αποτελεσμάτων στη Γραμμική Παλινδρόμηση	69
4.5.2 Κριτήρια Αποδοχής Αποτελεσμάτων στη Λογιστική Παλινδρόμηση	71
4.6 Το Likelihood Ratio Statistic Test	73
5. ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ – ΑΠΟΤΕΛΕΣΜΑΤΑ	75
5.1 Εισαγωγή	75
5.2 Εισαγωγή δεδομένων στο πρόγραμμα	77
5.3 Κατηγοριοποίηση και Συσχέτιση Μεταβλητών	83
5.4 Στατιστική Ανάλυση Δεδομένων	84
5.5 Ανάπτυξη Μαθηματικών Προτύπων	85
5.5.1 Πρότυπα σχετικά με τους Χρόνους Άφιξης	85
5.5.2 Πρότυπα σχετικά με τις καθυστερήσεις	93
6. ΣΥΜΠΕΡΑΣΜΑΤΑ	100
6.1 Εισαγωγή	100
6.2 Σύνοψη Αποτελεσμάτων	100
6.3 Προτάσεις για περαιτέρω έρευνα	105
Βιβλιογραφία	107
ΠΑΡΑΡΤΗΜΑ	109

1. ΕΙΣΑΓΩΓΗ

1.1 Γενική Ανασκόπηση

Οι χρόνοι διαδρομών και οι καθυστερήσεις των λεωφορείων Αστικών Συγκοινωνιών επιδρούν σημαντικά στις προτιμήσεις των χρηστών και στην αξιοπιστία των μετακινήσεων. Ειδικά σε μεγάλα αστικά κέντρα γεννάται η ανάγκη τα ΜΜΜ να εξασφαλίζουν στους πολίτες μια άνετη και αξιόπιστη μετακίνηση, κάτι που οδηγεί στη συνεχή μελέτη και βελτίωσή τους.

Στην παρούσα διπλωματική εργασία μελετάται μια συγκεκριμένη λεωφορειακή γραμμή στην πόλη της Θεσσαλονίκης. Η Θεσσαλονίκη είναι η δεύτερη μεγαλύτερη πόλη της Ελλάδας με συνολικό πληθυσμό περίπου 1 εκατομμύριο. Εντός του κέντρου ο πληθυσμός της είναι περίπου 320.000 κάτοικοι και η πλειοψηφία υπηρεσιών και επιχειρήσεων είναι συγκεντρωμένη εκεί, γεγονός που αποδεικνύει τη χρησιμότητα και αναγκαιότητα των αστικών συγκοινωνιών.

Στην πόλη της Θεσσαλονίκης το επιβατικό κοινό εξυπηρετείται από 78 συνολικά δρομολόγια τα οποία εκτείνονται σε απόσταση 50 χιλιομέτρων περιμετρικά της πόλης, καλύπτοντας έτσι το μεγαλύτερο κομμάτι της. Η μορφή του δικτύου είναι κυρίως ακτινωτή με πόλους τους σταθμούς μετεπιβίβασης στις παρυφές της πόλης και το ευρύτερο εμπορικό κέντρο. Το συνολικό μήκος του συγκοινωνιακού δικτύου ξεπερνά τα 970.000 χλμ. και εκτελούνται πάνω από 5.800 δρομολόγια ημερησίως. Το Σάββατο και την Κυριακή ο αριθμός αυτός μειώνεται κατά 9% περίπου. Υπάρχουν δύο ειδών δρομολόγια τα αστικά και τα περιαστικά, για το κέντρο και τους γύρω δήμους, τα οποία έχουν συχνότητες δρομολογίων 15 και 29 λεπτά αντίστοιχα. Επιπλέον ο Οργανισμός Αστικών Συγκοινωνιών Θεσσαλονίκης (ΟΑΣΘ) έχει εγκαταστήσει σύστημα τηλεματικής στην πόλη σε 200 στάσεις, από το σύνολο των 2000, για τον εντοπισμό των λεωφορείων και την άμεση πληροφόρηση των επιβαινόντων (οπτική και ηχητική).

Σύμφωνα με την πιο πρόσφατη έρευνα του Συμβουλίου Αστικών Συγκοινωνιών Θεσσαλονίκης (ΣΑΣΘ) που πραγματοποιήθηκε το 2008, συνολικά πραγματοποιούνται ετησίως περίπου 2 εκατομμύρια δρομολόγια και διανύονται 42 εκατομμύρια χιλιόμετρα εξυπηρέτησης. Επιπλέον η επιβατική κίνηση φαίνεται να παρουσιάζει αύξηση στην τελευταία πενταετία των μελετών (2003-2008) της τάξεως του 32%, φτάνοντας στα 177 εκατομμύρια επιβάτες ετησίως. Ο αριθμός αυτός αποτελεί το 42% της πληρότητας του συστήματος, το οποίο έχει ολική μεταφορική ικανότητα 422 εκατομμύρια θέσεις.

Στην παρούσα εργασία μελετάται συγκεκριμένα η λεωφορειακή γραμμή 3 Ανατολικός Σταθμός ΙΚΕΑ – Τερματικός Σταθμός Γ. Ειδικότερα, διερευνώνται οι παράγοντες που επηρεάζουν τους χρόνους άφιξης του λεωφορείου στην κάθε στάση,

καθώς επίσης και η πιθανότητα εμφάνισης καθυστερήσεων στην εκάστοτε στάση. Βάσει της παραπάνω μελέτης η γραμμή αυτή αποτέλεσε για το 2008 τη δημοφιλέστερη διαδρομή, εξυπηρετώντας συνολικά 12 εκατ. επιβάτες, δηλαδή το 7% περίπου της ολικής επιβατικής κίνησης. Για το λόγο αυτό, η μελέτη μιας τέτοιας γραμμής έχει ιδιαίτερη χρησιμότητα τόσο για τη βελτίωση και αναβάθμιση των αστικών συγκοινωνιών, όσο και για την άριστη εξυπηρέτηση των χρηστών της γραμμής.

Στον παρακάτω Πίνακα 1.1 παρουσιάζονται τα στοιχεία των μετρήσεων από το 2003-2008 για όλα τα δρομολόγια των Αστικών λεωφορείων, από τον ιστότοπο του ΣΑΣΘ www.thita.gov.gr. Παρουσιάζονται τα ετήσια δρομολόγια, χλμ εξυπηρέτησης, μεταφορική ικανότητα (στόλος-θέσεις), επιβατική κίνηση, πληρότητα και ποσοστιαία μεταβολή.

Πίνακας 1.1 Στοιχεία μετρήσεων ΣΑΣΘ

ΕΤΟΣ	Δρομολόγια	Χιλιόμετρα εξυπηρέτησης	Μεταφορική Ικανότητα (Θέσεις)	Επιβατική κίνηση	Μέση Πληρότητα (%)	Ποσοστιαία μεταβολή
2003			380.818.181	134.017.206	35,20%	
2004	1.859.381	40.651.345	411.331.096	159.825.104	38,90%	19,26%
2005	1.827.080	40.017.104	411.509.547	160.772.613	39,10%	0,59%
2006	1.917.060	42.809.753	434.273.091	173.606.596	40,00%	7,98%
2007	1.926.427	42.953.932	436.199.335	177.572.431	40,70%	2,28%
2008	1.861.969	42.058.121	422.069.873	176.814.922	41,90%	-0,43%

1.2 Σκοπός της Διπλωματικής Εργασίας

Με βάση όσα αναφέρθηκαν στη γενική ανασκόπηση, διαφαίνεται η σημαντικότητα να διεξαχθεί μια ολοκληρωμένη έρευνα σε ότι αφορά τους χρόνους άφιξης και τις καθυστερήσεις των λεωφορείων Αστικών Συγκοινωνιών στην πόλη της Θεσσαλονίκης.

Κατά τη μελέτη της Διαδρομής 3 λαμβάνονται υπόψη μεταξύ άλλων στοιχεία του συγκεκριμένου οδικού δικτύου, όπως οι αποστάσεις των στάσεων, οι φόρτοι και οι ταχύτητες που παρατηρούνται, οι διασταυρώσεις που παρεμβάλλονται ανάμεσα στις στάσεις. Για το λόγο αυτό δεν μπορεί να θεωρηθεί ότι οι ίδιες μεταβλητές γίνεται να εκφράσουν το σύνολο των δρομολογίων των αστικών λεωφορείων. Βέβαια αν γίνει η υπόθεση ότι παρόμοιες συνθήκες επικρατούν στο δίκτυο, καθώς η πόλη δεν έχει τόσο μεγάλη έκταση, είναι πιθανό ύστερα από κάποιες διαφοροποιήσεις τα τελικά μοντέλα να μπορούν να εκφράσουν και τα υπόλοιπα δρομολόγια.

Η παρούσα διπλωματική εργασία έχει σαν στόχο την εξαγωγή μαθηματικών προτύπων για τους χρόνους και τις καθυστερήσεις της λεωφορειακής γραμμής 3. Μέσα από αυτά τα πρότυπα διερευνάται η επίδραση ενός συνόλου παραγόντων στους χρόνους διαδρομής και στις καθυστερήσεις σε σχέση με το αναμενόμενο στην κάθε στάση. Επιπλέον, γίνεται μια πιο εκτενής διερεύνηση των δύο παραπάνω παραγόντων ανάλογα με το αν το λεωφορείο βρίσκεται στη συνολική διαδρομή ή στη διαδρομή εντός του κέντρου, όπου οι κυκλοφοριακές συνθήκες αλλάζουν. Ειδικότερα, σκοπός της παρούσας διπλωματικής εργασίας είναι η ποσοτική ανάλυση με στατιστικές μεθόδους της μεταβλητότητας των χρόνων για τη συνολική διαδρομή και για τη διαδρομή εντός του κέντρου. Και στη συνέχεια, σκοπός της είναι η ποσοτική ανάλυση με στατιστικές μεθόδους της εμφάνισης καθυστερήσεων για της δύο παραπάνω διαδρομές.

Τα στοιχεία που χρησιμοποιήθηκαν στην παρούσα έρευνα δόθηκαν από τον Οργανισμό Αστικών Συγκοινωνιών με τη μορφή Δελτίων Κίνησης Οχήματος. Ύστερα από κατάλληλη επεξεργασία της βάσης δεδομένων, τα στοιχεία μελετήθηκαν στατιστικά με σκοπό τη χρήση τους τόσο σε γραμμικά, όσο και σε μη γραμμικά πρότυπα για τη διεξαγωγή των αποτελεσμάτων. Για την τελική μελέτη τους λήφθηκαν υπόψη μια σειρά παραγόντων που αποτέλεσαν στοιχεία της διαδρομής (αποστάσεις, διασταυρώσεις), εξωγενείς παράγοντες (καιρικές συνθήκες, ημέρα, περίοδος της ημέρας), καθώς επίσης και κυκλοφοριακά στοιχεία (φόρτοι, ταχύτητες).

1.3 Δομή της Διπλωματικής Εργασίας

Τα κεφάλαια της παρούσας διπλωματικής εργασίας ακολουθούν την παρακάτω δομή:

Το **1^ο κεφάλαιο** είναι εισαγωγικό και ο σκοπός του είναι να δώσει στον αναγνώστη μια γενική εικόνα για το αντικείμενο που πραγματεύεται η παρούσα

διπλωματική εργασία. Για το λόγο αυτό ξεκινά με μια συνοπτική παρουσίαση του αντικειμένου μελέτης, της περιοχής μελέτης, καθώς και της παρουσίασης κάποιων δημογραφικών χαρακτηριστικών της περιοχής. Γίνεται εκτενέστερη αναφορά στην αναγκαιότητα ύπαρξης αξιόπιστων αστικών συγκοινωνιών, με δρομολόγια που πραγματοποιούνται εντός των προγραμματισμένων χρόνων και δεν εμφανίζουν ιδιαίτερες καθυστερήσεις. Στη συνέχεια καθορίζεται ο **στόχος** της συγκεκριμένης εργασίας. Το εισαγωγικό αυτό κεφάλαιο ολοκληρώνεται με την παρουσίαση της δομής της εργασίας.

Το **2^ο κεφάλαιο** περιλαμβάνει τη βιβλιογραφική ανασκόπηση, που αφορά στην ανάλυση ερευνών από τη διεθνή βιβλιογραφία σχετικά με τη μελέτη των χρόνων άφιξης σε συγκεκριμένες γραμμές, τους παράγοντες και το βαθμό στον οποίο τους επηρεάζουν. Η έρευνα του συγκεκριμένου αντικειμένου πραγματοποιείται με μια πληθώρα μεθόδων. Επίσης, η βιβλιογραφική ανασκόπηση περιλαμβάνει έρευνες που αφορούν λεωφορειακές γραμμές με πολλαπλές διαδρομές, καθώς και έρευνες πάνω στους χρόνους διαφόρων γραμμών σε μια κομβική στάση. Παρατίθενται, επιπλέον, και τα σχετικά αποτελέσματα των ερευνών αυτών. Τέλος, αναφέρεται η έλλειψη σχετικών στοιχείων στο Ελλαδικό χώρο και η αναγκαιότητα μιας πιο ολοκληρωμένης έρευνας σχετικής με τους χρόνους άφιξης και την εμφάνιση καθυστερήσεων. Με αυτό τον τρόπο παρουσιάζεται η σημαντικότητα της έρευνας που εκπονείται στα πλαίσια της παρούσας διπλωματικής εργασίας.

Το **3^ο κεφάλαιο** περιλαμβάνει την περιγραφή και επεξεργασία της βάσης δεδομένων που χρησιμοποιήθηκαν στην παρούσα εργασία. Αρχικά περιγράφονται τα δεδομένα της βάσης, αναλύονται τα στοιχεία των δελτίων κίνησης οχήματος και τα κυκλοφοριακά στοιχεία, καθώς και η αρχική επεξεργασία τους. Στη συνέχεια, γίνεται αναφορά στα περιγραφικά χαρακτηριστικά της βάσης δεδομένων με ανάλογους επεξηγηματικούς πίνακες και σχολιασμό των αποτελεσμάτων για όλες τις περιπτώσεις της έρευνας. Τέλος, πραγματοποιείται και σχολιάζεται η μελέτη κατανομής της βάσης δεδομένων.

Το **4^ο κεφάλαιο** περιλαμβάνει το θεωρητικό υπόβαθρο της διπλωματικής εργασίας. Το θεωρητικό υπόβαθρο αναφέρεται στην περιγραφή της επιλεγόμενης μεθοδολογίας. Παρουσιάζεται και αναλύεται σε βάθος η μέθοδος της γραμμικής, της λογαριθμοκανονικής και της λογιστικής παλινδρόμησης οι οποίες χρησιμοποιούνται ευρύτατα σε παρόμοιες έρευνες. Γίνεται, επιπλέον, μια εκτενής περιγραφή του τρόπου που πραγματοποιήθηκαν οι αναλύσεις των παραπάνω μεθόδων με τη χρήση του στατιστικού λογισμικού SPSS. Παρουσιάζονται αναλυτικά τα βήματα που πραγματοποιήθηκαν για την ανάπτυξη των μοντέλων στο πρόγραμμα, οι παράμετροι που ελέγχθηκαν και τα κριτήρια αποδοχής των τελικών προτύπων.

Το **5^ο κεφάλαιο** περιλαμβάνει τη διαδικασία ανάπτυξης και εφαρμογής των τελικών προτύπων. Ιδιαίτερη σημασία δίνεται στους στατιστικούς ελέγχους αξιοπιστίας που προκύπτουν από έρευνες, για την επιλογή των βέλτιστων μοντέλων. Επιπλέον παρατίθενται και σχολιάζονται τα αποτελέσματα που προκύπτουν τελικά με τη μορφή πινάκων, με σκοπό να συγκεντρωθούν και να παρουσιαστούν τα αποτελέσματα της έρευνας πεδίου. Τέλος, γίνεται αναλυτική ερμηνεία των αποτελεσμάτων των προτύπων.

Το **6^ο κεφάλαιο** είναι το τελευταίο κεφάλαιο της έρευνας και περιλαμβάνει την παρουσίαση των τελικών αποτελεσμάτων και την ερμηνεία τους. Γίνεται μια αναφορά στην χρησιμότητα των βασικών αποτελεσμάτων και συμπερασμάτων της παρούσας διπλωματικής εργασίας, ενώ ταυτόχρονα παρατίθενται προτάσεις και κατευθύνσεις για περαιτέρω έρευνα πάνω στο θέμα της διπλωματικής εργασίας. Σκοπός είναι να υπάρξει μια πιο ολοκληρωμένη εικόνα όσον αφορά τους παράγοντες που επηρεάζουν τους χρόνους άφιξης και τις καθυστερήσεις στην κάθε στάση.

Στο τέλος της διπλωματικής εργασίας παρατίθεται κατάλογος με όλες τις **βιβλιογραφικές αναφορές** που χρησιμοποιήθηκαν κατά την εκπόνηση της παρούσας έρευνας και **παράρτημα** των αποτελεσμάτων από το λογισμικό.

2.ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

2.1 Εισαγωγή

Όπως προαναφέρθηκε στο εισαγωγικό κεφάλαιο, η παρούσα διπλωματική εργασία έχει σαν στόχο την καθορισμό των παραγόντων που επηρεάζουν το χρόνο άφιξης και τις καθυστερήσεις λεωφορείων στις στάσεις του δρομολογίου 3 (Ανατολικός Σταθμός ΙΚΕΑ- Τερματικός Σταθμός Γ), στην πόλη της Θεσσαλονίκης. Πιο συγκεκριμένα, σκοπός της εργασίας είναι η ανάλυση και ο προσδιορισμός, με στατιστικές μεθόδους, των παραγόντων που τα επηρεάζουν σε μεγαλύτερο βαθμό. Πριν την παρουσίαση των στοιχείων, της βάσης δεδομένων, του τρόπου επεξεργασίας τους και της επιλεγόμενης μεθοδολογίας, κρίνεται απαραίτητο να γίνει αναφορά στις προσπάθειες ερευνητών στη διεθνή βιβλιογραφία οι οποίες είναι κοντά στο αντικείμενο που πραγματεύεται η συγκεκριμένη έρευνα.

Αρχικά γίνεται ανασκόπηση σε έρευνες σχετικές με την πρόβλεψη των χρόνων άφιξης λεωφορείων σε στάσεις και την ανάπτυξη στατιστικών μοντέλων για την καλύτερη προσέγγιση του θέματος. Παρουσιάζεται, επίσης, μια έρευνα στην οποία μελετάται η πρόβλεψη των χρόνων άφιξης σε συγκεκριμένη στάση η οποία χρησιμοποιείται από πολλαπλές λεωφορειακές γραμμές. Στις περισσότερες έρευνες τα στοιχεία έχουν συλλεχθεί είτε από μετρητές κυκλοφορίας, είτε από συστήματα εντοπισμού θέσης (GPS) που έχουν τοποθετηθεί πάνω στα λεωφορεία. Στη βιβλιογραφική ανασκόπηση που θα ακολουθήσει, και που θα ασχολείται με τις έρευνες αυτές που έχουν γίνει στο παρελθόν, παρατίθενται τα στοιχεία και η μεθοδολογία με την οποία διερευνούνται και προτυποποιούνται τα μεγέθη αυτά, καθώς και τα αποτελέσματα και συμπεράσματα που προκύπτουν από τις έρευνες αυτές.

Εάν αναλογιστεί κανείς τις μειώσεις των δρομολογίων, των οχημάτων και των λεωφορειακών γραμμών που καλούνται να εξυπηρετήσουν το σύνολο του επιβατικού κοινού στις ελληνικές συγκοινωνίες του σήμερα, η αναγκαιότητα μιας έγκυρης πρόβλεψης είναι μεγάλη.

2.2 Έρευνες από τη διεθνή βιβλιογραφία

Το 2009 πραγματοποιήθηκε από τους **Jing-nan Wang, Xu-mei Chen, Shu xia Guo** μια έρευνα με τίτλο «Bus travel time prediction model with v-support vector regression». **Σκοπός** αυτής της έρευνας είναι η πρόβλεψη χρόνου άφιξης των λεωφορείων στις στάσεις ώστε να μπορούν μακροπρόθεσμα να χρησιμοποιηθούν τα ρεαλιστικά αυτά στοιχεία σε συσκευές Intelligent Transportation Systems (ITS). Έτσι μέσω ηλεκτρονικών πινάκων στις στάσεις, οι επιβάτες θα διευκολύνονται και η χρήση λεωφορείου θα γίνεται πιο ελκυστική. Μελετήθηκε η μία διεύθυνση της εξπρές

λεωφορειακής γραμμής 104 στο Beijing της Κίνας, μήκους 14.7χλμ., στην οποία υπάρχουν 23 στάσεις λεωφορείου. Η συλλογή στοιχείων έγινε από ερευνητές με επί τόπου μετρήσεις, η διαδικασία αυτή διήρκησε σχεδόν ένα μήνα, με μετρήσεις 3 ημέρες την εβδομάδα σε ώρες αιχμής.

Στην συγκεκριμένη έρευνα των Jing-nan Wang, Xu-mei Chen, and Shu-xia Guo χρησιμοποιήθηκαν οι θεωρίες **Support Vector Machine (SVM)** και **Kernel function**, πιο συγκεκριμένα πρότειναν ένα ν - Support Vector Regression model. Οι παράμετροι που λαμβάνονται υπόψη συνδέονται έμμεσα με τα αποτελέσματα που προβλέπονται, όπως: χρόνος αναχώρησης από συγκεκριμένη στάση, μήκος διαδρομής, αριθμός διασταυρώσεων, σηματοδότηση ή μη των κόμβων στις διασταυρώσεις. Επιπλέον χρησιμοποιήθηκαν οι εξής παραδοχές:

- 1) Οι κυκλοφοριακές συνθήκες που επικρατούν στο αστικό οδικό δίκτυο είναι παρόμοιες για όλες τις ημέρες της εβδομάδας.
- 2) Ο χρόνος διαδρομής εξαρτάται από το μήκος της διαδρομής, καθώς και από τις καθυστερήσεις που προκαλεί ο σηματοδοτούμενος κόμβος.
- 3) Ο διάρκεια διαδρομής συνδέεται άμεσα με το χρόνους του λεωφορείου σε συγκεκριμένη λεωφορειακή γραμμή (historical travel time).

Μετά την ολοκλήρωση της έρευνας έγινε σύγκριση των χρόνων που προέκυψαν απ' το ν -Support Vector Regression model και των πραγματικών. Υπολογίστηκαν τα εξής:

- Η απόλυτη τιμή της μέσης διαφοράς δυο ανεξάρτητων μεταβλητών (MAD).
- Η απόλυτη τιμή του μέσου ποσοστού σφάλματος(MAPE).
- Η απόλυτη τιμή του αριθμητικού μέσου όρου του συνόλου των σφαλμάτων της πρόβλεψης (MAE).
- Το μέγιστο σχετικό σφάλμα (MRE).

Από τα **αποτελέσματα** προέκυψε ότι οι τιμές των Maximum Absolute Error (MAE) και Maximum Relative Error (MRE) είναι σε αποδεκτά όρια και συνεπώς η πρόβλεψη είχε σχετικά υψηλή ακρίβεια δεδομένου ότι το δείγμα ήταν μικρό.

Τα **συμπεράσματα** της έρευνας που διεξήχθησαν ήταν ότι παρόλο που το μοντέλο ν -SVR έδωσε ικανοποιητικά αποτελέσματα υπήρχαν ακόμα κάποιες παράμετροι που δεν είχαν εξετασθεί. Αρχικά, δεν έγινε επαρκής σύγκριση με άλλα μοντέλα όπως για παράδειγμα τα Artificial Neural Networks ή ARIMA. Επιπλέον, η εισαγωγή άλλων παραμέτρων και η συσχέτιση των στοιχείων με δεδομένα της ίδιας γραμμής κάποιας άλλης ημέρας ενδεχομένως θα επηρέαζαν τα αποτελέσματα. Τέλος, η χρήση μοντέλου Kernel και η εύρεση των κατάλληλων παραμέτρων θα αποτελούσε μια πολύ χρονοβόρα διαδικασία στην περίπτωση μεγαλύτερου δείγματος.

Το 2009 διεξήχθη μια έρευνα από τους **YU Bo, LU Jing, YU Bin, YANG Zhongzhen** με τίτλο «An Adaptive Bus Arrival Time Prediction Model». Σκοπό είχε την πρόβλεψη χρόνου άφιξης των λεωφορείων σε συγκεκριμένες στάσεις για βελτίωση των συστημάτων πληροφόρησης των επιβατών, όπως τα Advanced Traveler Information Systems (ATIS) και Advanced Public Transportation Systems (APTS), για το άμεσο μέλλον, προκειμένου να μειωθούν οι χρόνοι αναμονής, οι καθυστερήσεις, οι αποκλίσεις από το πρόγραμμα, αλλά και τα κόστη λειτουργίας.

Στη συγκεκριμένη έρευνα χρησιμοποιούνται μοντέλα **Γραμμικής Παλινδρόμησης (Linear Regression)**, καθώς και ένας **προσαρμοζόμενος αλγόριθμος**. Μέσω των μοντέλων γραμμικής παλινδρόμησης εκτιμάται ο χρόνος αναφοράς μεταξύ γειτονικών στάσεων βάσει των κυκλοφοριακών συνθηκών που επικρατούν στο επόμενο οδικό τμήμα, ενώ για τον προσαρμοσμένο αλγόριθμο χρησιμοποιούνται οι πιο πρόσφατες πληροφορίες χρόνων άφιξης λεωφορείων μαζί με τους εκτιμώμενους χρόνους αναφοράς από τη γραμμική παλινδρόμηση για την πρόβλεψη χρόνων άφιξης στις κατάντη στάσεις.

Η μελέτη πραγματοποιήθηκε στην Κίνα, για την πόλη Dalian και τα δεδομένα του λεωφορείου με αριθμό 23. Η συγκεκριμένη λεωφορειακή γραμμή πραγματοποιεί τη διαδρομή από τα προάστια ως το κέντρο της πόλης, έχει μήκος 14,5χλμ. και 19 στάσεις σε κάθε κατεύθυνση, έχει υψηλή συμφόρηση στην πρωινή και απογευματινή αιχμή και ο χρόνος διαδρομής στις αιχμές είναι περίπου 45 λεπτά. Η συλλογή στοιχείων έγινε για 16 στάσεις σε κάθε διαδρομή στις ώρες αιχμής, για ένα μήνα μετρήσεων, 520 διαδρομές συνολικά.

Από τα **αποτελέσματα** των μοντέλων της **γραμμικής παλινδρόμησης (LR)** και τους πραγματικούς χρόνους άφιξης των λεωφορείων στις στάσεις υπολογίστηκε η ρίζα του μέσου τετραγωνικού σφάλματος (Root Mean Squared Error) για κάθε τμήμα. Με βάση τη θεώρηση ότι η κίνηση πεζών είναι αυξημένη και οι δρόμοι της περιοχής στενοί, προέκυψε ότι η χρήση δύο ανεξάρτητων μεταβλητών στο μοντέλο δίνει πιο έγκυρα αποτελέσματα. Αυτό μπορεί να αποδοθεί στο ότι τα μοντέλα με δύο ανεξάρτητες μεταβλητές λαμβάνουν υπόψη τις κυκλοφοριακές συνθήκες που επικρατούν τόσο στο μελετώμενο κομμάτι, όσο και στο επόμενο. Επίσης, αναφέρεται ότι οι συνθήκες λειτουργίας στο επόμενο κομμάτι είναι σημαντικότερες από αυτές του παρόντος, καθώς επίσης ότι τα RSMEs είναι υψηλότερα. Ακόμη, υψηλές τιμές RSMEs εμφανίστηκαν στο δέκατο και δωδέκατο τμήμα της γραμμής, λόγω του ότι βρίσκονται στο κέντρο της πόλης Dalian, Central Business District (CBD), όπου σε ώρες αιχμής οι ταχύτητες μειώνονται και φτάνουν σε μια σχεδόν στασιμότητα.

Το **ενισχυμένο μοντέλο γραμμικής παλινδρόμησης** το οποίο περιλαμβάνει και τον **προσαρμοζόμενο αλγόριθμο** έχει σχεδιαστεί για να δίνει βελτιωμένα τα αποτελέσματα της γραμμικής παλινδρόμησης. Η διαδικασία που ακολουθήθηκε είναι η εξής: Κάθε φορά που φτάνει το λεωφορείο σε μια στάση, οι χρόνοι άφιξης που έχουν προβλεφθεί ρυθμίζονται με βάση τα καινούργια στοιχεία. Για να διεξαχθούν συμπεράσματα για την αποτελεσματικότητα του μοντέλου έγινε σύγκριση μεταξύ των συγκεκριμένων αποτελεσμάτων, της γραμμικής παλινδρόμησης και του καθορισμένου χρονοδιαγράμματος. Η σύγκριση έδειξε ότι τα μοντέλα γραμμικής παλινδρόμησης δίνουν τις καλύτερες ενδείξεις για το χρόνο άφιξης μεταξύ δύο γειτονικών στάσεων. Ενώ για περισσότερα τμήματα διαδρομής τα RMSEs του ενισχυμένου μοντέλου γραμμικής παλινδρόμησης είναι μικρότερα.

Συνεπώς, τα **βασικά συμπεράσματα** της έρευνας είναι τα εξής: Το ενισχυμένο μοντέλο γραμμικής παλινδρόμησης μπορεί να προσομοιάσει καλύτερα τους πραγματικούς χρόνους, έχει τη δυνατότητα προσαρμογής σε δυναμικές αλλαγές της κυκλοφορίας και μπορεί να εφαρμοστεί σε πραγματικές λεωφορειακές γραμμές, παρέχοντας έγκυρες πληροφορίες για το χρόνο άφιξης του λεωφορείου στη στάση.

Το **2011** πραγματοποιήθηκε μια έρευνα από τους **Bin Yu, William H.K. Lam, Mei Lam Tam** με τίτλο «Bus arrival time prediction at bus stop with multiple routes». Στην παρούσα έρευνα προτείνονται μοντέλα για την πρόβλεψη του χρόνου άφιξης λεωφορείου σε συγκεκριμένη στάση με πολλαπλές διαδρομές. Κάτι τέτοιο θα έδινε τη δυνατότητα επιλογής κατάλληλης λεωφορειακής γραμμής στους επιβάτες, με σκοπό την ταχύτερη άφιξη στον προορισμό τους.

Χρησιμοποιήθηκαν τέσσερις διαφορετικές μέθοδοι: **Support Vector Machine (SVM), Artificial Neural Network (ANN), k-Nearest Neighbors algorithm (k-NN) and Linear Regression (LR)**.

Τα μοντέλα μελετήθηκαν στη στάση Cross Harbour Tunnel (CHT) στην αστική περιοχή Kowloon Central του Hong-Kong για δύο διαδρομές. Η μια ήταν μεταξύ της δυτικής κατεύθυνσης Chatham Road North (CRN) προς τη στάση Cross Harbour Tunnel (CHT) και η άλλη από την ανατολική κατεύθυνση Ping Chi Street (PCS) προς τη μελετώμενη στάση. Στην πρώτη διαδρομή μελετήθηκαν οκτώ λεωφορειακές γραμμές και η απόσταση μεταξύ των δύο στάσεων ήταν 0.62 χλμ, ενώ στη δεύτερη διαδρομή μελετήθηκαν έξι λεωφορειακές γραμμές και η αντίστοιχη απόσταση ήταν 0.72 χλμ.

Αρχικά διεξήχθησαν έρευνες για τη συλλογή στοιχείων για τους χρόνους διαδρομής και χρόνους άφιξης των λεωφορείων στη στάση σε δύο τυπικές ημέρες της εβδομάδας, 11 – 12 Μαΐου 2010 (Τρίτη προς Τετάρτη) και 8 Ιουνίου 2010 (Τρίτη) στην ώρα της πρωινής αιχμής (08:00 – 10:00). Από αυτά τα στοιχεία υπολογίστηκαν ο

μέγιστος, ο ελάχιστος και ο μέσος χρόνος κάθε διαδρομής καθώς και η ρίζα του μέσου τετραγωνικού σφάλματος. Στη συγκεκριμένη έρευνα εκτιμήθηκαν οι χρόνοι άφιξης των λεωφορείων και υπολογίστηκαν μέσω των προαναφερθέντων μεθόδων τα MAE, MAPE, RMSE για την εύρεση του μοντέλου που προσομοιάζει καλύτερα το δείγμα.

Το **συμπέρασμα** αυτής της έρευνας ήταν ότι τα μοντέλα που χρησιμοποιήθηκαν είναι πιο ακριβή για λεωφορεία με πολλαπλές διαδρομές παρά για λεωφορεία που πραγματοποιούσαν μια διαδρομή. Επίσης βρέθηκε ότι το **Support Vector Machine (SVM)** model έδωσε τα καλύτερα αποτελέσματα από τα τέσσερα μοντέλα που χρησιμοποιήθηκαν για την πρόβλεψη χρόνου άφιξης λεωφορείου σε συγκεκριμένη στάση με πολλαπλές διαδρομές.

Το **2012** πραγματοποιήθηκε μια έρευνα από τους **Mazloumi E., Moridpour S., Currie G., Rose G.**, με τίτλο «Exploring the value of traffic flow data in bus travel time prediction». **Σκοπός** της έρευνας αυτής είναι η πρόβλεψη των χρόνων διαδρομής των λεωφορείων βάσει στοιχείων κυκλοφοριακής ροής που έχουν συλλεχθεί στην λεωφορειακή γραμμή 246 στη Μελβούρνη, Αυστραλία. Η γραμμή αυτή είναι 8 χλμ., έχει χωριστεί σε τέσσερα ισομήκη τμήματα και έχουν ορισθεί σε αυτήν πέντε χρονικά σημεία (timing points). Για την έρευνα χρησιμοποιήθηκαν στοιχεία από πραγματικές μετρήσεις με συσκευές GPS που τοποθετήθηκαν σε συγκεκριμένα λεωφορεία αυτής της γραμμής και κατέγραφαν τους χρόνους άφιξης/ αναχώρησης σε κάθε στάση για 6 μήνες.

Στη συγκεκριμένη έρευνα χρησιμοποιήθηκε **Regression Analysis** για να βρεθεί η επιρροή κάθε εισερχόμενης μεταβλητής στο μοντέλο και στη συνέχεια, αφού βρέθηκαν οι στατιστικά σημαντικές μεταβλητές, αναπτύχθηκαν **Artificial Neural Network (ANN) models**. Επίσης, έγινε χρήση τριών εναλλακτικών μοντέλων για την πρόβλεψη του χρόνου άφιξης: **Traffic Flow Data-Based Model, Historical Data-Based Model, Timetable-Based Model**.

Οι μεταβλητές που χρησιμοποιήθηκαν τελικά στο Traffic Flow Data – Based Model ήταν οι εξής:

- 1) Οι χρόνοι άφιξης στη στάση συγκρινόμενοι με τους προγραμματισμένους χρόνους (schedule adherence). Αυτό γίνεται με αφαίρεση του πραγματικού χρόνου άφιξης στο ανάντη χρονικό σημείο από τον εκτιμώμενο χρόνο.
- 2) Η βροχόπτωση (rainfall) μετρούμενη σε χιλιοστά για ώρα που γίνεται η μέτρηση.
- 3) Οι κυκλοφοριακοί φόρτοι στο συγκεκριμένο κομμάτι ,οι οποίοι μετρήθηκαν μέσω μετρητών κυκλοφορίας (inductive loop detectors) συνδεδεμένων στο σύστημα Sydney Coordinated Adaptive Traffic Systems (SCATS). Οι κυκλοφοριακοί φόρτοι και ο βαθμός

κορεσμού μετρήθηκαν τόσο για την μελετώμενη γραμμή, όσο και για τις διασταυρούμενες οδούς.

Επιπλέον για κάθε διασταύρωση χρησιμοποιήθηκαν κάποια προκαθορισμένα σύνολα (predefined aggregation period) κ λεπτών (2, 15, 30, 60) πριν την αναχώρηση του κάθε λεωφορείου από το ανάντη χρονικό σημείο.

Οι αντίστοιχες μεταβλητές όσον αφορά το Historical Data-Based Model ήταν η ώρα της ημέρας, η ημέρα της εβδομάδας, ο μήνας του χρόνου, καθώς και το κατά πόσο είναι εντός του χρονοδιαγράμματος.

Τέλος το Timetable-Based Model χρησιμοποιεί τους προγραμματισμένους χρόνους για να προβλέψει τους χρόνους διαδρομής.

Τα **συμπεράσματα** που διεξήχθησαν από την έρευνα ήταν ότι το Timetable-Based Model έδωσε τα πιο ανακριβή αποτελέσματα, ενώ το Traffic Flow Data-Based Model έδωσε τα πιο έγκυρα. Παρατηρήθηκε ότι τα δεδομένα της κίνησης μπορούν να προσεγγιστούν καλά και από χρονικές μεταβλητές, καθώς τα αποτελέσματα των Traffic Flow Data-Based Model και Historical Data-Based Model δεν διέφεραν αρκετά. Η μόνη περίπτωση που υπερέχει το Traffic Flow Data-Based Model, λόγω της ευαισθησίας του στις δυναμικές αλλαγές, είναι όταν παρουσιάζεται ξαφνικό κύμα ή αύξηση της επιβατικής κίνησης, όπου η χρήση του συνίσταται.

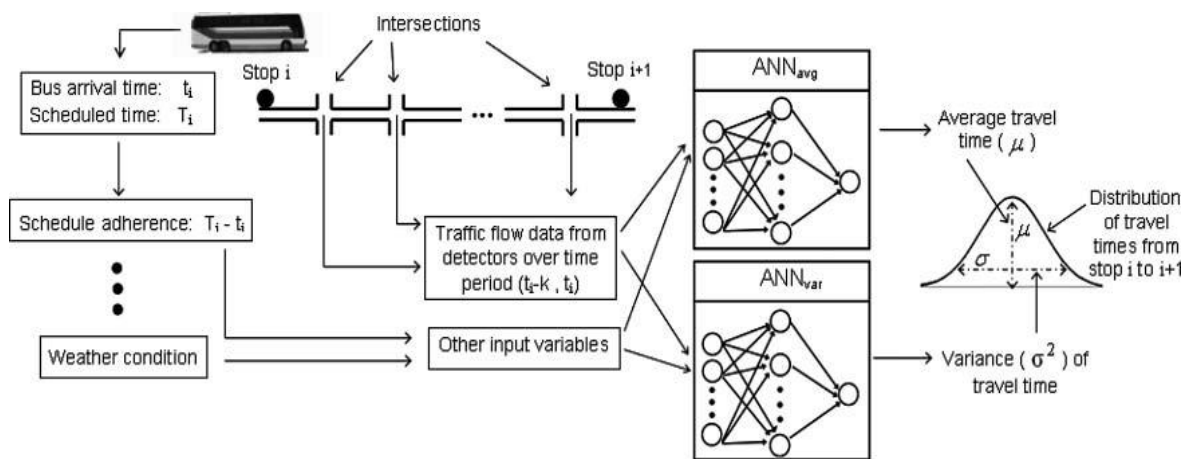
Το **2012** πραγματοποιήθηκε μια έρευνα από τους **Mazloumi E., Rose G., Currie G. & Sarvi M.** με τίτλο «An Integrated Framework to predict Bus Travel Time and its Variability Using Traffic Flow Data». **Σκοπός** της συγκεκριμένης έρευνας είναι να δώσει πληροφορίες για το μέσο χρόνο μετακίνησης με λεωφορεία και για τη μεταβλητότητά του.

Τα στοιχεία της βάσης δεδομένων, που χρησιμοποιήθηκαν στην παρούσα έρευνα, συλλέχθηκαν από τη λεωφορειακή γραμμή 246 στη Μελβούρνη της Αυστραλίας. Η γραμμή αυτή έχει μήκος 8χλμ. και αποτελείται από τέσσερα ισομήκη τμήματα και έχουν ορισθεί σε αυτήν πέντε χρονικά σημεία (timing points). Συλλέχθηκαν στοιχεία για περίοδο έξι μηνών μέσω συσκευών GPS και η συχνότητα των δρομολογίων της γραμμής ήταν ανά δέκα λεπτά στις ώρες αιχμής και ανά τριάντα τις υπόλοιπες.

Αρχικά, μελετήθηκε η μεταβλητότητα του χρόνου διαδρομής και ως παράγοντες που την επηρεάζουν θεωρήθηκαν η ζήτηση (demand) και η ικανότητα (capacity). Όπως φαίνεται και στο παρακάτω σχήμα οι παράγοντες που επηρεάζουν τη ζήτηση είναι οι κυκλοφοριακοί φόρτοι και η ζήτηση των επιβατών, ενώ για την ικανότητα ο καιρός, ένα πιθανό ατύχημα, η οδηγική συμπεριφορά του οδηγού λεωφορείου, τα χαρακτηριστικά

της διαδρομής κ.α. Στη συγκεκριμένη έρευνα μελετήθηκε η μεταβλητότητα των χρόνων από μέρα-σε-μέρα, η οποία μάλλον οφείλεται στις διακυμάνσεις των ανωτέρω παραγόντων.

Στη συνέχεια, έγινε η πρόβλεψη του χρόνου άφιξης του λεωφορείου με τη χρήση μοντέλων **Artificial Neural Network (ANN)**. Συγκεκριμένα αναπτύχθηκαν δυο μοντέλα, για την εύρεση των μέσων χρόνων διαδρομής **ANN_{avg}** και τη διακύμανση τους **ANN_{var}**, ανάμεσα σε δυο διαδοχικά χρονικά σημεία.



Πρόβλεψη χρόνων άφιξης λεωφορείου με μοντέλα ANN.

Έγινε η αξιολόγηση των μοντέλα που προέκυψαν. Το μοντέλο **ANN_{avg}** συγκρίθηκε με historical data - based models και timetable - based models. Τα **αποτελέσματα** της σύγκρισης έδειξαν ότι τα timetable-based models έδωσαν τα χειρότερα αποτελέσματα λόγω της σταθερότητας των προγραμματισμένων χρόνων διαδρομής και της μη επιρροής τους από τις δυναμικές αλλαγές, ζήτησης και κυκλοφορίας. Από την άλλη οι διαφορές μεταξύ των historical data- based models και του **ANN_{avg}** είναι ελάχιστες. Από την αξιολόγηση του μοντέλου **ANN_{var}** προέκυψε ότι οι διακυμάνσεις των χρόνων διαδρομής είναι εντός των επιτρεπτών ορίων για επίπεδο εμπιστοσύνης 95%.

2.3 Συμπεράσματα

Η πλειοψηφία αυτού του είδους της έρευνας έχει πραγματοποιηθεί σε μεγάλα αστικά κέντρα με αυξημένους φόρτους και έντονα προβλήματα συμφόρησης, όπως σε πόλεις της Κίνας και της Αυστραλίας. Για το λόγο αυτό είναι φανερό από την παραπάνω ανάλυση και παρουσίαση των σχετικών ερευνών που έχουν γίνει στο εξωτερικό, ότι είναι αναγκαίο να γίνει μια πιο ολοκληρωμένη προσπάθεια πάνω στην έρευνα της πρόβλεψης των χρόνων άφιξης των λεωφορείων στις στάσεις κάθε διαδρομής, έτσι

ώστε να παρουσιαστεί μια πιο αντιπροσωπευτική εικόνα των αναγκών των συγκοινωνιών στον Ελλαδικό χώρο.

Από την παρούσα βιβλιογραφική ανασκόπηση προκύπτει μια πληθώρα μεθόδων που χρησιμοποιούνται για την προσέγγιση του προβλήματος, τόσο στατιστικές, όσο και υπολογιστικές, με σκοπό τη χρήση των αποτελεσμάτων αυτών σε «έξυπνες εφαρμογές» που εξυπηρετούν στη βελτίωση των συγκοινωνιακών δικτύων (ITS) και στη διευκόλυνση των χρηστών τους. Σε όλες τις παραπάνω έρευνες την περιοχή μελέτης αποτέλεσε συγκεκριμένο δρομολόγιο λεωφορειακής γραμμής της πόλης, στο οποίο εξετάσθηκαν οι παράμετροι που πιθανόν να επηρεάζουν τους χρόνους διαδρομής. Οι βάσεις δεδομένων αποτελούνται από κυκλοφοριακά στοιχεία που συλλέχθηκαν από μετρητές, μετρήσεις διάρκειας ταξιδιών σε πραγματικούς χρόνους από συσκευές GPS τοποθετημένες πάνω στα λεωφορεία, ή συνδυασμό των δυο παραπάνω.

Ειδικότερα οι μελέτες που πραγματοποιήθηκαν στο Beijing της Κίνας και στη Μελβούρνη της Αυστραλίας είναι πιο κοντά στη φιλοσοφία της παρούσας διπλωματικής εργασίας, καθώς μελετούν μια σειρά παραμέτρων στο κατά πόσο επιδρούν στους χρόνους διαδρομής. Οι παράμετροι αυτοί έχουν να κάνουν με τις αποστάσεις μεταξύ των στάσεων, τη βροχόπτωση, τους κυκλοφοριακούς φόρτους, το χρόνο αναχώρησης από συγκεκριμένη στάση, των αριθμό των διασταυρώσεων.

Λόγω της έλλειψης τέτοιου είδους έρευνας στις λεωφορειακές γραμμές στον Ελλαδικό χώρο εμφανίζεται η αναγκαιότητα περαιτέρω μελέτης. Η παρούσα διπλωματική εργασία έχει ως σκοπό την όσο το δυνατό καλύτερη προσέγγιση του προβλήματος για την πόλη της Θεσσαλονίκης, που αποτελεί το δεύτερο μεγαλύτερο αστικό κέντρο της Ελλάδας.

	ΤΙΤΛΟΣ	ΟΝΟΜΑ	ΕΤΟΣ	ΣΚΟΠΟΣ	ΜΕΘΟΔΟΛΟΓΙΑ	INPUT DATA	OUTPUT DATA
1	Bus travel time prediction model with v-support vector regression	Wang, J.-N. , Chen, X.-M. , Guo, S.-X.	2009	Πρόβλεψη χρόνου άφιξης λεωφορείου στη στάση σε εξπρές γραμμή στην Κίνα. Γίνεται σύγκριση δύο μεθόδων για την εύρεση της πιο αποτελεσματικής.	v-support vector regression	Χρόνος αναχώρησης από συγκεκριμένη στάση, μήκος διαδρομής, αριθμός διασταυρώσεων, σηματοδότηση ή μη των κόμβων στις διασταυρώσεις	MAD ,MAPE ,MAE, MRE
					support vector machine model		
2	An Adaptive Bus Arrival Time Prediction Model	YU Bo, LU Jing, YU Bin, YANG Zhongzhen	2009	Πρόβλεψη χρόνου άφιξης λεωφορείου στη στάση σε γραμμή στην Κίνα. Γίνεται χρήση δυο μεθόδων, όπου στην πρώτη υπολογίζονται οι χρόνοι μεταξύ στάσεων βάσει των κυκλοφοριακών συνθηκών στο επόμενο κομμάτι, ενώ στη δεύτερη με βάση τους πραγματικούς χρόνους άφιξης.	Multiple Linear Regressions	Μήκος διαδρομής, εκτιμώμενος χρόνος διαδρομής, ταχύτητα	RMSE των δυο μεθόδων και σύγκριση με τις προγραμματισμένες
					Enhanced Regression Models (adaptive algorithm)		
3	Bus arrival time prediction at bus stop with multiple routes	Bin Yu, William H.K. Lam, Mei Lam Tam	2011	Πρόβλεψη χρόνων άφιξης λεωφορείου σε συγκεκριμένη στάση, μελέτη πολλαπλών διαδρομών. Υπολογισμός των χρόνων με τέσσερεις διαφορετικές μεθόδους και επιλογή της πιο αξιόπιστης .	Support vector machine (SVM) Artificial neural network (ANN) k-nearest neighbors algorithm (k-NN) Linear Regression (LR)	Χρονικό διάστημα για το σύνολο της διαδρομής, χρονικό διάστημα για την κάθε διαδρομή, σταθμισμένος μέσος όρος του χρόνου ταξιδιού του συνόλου της διαδρομής, χρόνος ταξιδιού για κάθε διαδρομή.	MAE, MAPE, RMSE

4	Exploring the value of traffic flow data in bus travel time prediction	Mazloumi, E., Moridpour, S., Currie, G., Rose, G.	2012	Αρχικά χρησιμοποιείται Regression Analysis για να βρεθούν οι στατιστικά σημαντικές μεταβλητές που θα εισαχθούν στα ANN models και στη συνέχεια εφαρμόζονται τα ANN-models.	Regression Analysis	Βροχόπτωση, τήρηση ή μη του χρονοδιαγράμματος, βαθμός κορεσμού σε προκαθορισμένες περιόδους σηματοδοτούμενων διασταυρώσεων.	RMSE των τριών μεθόδων και σύγκριση με τις προγραμματισμένες
					Artificial Neural Network (traffic flow data-based model, historical data-based model, timetable-based model)		
5	An integrated framework to predict bus travel time and its variability using traffic flow data	Mazloumi, E., Currie, G., Rose, G., Sarvi M.	2012	Στατιστική προσέγγιση για την πρόβλεψη του μέσου χρόνου άφιξης λεωφορείου σε στάση και της διακύμανσής του με χρήση τεχνητών νευρωνικών δικτύων	Artificial Neural Network (historical data-based model, timetable-based model)	Βροχόπτωση, τήρηση ή μη του χρονοδιαγράμματος, βαθμός κορεσμού σε προκαθορισμένες περιόδους σηματοδοτούμενων διασταυρώσεων.	RMSE των δυο μεθόδων και σύγκριση με τις προγραμματισμένες

3. ΠΕΡΙΓΡΑΦΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ

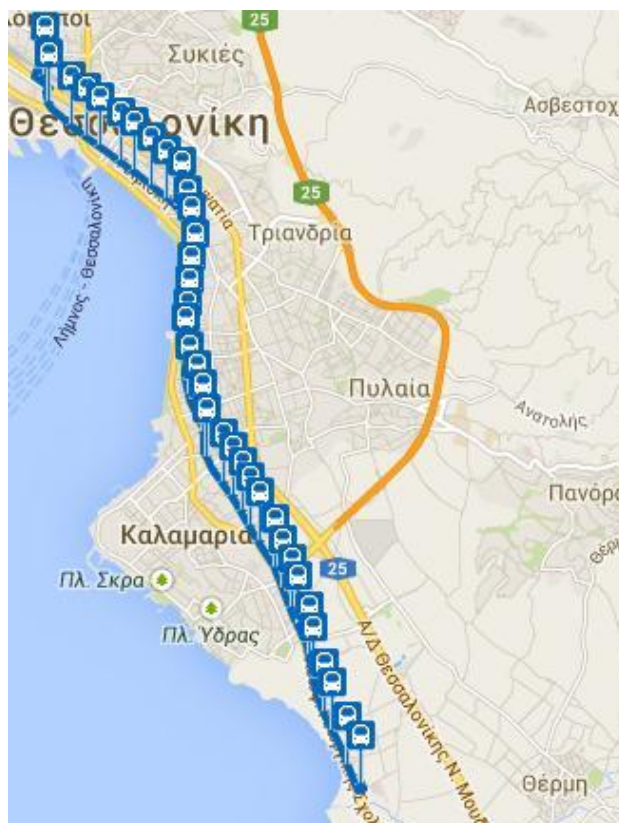
3.1 Περιγραφή βάσης δεδομένων

3.1.1 Εισαγωγή

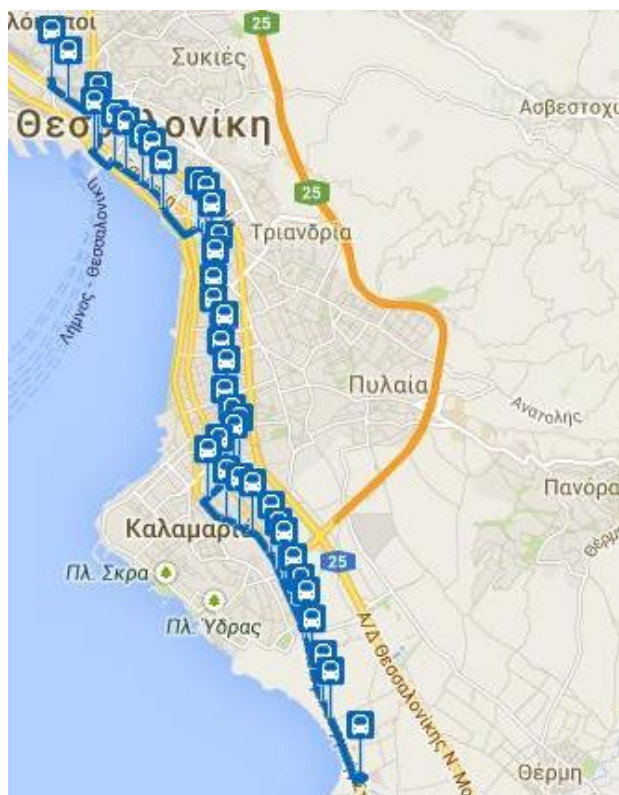
Η παρούσα διπλωματική εργασία μελετά τους χρόνους άφιξης και τις καθυστερήσεις των λεωφορείων στην εκάστοτε στάση, καθώς επίσης και τους παράγοντες που επιδρούν σε αυτά για τη λεωφορειακή γραμμή 3 στην πόλη της Θεσσαλονίκης. Το κεφάλαιο αυτό αφορά στη συλλογή και επεξεργασία των στοιχείων, με σκοπό να δώσει στον αναγνώστη μια πιο ολοκληρωμένη εικόνα για το μέγεθος του δείγματος, την ποιότητα και αξιοπιστία των στοιχείων.

Τη βασικότερη πηγή παροχής των δεδομένων της διπλωματικής εργασίας αποτέλεσε ο **Οργανισμός Αστικών Συγκοινωνιών Θεσσαλονίκης (ΟΑΣΘ)**. Έπειτα από επικοινωνία και συνεργασία με τους υπεύθυνους της αρμόδιας υπηρεσίας, παραχωρήθηκε μια βάση δεδομένων, με πληροφορίες μετρήσεων, για το Δρομολόγιο 3 Ανατολικός Σταθμός ΙΚΕΑ – Τερματικός Σταθμός Γ, σε πραγματικό χρόνο. Επιπρόσθετα, συλλέχθηκαν και κυκλοφοριακά στοιχεία από φωρατές εντός του κέντρου της πόλης, τα οποία προστέθηκαν στη βάση δεδομένων. Τα στοιχεία αυτά δόθηκαν σε ηλεκτρονική μορφή (Excel) και διαμορφώθηκαν κατάλληλα για να αποτελέσουν τη τελική προς μελέτη βάση δεδομένων.

Οι μετρήσεις αφορούν το έτος 2013 και συγκεκριμένα τους μήνες Ιανουαρίου και Φεβρουάριο. Αναλυτικότερα, τη βάση αποτέλεσαν μετρήσεις από επτά συνολικά ημέρες: 24, 28, 29, 30, 31 Ιανουαρίου και 7, 13 Φεβρουαρίου. Οι τρεις ημέρες με «βροχή» και τέσσερις με «ηλιοφάνεια». Επίσης, το μήκος της γραμμής 3 είναι 13,1 χλμ. κατά τη μετάβαση από την αφετηρία ως το τέρμα και 14,5 χλμ. κατά την επιστροφή και εξυπηρετεί περιοχές τόσο εκτός, όσο και εντός κέντρου. Παρακάτω φαίνονται δυο χάρτες του δρομολογίου 3 από την επίσημη ιστοσελίδα του ΟΑΣΘ, ο πρώτος αφορά τη μετάβαση και ο δεύτερος την επιστροφή.



Χάρτης 3.1 Μετάβαση Δρομολογίου 3



Χάρτης 3.2 Επιστροφή Δρομολογίου 3

3.1.2 Στοιχεία μετρήσεων σε πραγματικό χρόνο

Το πρώτο μέρος των στοιχείων που μας δόθηκαν από τον Οργανισμό Αστικών Συγκοινωνιών Θεσσαλονίκης περιείχε τα **Δελτία Κίνησης των Οχημάτων** για επτά συνολικά ημέρες σε πραγματικό χρόνο, τα οποία προέκυψαν από μετρήσεις συσκευών GPS τοποθετημένων πάνω στα οχήματα. Ύστερα από συνεννόηση με τους αρμόδιους της υπηρεσίας, αποφασίστηκε από το σύνολο των οχημάτων που εξυπηρετούν σε καθημερινή βάση τη συγκεκριμένη γραμμή να μελετηθεί το 50% αυτών, δηλαδή 10 λεωφορεία την ημέρα (από το σύνολο των 20), έτσι ώστε η βάση δεδομένων να είναι πιο εύχρηστη για τον μελετητή. Το σύνολο των οχημάτων της γραμμής 3 είναι αρθρωτά Mercedes χωρητικότητας 160 θέσεων.

Συγκεκριμένα, στο κάθε **Δελτίο Κίνησης Οχήματος** περιγράφονται τα εξής:

- Η ημερομηνία και η χρονική στιγμή της μέτρησης. Οι μετρήσεις έχουν νωρίτερη έναρξη στις 05:00 πμ. και αργότερη λήξη στις 02:00 μμ.
- Η κατάσταση στην οποία βρίσκεται το όχημα, η οποία μπορεί να είναι: Αναχώρηση από αμαξοστάσιο – Άφιξη σε αφετηρία – Αναχώρηση από αφετηρία – Άφιξη σε στάση – Άφιξη σε τέρμα – Αναχώρηση από τέρμα – Άφιξη σε αμαξοστάσιο.
- Ο κωδικός της στάσης.
- Η περιγραφή – ονοματολογία της στάσης.
- Η χρονική διάρκεια μετάβασης από την μία στάση στην επόμενη.
- Ο καιρός, δηλαδή αν έχουμε «βροχή» ή «ηλιοφάνεια».

Αρχικά, έγινε διαχωρισμός της ημέρας σε τέσσερις χρονικές περιόδους, καθώς οι διακυμάνσεις στους χρόνους άφιξης ήταν αισθητές από την μια περίοδο στην άλλη. Η ημέρα μοιράστηκε ως εξής:

1^η περίοδος: 05:00 – 11:00

2^η περίοδος: 12:00 – 15:00

3^η περίοδος: 16:00 – 19:00

4^η περίοδος: 20:00 – 02:00

Το γεγονός ότι η πρώτη και η τέταρτη περίοδος έχουν πολύ μεγαλύτερο εύρος από τις άλλες οφείλεται στο ότι τις πολύ πρωινές και βραδινές ώρες τα δρομολόγια έχουν μικρές συχνότητες και έτσι ικανοποιείται η απαίτηση η κάθε περίοδος να εκφράζει περίπου το ίδιο μέγεθος δείγματος. Τέλος, παρατηρήθηκε ότι κατά τις βραδινές ώρες μειώνεται ο αριθμός των στάσεων που πραγματοποιεί το

κάθε λεωφορείο πιθανόν λόγω της μειωμένης ζήτησης και συχνά παραλείπεται η στάση ΒΙΑΜΥΛ.

Επίσης στα αρχικά στοιχεία απεικονίζονταν το σύνολο των στάσεων κατά τη μετάβαση και την επιστροφή, οι αποστάσεις από την αρχή και το συνολικό μήκος κάθε διαδρομής. Στους κάτωθι πίνακες αναγράφονται αναλυτικά το σύνολο των στάσεων και οι αποστάσεις από την αρχή (πίνακας 3.1), καθώς και τα χαρακτηριστικά ενός Επίσημου Δελτίου Κίνησης Οχήματος από τον ΟΑΣΘ (πίνακας 3.2).

Πίνακας 3.1 Στάσεις μετάβασης και επιστροφής με αποστάσεις από αρχή

ΜΕΤΑΒΑΣΗ		ΕΠΙΣΤΡΟΦΗ	
Περιγραφή Στάσης	Km	Τ.Σ. ΣΤΑΘΜΟΣ Γ	0,00
ΑΝΑΤΟΛΙΚΟΣ ΣΤΑΘΜΟΣ ΙΚΕΑ Α	0,00	ΖΩΓΡΑΦΟΥ	524,19
ΙΚΕΑ	631,07	ΔΗΜΟΚΡΙΤΟΣ	1288,68
Σ.Α.Σ.Θ	1203,53	ΛΙΜΑΝΙ	1551,88
ΒΙΑΜΥΛ	1515,85	ΒΕΝΙΖΕΛΟΥ	2100,99
ΚΑΛΑΜΑΡΙ	2100,50	ΠΛΑΤΕΙΑ ΑΡΙΣΤΟΤΕΛΟΥΣ	2310,80
ΚΛΕΙΣΤΟ ΓΗΠΕΔΟ ΜΠΑΣΚΕΤ ΠΥΛΑΙΑΣ	2418,98	ΜΗΤΡΟΠΟΛΗ	2624,96
ΠΡΟΝΟΙΑ	2916,87	ΔΙΑΓΩΝΙΟΣ	2827,82
ΑΓΟΡΑ	3184,92	ΛΕΥΚΟΣ ΠΥΡΓΟΣ	3190,41
ΑΜΑΞΟΣΤΑΣΙΟ	3518,95	ΜΟΥΣΕΙΟ ΒΥΖΑΝΤΙΝΟΥ ΠΟΛΙΤΙΣΜΟΥ	3954,88
ΝΟΣΟΚΟΜΕΙΟ ΑΓΙΟΣ ΠΑΥΛΟΣ	3850,01	ΣΤΡΑΤΗΓΕΙΟ	4167,64
ΣΤΡΑΤΟΠΕΔΟ	4322,09	ΕΥΖΩΝΩΝ-ΓΑΛΛΙΚΟ ΙΝΣΤΙΤΟΥΤΟ	4456,82
ΑΓΙΟΣ ΠΑΝΤΕΛΕΗΜΩΝ	4592,76	ΦΑΛΗΡΟ	4966,48
ΠΕΡΙΠΤΕΡΟ	4855,15	ΚΑΡΑΙΣΚΑΚΗ	5213,48
ΚΡΙΚΕΛΑ	5131,98	ΜΕΤΑΜΟΡΦΩΣΗ ΤΟΥ ΣΩΤΗΡΟΣ	5658,07
ΒΥΖΑΝΤΙΟ	5380,68	ΜΠΟΤΣΑΡΗ	5969,62
ΠΕΡΙΦΕΡΕΙΑΚΗ ΕΝΟΤΗΤΑ ΘΕΣΣΑΛΟΝΙΚΗΣ	5837,98	ΑΝΑΛΗΨΗ	6444,04
ΓΕΦΥΡΑ	6235,02	ΓΕΩΡΓΙΟΥ	6908,75
ΖΗΣ ΜΑΡΤΙΟΥ	6522,80	ΖΗΣ ΜΑΡΤΙΟΥ	7204,35
ΓΕΩΡΓΙΟΥ	6847,63	ΑΓΙΟΣ ΕΛΕΥΘΕΡΙΟΣ	7740,64
ΑΝΑΛΗΨΗ	7286,12	ΚΗΦΙΣΙΑΣ	8111,22
ΜΠΟΤΣΑΡΗ	7521,19	ΒΡΥΣΑΚΙ	8291,11
ΛΑΟΓΡΑΦΙΚΟ ΜΟΥΣΕΙΟ	7883,95	ΑΙΓΑΙΟΥ	8472,67
ΣΧΟΛΗ ΤΥΦΛΩΝ	8237,06	ΘΕΡΜΑΙΚΟΥ	8762,67
ΦΑΛΗΡΟ	8581,09	ΒΥΖΑΝΤΙΟ	9036,44
ΕΥΖΩΝΩΝ	8995,68	ΚΡΙΚΕΛΑ	9561,06
ΔΗΜΑΡΧΙΑΚΟ ΜΕΓΑΡΟ	9260,07	ΑΓΙΟΣ ΠΑΝΤΕΛΕΗΜΩΝ	9813,45
Χ.Α.Ν.Θ.	9713,45	ΣΤΡΑΤΟΠΕΔΟ 1	10023,09
ΚΕΝΤΡΟ ΙΣΤΟΡΙΑΣ ΘΕΣΣΑΛΟΝΙΚΗΣ	10002,45	ΣΤΡΑΤΟΠΕΔΟ 2	10473,96
ΔΙΑΓΩΝΙΟΣ	10302,76	ΝΟΣΟΚΟΜΕΙΟ ΑΓΙΟΣ ΠΑΥΛΟΣ	10764,29
ΑΓΙΑΣ ΣΟΦΙΑΣ	10663,51	ΑΜΑΞΟΣΤΑΣΙΟ	10944,20
ΠΛΑΤΕΙΑ ΑΡΙΣΤΟΤΕΛΟΥΣ	10921,11	ΑΓΟΡΑ	11380,42
ΠΛΑΤΕΙΑ ΕΜΠΟΡΙΟΥ	11316,42	ΠΡΟΝΟΙΑ	11703,84
ΔΙΟΙΚΗΤΙΚΑ ΔΙΚΑΣΤΗΡΙΑ	11558,74	ΚΛΕΙΣΤΟ ΓΗΠΕΔΟ ΜΠΑΣΚΕΤ ΠΥΛΑΙΑΣ	11931,43
Ι.Κ.Α	11835,44	ΚΑΛΑΜΑΡΙ	12381,21
ΑΝΑΓΕΝΝΗΣΕΩΣ	12356,66	ΒΙΑΜΥΛ	12944,24
Τ.Σ. ΣΤΑΘΜΟΣ Γ	13122,86	Σ.Α.Σ.Θ	13250,07
		ΑΝΑΤΟΛΙΚΟΣ ΣΤΑΘΜΟΣ ΙΚΕΑ Α	14491,48

Πίνακας 3.2 Δελτίο Κίνησης Οχήματος

A	B	C	D	E	F	G	H	I	J	K	L
	3/8/2013				ΔΕΛΤΙΟ ΚΙΝΗΣΗΣ ΟΧΗΜΑΤΟΣ						
					Όχημα 0771 Από: 7/2/2013 04:00:00 Έως: 8/2/2013 02:00:00						
Ημ/νία	Κατάσταση	Κωδ. Στάσης	Περιγραφή Στάσης	Διάρκει	Γραμμή						
07/02/2013 07:24:57	Αναχώρηση από Αμαξοστάσιο	Γ-11/004ΤΣ	ΑΜΑΞΟΣΤΑΣΙΟ ΦΟΙΝΙΚΑ		03						
07/02/2013 07:32:23	Αφίξη σε Αφετηρία	Ι-05/001 Α	ΑΝΑΤΟΛΙΚΟΣ ΣΤΑΘΜΟΣ ΙΚΕΑ Α	+00:07:26	03						
07/02/2013 07:51:26	Αναχώρηση από Αφετηρία	Ι-05/001 Α	ΑΝΑΤΟΛΙΚΟΣ ΣΤΑΘΜΟΣ ΙΚΕΑ Α	+00:19:03	03						
07/02/2013 07:51:52	Αφίξη σε Στάση	Ι-05/219	ΙΚΕΑ	+00:00:26							
07/02/2013 07:52:30	Αφίξη σε Στάση	Ι-05/220	Σ.Α.Σ.Θ	+00:00:38							
07/02/2013 07:53:10	Αφίξη σε Στάση	Ι-05/221	ΒΙΑΜΥΛ	+00:00:40							
07/02/2013 07:54:56	Αφίξη σε Στάση	Γ-12/008	ΚΑΛΑΜΑΡΙ	+00:01:46							
07/02/2013 07:55:36	Αφίξη σε Στάση	Γ-12/006	ΚΛΕΙΣΤΟ ΓΗΠΕΔΟ ΜΠΑΣΚΕΤ ΠΥΛΑΙΑΣ	+00:00:40							
07/02/2013 07:57:08	Αφίξη σε Στάση	Γ-12/003	ΠΡΟΝΟΙΑ	+00:01:32							
07/02/2013 07:58:04	Αφίξη σε Στάση	Γ-11/005	ΑΓΟΡΑ	+00:00:56							
07/02/2013 07:58:48	Αφίξη σε Στάση	Γ-11/004	ΑΜΑΞΟΣΤΑΣΙΟ	+00:00:44							
07/02/2013 07:59:34	Αφίξη σε Στάση	Γ-11/002	ΝΟΣΟΚΟΜΕΙΟ ΑΓΙΟΣ ΠΑΥΛΟΣ	+00:00:46							
07/02/2013 08:00:28	Αφίξη σε Στάση	Γ-11/006	ΣΤΡΑΤΟΠΕΔΟ	+00:00:54							
07/02/2013 08:01:20	Αφίξη σε Στάση	Γ-10/022	ΑΓΙΟΣ ΠΑΝΤΕΛΕΗΜΩΝ	+00:00:52							
07/02/2013 08:02:18	Αφίξη σε Στάση	Γ-10/021	ΠΕΡΙΠΤΕΡΟ	+00:00:58							
07/02/2013 08:03:26	Αφίξη σε Στάση	Γ-10/020	ΚΡΙΚΕΛΑ	+00:01:08							
07/02/2013 08:04:42	Αφίξη σε Στάση	Γ-10/019	ΒΥΖΑΝΤΙΟ	+00:01:16							
07/02/2013 08:05:58	Αφίξη σε Στάση	Γ-09/013	ΠΕΡΙΦΕΡΕΙΑΚΗ ΕΝΟΤΗΤΑ	+00:01:16							
07/02/2013 08:07:38	Αφίξη σε Στάση	Γ-09/014	ΓΕΦΥΡΑ	+00:01:40							
07/02/2013 08:08:26	Αφίξη σε Στάση	Γ-09/015	25ΗΣ ΜΑΡΤΙΟΥ	+00:00:48							
07/02/2013 08:09:52	Αφίξη σε Στάση	Γ-09/016	ΓΕΩΡΓΙΟΥ	+00:01:26							
07/02/2013 08:11:42	Αφίξη σε Στάση	Γ-08/001	ΑΝΔΑΛΗΣΗ	+00:01:50							
07/02/2013 08:13:06	Αφίξη σε Στάση	Γ-08/002	ΜΠΟΤΣΑΡΗ	+00:01:24							
07/02/2013 08:14:52	Αφίξη σε Στάση	Γ-08/003	ΛΑΟΓΡΑΦΙΚΟ ΜΟΥΣΕΙΟ	+00:01:46							
07/02/2013 08:16:16	Αφίξη σε Στάση	Γ-08/004	ΣΧΟΛΗ ΤΥΦΛΩΝ	+00:01:24							
07/02/2013 08:18:02	Αφίξη σε Στάση	Γ-08/005	ΦΑΛΗΡΟ	+00:01:46							
07/02/2013 08:19:56	Αφίξη σε Στάση	Β-07/001	ΕΥΣΩΝΩΝ	+00:01:54							
07/02/2013 08:20:34	Αφίξη σε Στάση	Β-07/057	ΔΗΜΑΡΧΙΑΚΟ ΜΕΓΑΡΟ	+00:00:38							
07/02/2013 08:22:10	Αφίξη σε Στάση	Β-07/047	Χ.Α.Ν.Θ.	+00:01:36							
07/02/2013 08:23:10	Αφίξη σε Στάση	Β-07/048	ΚΕΝΤΡΟ ΙΣΤΟΡΙΑΣ ΘΕΣΣΑΛΟΝΙΚΗΣ	+00:01:00							

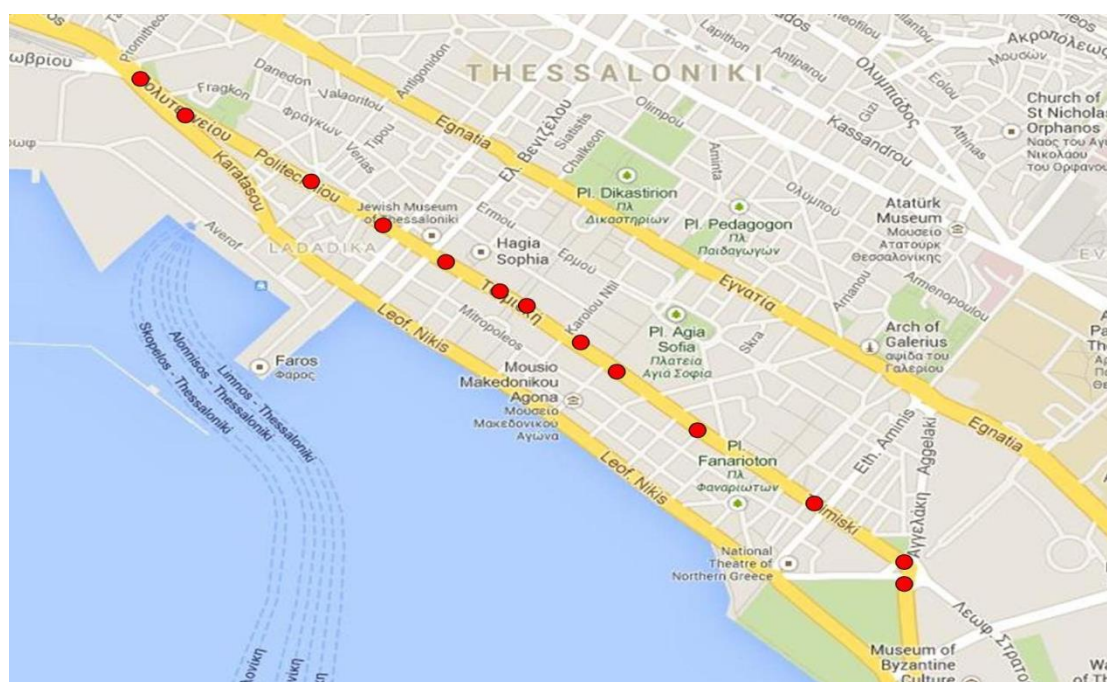
3.1.3 Κυκλοφοριακά Στοιχεία

Το δεύτερο σκέλος των στοιχείων της βάσης δεδομένων αποτέλεσαν τα κυκλοφοριακά στοιχεία. Ύστερα από συνεννόηση για τις επιλεγμένες ημέρες όπου πραγματοποιείται η μελέτη, μας παραχωρήθηκαν οι μετρήσεις των φόρτων και των ταχυτήτων του συνόλου των μετρητών κυκλοφορίας της Θεσσαλονίκης. Οι μετρητές αυτοί είναι εγκατεστημένοι στο κέντρο της πόλης, όπου εντείνεται και το πρόβλημα της κυκλοφοριακής συμφόρησης και οι καθυστερήσεις αυξάνονται. Οι μετρήσεις των εν λόγω ανιχνευτών κυκλοφορίας πραγματοποιούνται ανά πέντε λεπτά, ξεκινούν στις 05:00 πμ. και τελειώνουν στις 02:00 πμ. και οι τιμές τόσο του φόρτου, όσο και της ταχύτητας είναι ανηγμένες στην ώρα. Ο φόρτος δίνεται σε **οχήματα/ώρα** και η ταχύτητα σε **χιλιόμετρα/ώρα**.

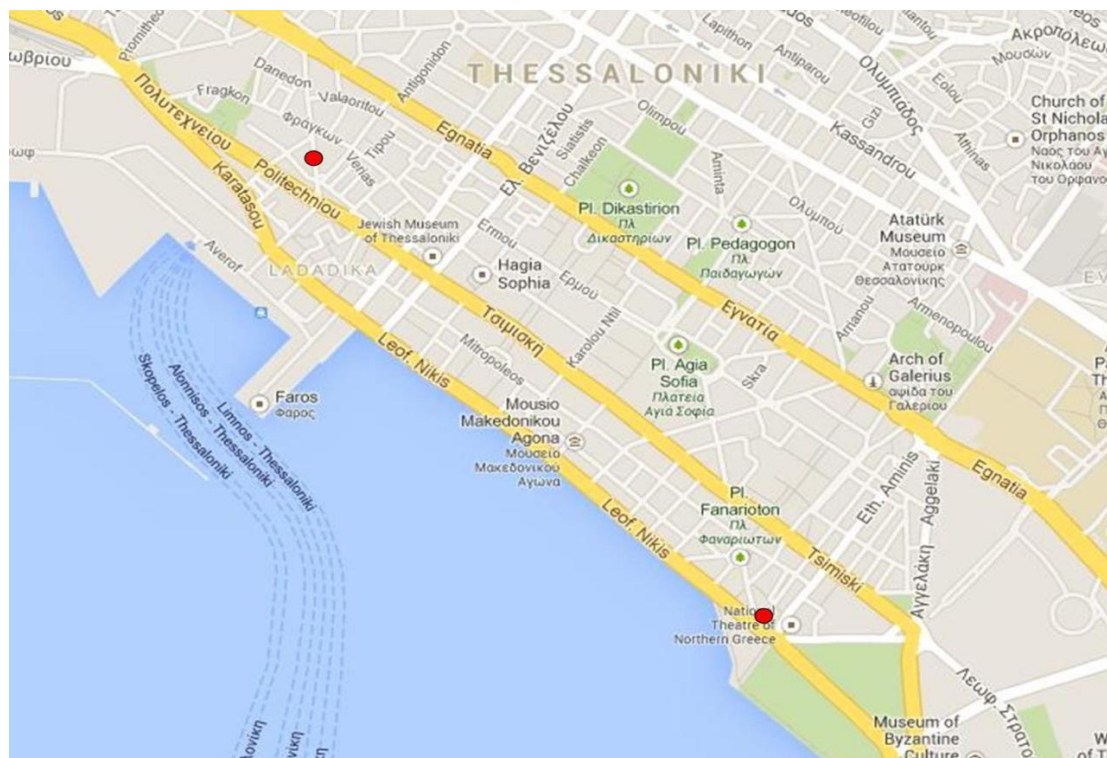
Η επιλογή των μετρητών κυκλοφορίας έγινε με μόνο κριτήριο αυτοί να βρίσκονται επάνω στη διαδρομή της γραμμής 3, κατά τη μετάβαση ή κατά την επιστροφή. Εδώ πρέπει να σημειωθεί ότι από το σύνολο των τριάντα τεσσάρων ανιχνευτών κίνησης που είναι εγκατεστημένοι στην πόλη της Θεσσαλονίκης εξυπηρετούσαν μόνο οι δεκαπέντε, εκ των οποίων οι δεκατρείς βρίσκονται στη

μετάβαση της διαδρομής και οι δυο στην επιστροφή. Επίσης, το κομμάτι της διαδρομής 3 για το οποίο μας παρέχουν πληροφορίες φόρτου-ταχύτητας είναι 2,7 χλμ. από το σύνολο των 13,1 χλμ. της μετάβασης και 2,1 χλμ. από το σύνολο των 14,5 χλμ. της επιστροφής. Από τα δοθέντα στοιχεία ιδιαίτερα αυξημένοι ήταν φόρτοι κατά τη μετάβαση, Ανατολικός Σταθμός ΙΚΕΑ – Τερματικός Σταθμός, όπου το λεωφορείο διασχίζει τη Τσιμισκή συγκρινόμενοι με την επιστροφή μέσω της Μητροπόλεως. Ακόμη, έχουμε ιδιαίτερη μείωση των φόρτων, τουλάχιστον στο 50% στην 4^η περίοδο της ημέρας. Τέλος, διαπιστώνεται πως οι ταχύτητες δεν παρουσιάζουν αύξηση όσο μειώνονται οι φόρτοι και διατηρούνται σε σχετικά σταθερά επίπεδα.

Στους παρακάτω χάρτες της πόλης της Θεσσαλονίκης φαίνονται με κόκκινη κουκίδα οι ανιχνευτές κυκλοφορίας που έχουν χρησιμοποιηθεί στη παρούσα μελέτη κατά τη μετάβαση και την επιστροφή του Δρομολογίου 3.



Χάρτης 3.3 Ανιχνευτές Κυκλοφορίας κατά τη Μετάβαση του Δρομολογίου 3



Χάρτης 3.4 Ανιχνευτές Κυκλοφορίας κατά την Επιστροφή του Δρομολογίου 3

Το γεγονός ότι οι φόρτοι και οι ταχύτητες αφορούν μόνο το κέντρο οδήγησε στη δημιουργία μιας ξεχωριστής βάσης δεδομένων που αναφέρεται μόνο τις στάσεις εντός κέντρου των οποίων οι φόρτοι και οι ταχύτητες μπορούν να εκφραστούν από τις μετρήσεις των φωρατών. Η βάση αυτή εξετάσθηκε ανεξάρτητα ως προς τα χαρακτηριστικά που επιδρούν στους χρόνους και τις καθυστερήσεις του δρομολογίου.

3.2 Επεξεργασία Βάσης Δεδομένων

Με σκοπό την εξαγωγή των τελικών αρχείων excel για την στατιστική μελέτη, πραγματοποιήθηκαν κάποιες διορθώσεις, αλλαγές και παραδοχές στις αρχικές βάσεις δεδομένων για την εξάλειψη λαθών και ακραίων αποτελεσμάτων. Επιπλέον πραγματοποιήθηκαν κάποιοι υπολογισμοί για τη δημιουργία επιπλέον στοιχείων που συμπεριλήφθηκαν στην τελική βάση δεδομένων. Επεξεργάστηκαν τόσο τα στοιχεία των Δελτίων Κίνησης Οχημάτων, όσο και τα κυκλοφοριακά στοιχεία με τρόπο που εξηγείται παρακάτω.

3.2.1 Επεξεργασία στοιχείων μετρήσεων σε πραγματικό χρόνο

Η πρώτη επεξεργασία του δείγματος πραγματοποιήθηκε με τα στοιχεία όπως ακριβώς δόθηκαν. Στα αποτελέσματα που προέκυψαν από τα περιγραφικά στατιστικά εμφανίστηκαν κάποιες ακραίες μέγιστες και ελάχιστες τιμές οι οποίες δεν είχαν φυσική σημασία. Τέτοιου είδους τιμές οφείλονται σε σφάλματα των μηχανημάτων και αλλοιώνουν πολύ τα αποτελέσματα της μελέτης για αυτό και αφαιρέθηκαν από τη βάση δεδομένων. Η διαδικασία που ακολουθήθηκε περιγράφεται αναλυτικά παρακάτω.

Αρχικά μελετήθηκαν οι ελάχιστοι και οι μέγιστοι χρόνοι άφιξης που παρατηρήθηκαν. Οι ελάχιστοι χρόνοι συναντήθηκαν κυρίως κατά τη μετάβαση από την πρώτη στη δεύτερη στάση, δηλαδή κατά την έναρξη των μετρήσεων, και οφείλονταν στην καθυστέρηση της συσκευής προσδιορισμού θέσης (GPS) να στείλει το σήμα. Θεωρώντας ως μέση ταχύτητα λεωφορείου 9-13 χλμ./ ώρα (Παπαϊωάννου, 2010) χρόνοι μεταξύ στάσεων μικρότεροι των 16 δευτερολέπτων αφαιρέθηκαν από το δείγμα. Οι τιμές αυτές αντικαταστάθηκαν με νέες μέσες τιμές οι οποίες διαφοροποιούνταν ανά στάση/ ημέρα/ περίοδο. Οι μέγιστοι χρόνοι που παρατηρήθηκαν ήταν σε περιπτώσεις που το λεωφορείο δεν πραγματοποιούσε τις αναμενόμενες στάσεις σε κάθε σταθμό, γεγονός που έδινε λανθασμένα αποτελέσματα. Στην περίπτωση αυτή αφαιρέθηκαν όλες οι μετρήσεις με απόσταση μεγαλύτερη των 1000 μέτρων (εκτός της τελευταίας στάσης κατά την επιστροφή όπου η απόσταση είναι 1241,41 μέτρα). Επιπλέον όσον αφορά τους μέγιστους χρόνους, δεδομένου ότι οι μέσες αποστάσεις μεταξύ στάσεων είναι 375 και 392 μέτρα κατά τη μετάβαση και την επιστροφή αντίστοιχα, αφαιρέθηκαν όλοι οι χρόνοι μεγαλύτεροι των 360 δευτερολέπτων με τη θεώρηση μιας ελάχιστης ταχύτητας 4 χλμ./ ώρα σε ώρες αιχμής.

Ένα ακόμα λάθος που παρατηρήθηκε ήταν μερικές επιπλέον διαδρομές που πραγματοποιούσαν κάποια από τα οχήματα, που εξυπηρετούσαν και άλλα δρομολόγια. Όπου εμφανίστηκαν αυτές οι επιπλέον διαδρομές, αφαιρέθηκαν από το δείγμα.

Αφού έγιναν οι απαραίτητες διορθώσεις στο αρχικό δείγμα, πραγματοποιήθηκαν κάποιοι υπολογισμοί για τη διεξαγωγή χρήσιμων για τη μελέτη στοιχείων. Αρχικά υπολογίστηκαν οι αποστάσεις μεταξύ των στάσεων, καθώς οι χρόνοι άφιξης αφορούσαν τον πραγματικό χρόνο από στάση-σε-στάση. Στη συνέχεια, υπολογίστηκαν οι μέσες ταχύτητες ανά στάση για τη συνολική διαδρομή του λεωφορείου. Επιπρόσθετα υπολογίστηκαν οι αναμενόμενες καθυστερήσεις. Η διαδικασία που ακολουθήθηκε για τον υπολογισμό των καθυστερήσεων ήταν να ομαδοποιηθούν οι χρόνοι άφιξης ανά στάση/ ημέρα/ περίοδο, να βρεθεί ο μέσος

όρος αυτών και να υπολογιστεί η διαφορά χρόνου-μέσου χρόνου για κάθε στάση. Η φυσική σημασία των αρνητικών τιμών είναι η νωρίτερη άφιξη, των θετικών τιμών η καθυστέρηση και των μηδενικών τιμών η άφιξη στον αναμενόμενο χρόνο. Καθώς οι αρνητικές τιμές δεν έχουν κάποιο νόημα όταν μιλάμε για καθυστέρηση, οι τιμές τους αντικαταστάθηκαν με 0 στην τελική βάση δεδομένων. Τέλος στη βάση δεδομένων προστέθηκε από οδικούς χάρτες της Θεσσαλονίκης ο αριθμός των διασταυρώσεων μεταξύ των στάσεων.

3.2.2 Επεξεργασία Κυκλοφοριακών Στοιχείων

Όπως αναφέρθηκε προηγουμένως το δεύτερο σκέλος της βάσης δεδομένων αποτέλεσαν τα κυκλοφοριακά στοιχεία, οι φόρτοι και οι ταχύτητες στο κέντρο της πόλης ανηγμένες στην ώρα. Αρχικά οι στάσεις του κέντρου ορίστηκαν χωροταξικά, έτσι ώστε οι μέσοι φόρτοι και ταχύτητες των φωρατών κυκλοφορίας να τις εκφράζουν. Κατά τη μετάβαση επιλέχθηκαν οι στάσεις από το Δημαρχιακό Μέγαρο έως τη στάση Ι.Κ.Α και αντιστοιχίστηκαν στους δεκατρείς μετρητές κυκλοφορίας όπως φαίνεται στον παρακάτω πίνακα (Πίνακας 3.3). Η απόσταση μεταξύ στάσης και φωρατή δεν υπερβαίνει σε καμία περίπτωση τα 150 μέτρα.

Κατά την επιστροφή επιλέχθηκαν οι στάσεις από το Δημόκριτο έως το Λευκό Πύργο και αντιστοιχίστηκαν στους δύο φωρατές που είναι εντός της διαδρομής (Πίνακας 3.3). Εδώ λόγω των μειωμένων φωρατών οι αποστάσεις που έχουν θεωρηθεί είναι αρκετά μεγαλύτερες, με μέγιστη τιμή τα 1000 μέτρα ανάμεσα στη στάση Βενιζέλου και το μετρητή που βρίσκεται επί της Δωδεκανήσου.

Μετά την αντιστοίχιση των φωρατών κυκλοφορίας στις στάσεις επεξεργάστηκαν τα στοιχεία των μετρήσεων για να εισαχθούν στη βάση. Αυτό έγινε ως εξής, οι αρχικές μετρήσεις οι οποίες πραγματοποιούνταν ανά πέντε λεπτά της ώρας ομαδοποιήθηκαν ανά ημέρα/ περίοδο/ στάση, ύστερα υπολογίστηκαν οι μέσοι όροι των μετρήσεων και οι τιμές αυτές φόρτου (οχήματα/ώρα) και ταχύτητας (χλμ./ώρα) εισήχθησαν στη βάση δεδομένων.

Οι συγκεκριμένες στάσεις τόσο στη μετάβαση όσο και στην επιστροφή απεικονίζονται στον Πίνακα 3.3 και είναι αριθμημένες από 1-16, οι αντίστοιχοι φωρατές κυκλοφορίας και οι θέσεις τους ως προς τις στάσεις φαίνονται μέσα σε αστερίσκους **.

Πίνακας 3.3 Στάσεις εντός κέντρου και αποστάσεις από αντίστοιχους φωρατές κυκλοφορίας

ΜΕΤΑΒΑΣΗ		ΕΠΙΣΤΡΟΦΗ	
ΜΑΝΩΛΗ ΑΝΔΡΟΝΙΚΟΥ	10m	*ΔΩΔΕΚΑΝΗΣΟΥ*	190m
1) ΔΗΜΑΡΧΙΑΚΟ ΜΕΓΑΡΟ		10) ΔΗΜΟΚΡΙΤΟΣ	450 m
ΑΓΓΕΛΑΚΗ & ΔΕΣΠΑΙΡΕ	55m	11) ΛΙΜΑΝΙ	1000m
2) Χ.Α.Ν.Θ.		12) ΒΕΝΙΖΕΛΟΥ	
ΕΘΝΙΚΗΣ ΑΜΥΝΑΣ & ΦΙΛ.ΕΤΑΙΡΕΙΑΣ	70m	13) ΠΛΑΤΕΙΑ ΑΡΙΣΤΟΤΕΛΟΥΣ	800m
3) ΚΕΝΤΡΟ ΙΣΤΟΡΙΑΣ ΘΕΣΣΑΛΟΝΙΚΗΣ		14) ΜΗΤΡΟΠΟΛΗ	450m
ΠΑΥΛΟΥ ΜΕΛΑ & ΠΑΛΑΜΑ	40m	15) ΔΙΑΓΩΝΙΟΣ	270m
4) ΔΙΑΓΩΝΙΟΣ		*ΠΑΥΛΟΥ ΜΕΛΑ*	100m
ΙΚΤΙΝΟΥ & ΑΓΙΑΣ ΣΟΦΙΑΣ	150m	16) ΛΕΥΚΟΣ ΠΥΡΓΟΣ	
ΑΓΙΑΣ ΣΟΦΙΑΣ & ΝΤΗΛ	60m		
5) ΑΓΙΑΣ ΣΟΦΙΑΣ	120m		
ΝΤΗΛ & ΑΡΙΣΤΟΤΕΛΟΥΣ			
ΑΡΙΣΤΟΤΕΛΟΥΣ	73m		
6) ΠΛΑΤΕΙΑ ΑΡΙΣΤΟΤΕΛΟΥΣ	120m		
ΚΟΜΝΗΝΩΝ & ΒΕΝΙΖΕΛΟΥ			
Ι. ΔΡΑΓΟΥΜΗ & ΚΑΤΟΥΝΗ	100m		
7) ΠΛΑΤΕΙΑ ΕΜΠΟΡΙΟΥ			
ΤΣΙΜΙΣΚΗ & ΔΩΔΕΚΑΝΗΣΟΥ	140m		
8) ΔΙΟΙΚΗΤΙΚΑ ΔΙΚΑΣΤΗΡΙΑ			
ΤΣΙΜΙΣΚΗ ΠΡΟ ΦΡΑΓΚΩΝ	35m		
9) Ι.Κ.Α	100m		
ΣΤΑΘΜΟΥ ΠΡΟ ΔΙΚΑΣΤΗΡΙΩΝ			

3.3 Περιγραφικά στατιστικά χαρακτηριστικά βάσης δεδομένων

Στο παρών υποκεφάλαιο περιγράφεται η αρχική Στατιστική Επεξεργασία των δεδομένων της βάσης και η ανάπτυξη μεθόδων για τη συνοπτική και αποτελεσματική παρουσίαση των δεδομένων αυτών. Αναλύοντας τη βάση προκύπτουν διάφορα συμπεράσματα ιδιαίτερου ενδιαφέροντος και χρησιμότητας. Παρακάτω απεικονίζονται τα στοιχεία πινακοποιημένα και γραφικά, ενώ ταυτόχρονα παρατίθενται αριθμητικά περιγραφικά μέτρα για την ευχερέστερη κατανόηση των αποτελεσμάτων. Το δείγμα μελετήθηκε αρχικά συνολικά και στη συνέχεια μόνο για την περιοχή του κέντρου, τόσο για τη μετάβαση όσο και για την επιστροφή.

Βασικές έννοιες στατιστικής θεωρίας:

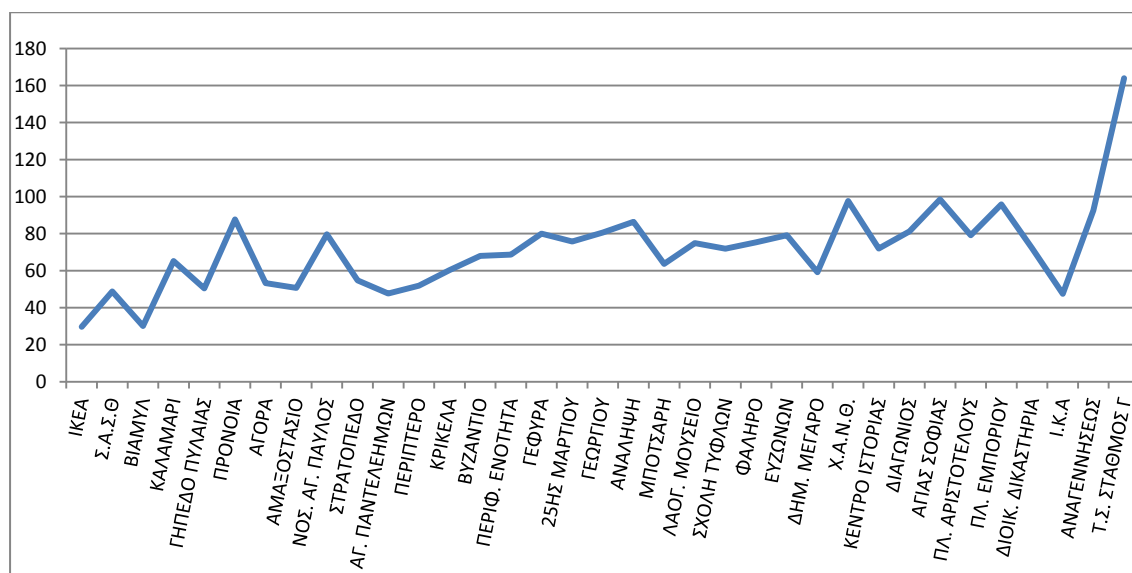
- **Στατιστικός πληθυσμός ή Πληθυσμός:** Ονομάζεται ένα σύνολο στοιχείων που είναι τελείως καθορισμένα, τα οποία εξετάζουμε ως προς ένα ή περισσότερα χαρακτηριστικά τους. Τα στοιχεία του πληθυσμού ονομάζονται μονάδες ή άτομα.
- **Δείγμα:** είναι ένα υποσύνολο του στατιστικού πληθυσμού με σκοπό τη συλλογή κάποιων παρατηρήσεων. Έτσι, όλα τα στοιχεία που ανήκουν στο δείγμα ανήκουν και στον πληθυσμό, χωρίς να ισχύει το αντίστροφο. Το δείγμα θα πρέπει να είναι αντιπροσωπευτικό του πληθυσμού, έτσι ώστε τα αποτελέσματα που θα προκύψουν από τη μελέτη του να είναι ικανοποιητικά. Προφανώς η αντιπροσωπευτικότητα ή όχι του δείγματος εξαρτάται από τον τρόπο συλλογής των μονάδων (Κοκολάκης, Σπηλιώτης, 1999).
- **Μεταβλητές:** τα χαρακτηριστικά ή οι ιδιότητες ως προς τα οποία εξετάζουμε ένα πληθυσμό ονομάζονται μεταβλητές (variables). Οι δυνατές τιμές που μπορεί να πάρει μια μεταβλητή λέγονται τιμές της μεταβλητής. Τις μεταβλητές τις διακρίνουμε σε:
 - i) **Ποιοτικές ή Κατηγορικές (qualitative),** οι οποίες είναι μεταβλητές που δεν επιδέχονται αριθμητική μέτρηση. Τέτοιες είναι για παράδειγμα η ομάδα αίματος, το φύλλο, η οικονομική κατάσταση των ανθρώπων
 - ii) **Ποσοτικές (quantitative),** οι οποίες επιδέχονται αριθμητική μέτρηση και με τη σειρά τους χωρίζονται σε:

Διακριτές μεταβλητές (discrete variables), που παίρνουν μόνο ακέραιες τιμές, όπως ο αριθμός των μεταβάσεων, των ημερών της εβδομάδας.

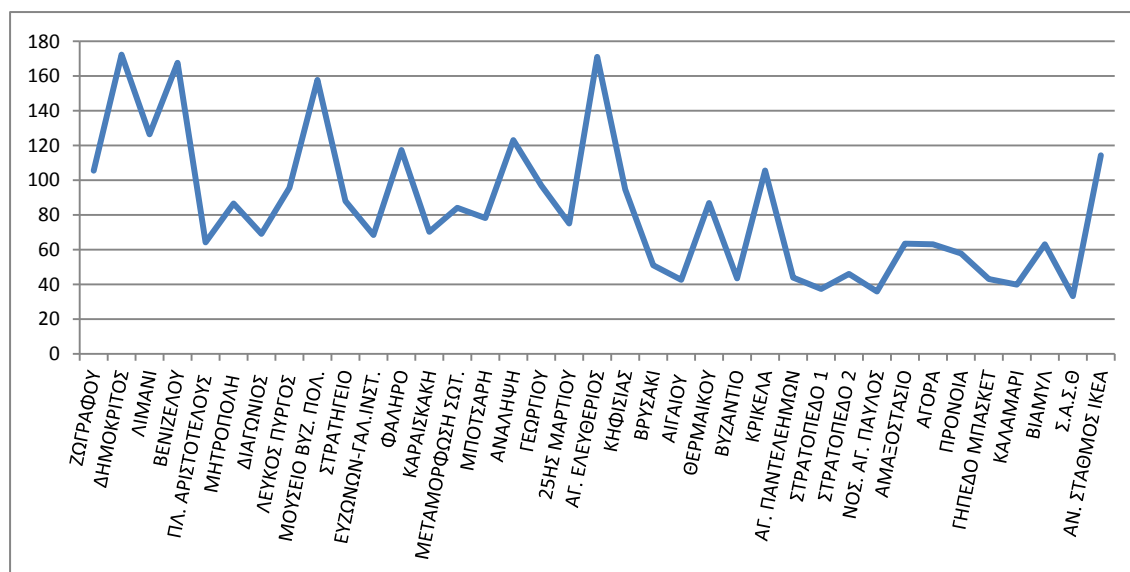
Συνεχείς μεταβλητές (continuous variables), που μπορούν να πάρουν οποιαδήποτε τιμή εντός ενός διαστήματος πραγματικών αριθμών (α, β) , όπως οι διάρκειες μετακίνησης, οι αποστάσεις μεταξύ στάσεων.

3.3.1 Μελέτη συνολικής διαδρομής

Αρχικά παρουσιάζονται τα διαγράμματα των μέσων χρόνων άφιξης στην κάθε στάση για το Δρομολόγιο 3, Ανατολικός Σταθμός ΙΚΕΑ – Τ.Σ Σταθμός Γ, κατά τη μετάβαση και την επιστροφή για να δοθεί μια γενική εικόνα της τάσης που ακολουθούν οι χρόνοι κατά τη διάρκεια της διαδρομής. Στο πρώτο διάγραμμα (Διάγραμμα 3.3.1) φαίνεται ότι οι χρόνοι ακολουθούν μια ανοδική πορεία κατά τη διάρκεια της διαδρομής. Οι διακυμάνσεις είναι σχετικά μικρές καθώς το μεγαλύτερο μέρος των τιμών κυμαίνεται στα 50s έως 90s, με εξαίρεση την αρχή του δρομολογίου όπου έχουμε χαμηλότερες τιμές και το τέλος όπου εμφανίζονται υψηλότερες, ιδιαίτερα στις δυο τελευταίες στάσεις. Οι αρκετά μεγαλύτερες τιμές που εμφανίζονται στο τέλος της διαδρομής οφείλονται στις αυξημένες αποστάσεις μεταξύ των συγκεκριμένων στάσεων σε συνδυασμό με τους υψηλούς φόρτους του κέντρου από το οποίο το λεωφορείο έχει μόλις εξέλθει. Αντιθέτως, οι μέσοι χρόνοι κατά την επιστροφή (Διάγραμμα 3.3.2) ακολουθούν καθοδική πορεία με πολύ έντονες διακυμάνσεις. Οι χρόνοι κυμαίνονται σε αντίστοιχα επίπεδα με προηγουμένως με τιμές από 30s έως 170s, όμως εδώ οι εναλλαγές είναι εντονότερες με δυο κορυφές στις στάσεις Μουσείο Βυζαντινού Πολιτισμού και Άγιος Ελευθέριος.

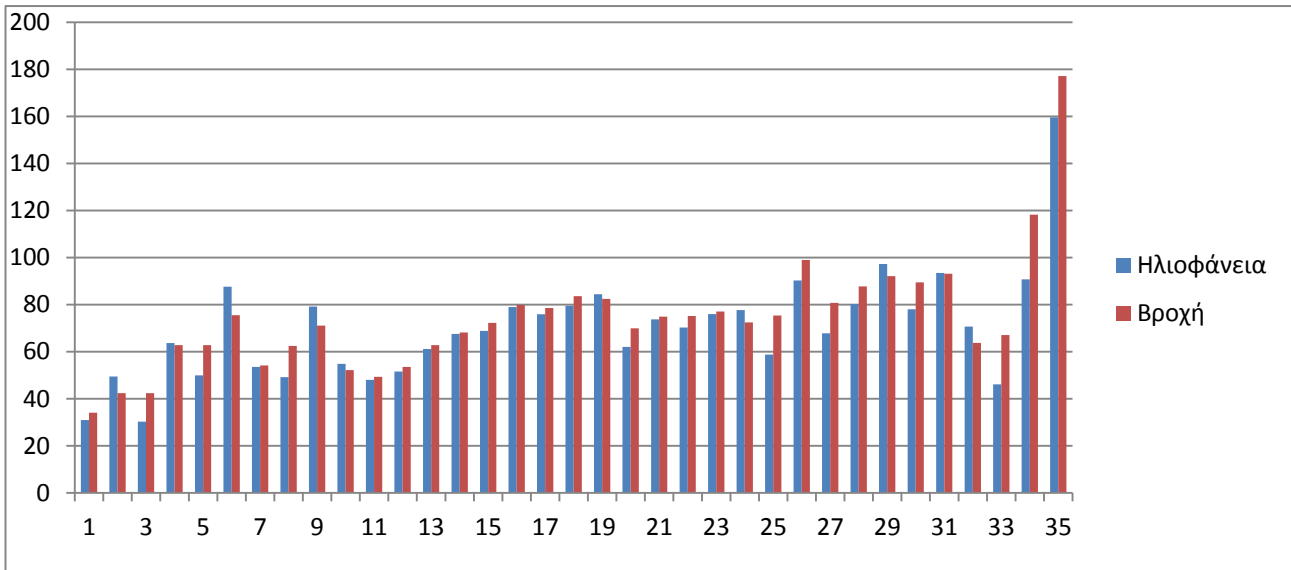


Διάγραμμα 3.3.1 Μέσοι χρόνοι κατά τη μετάβαση

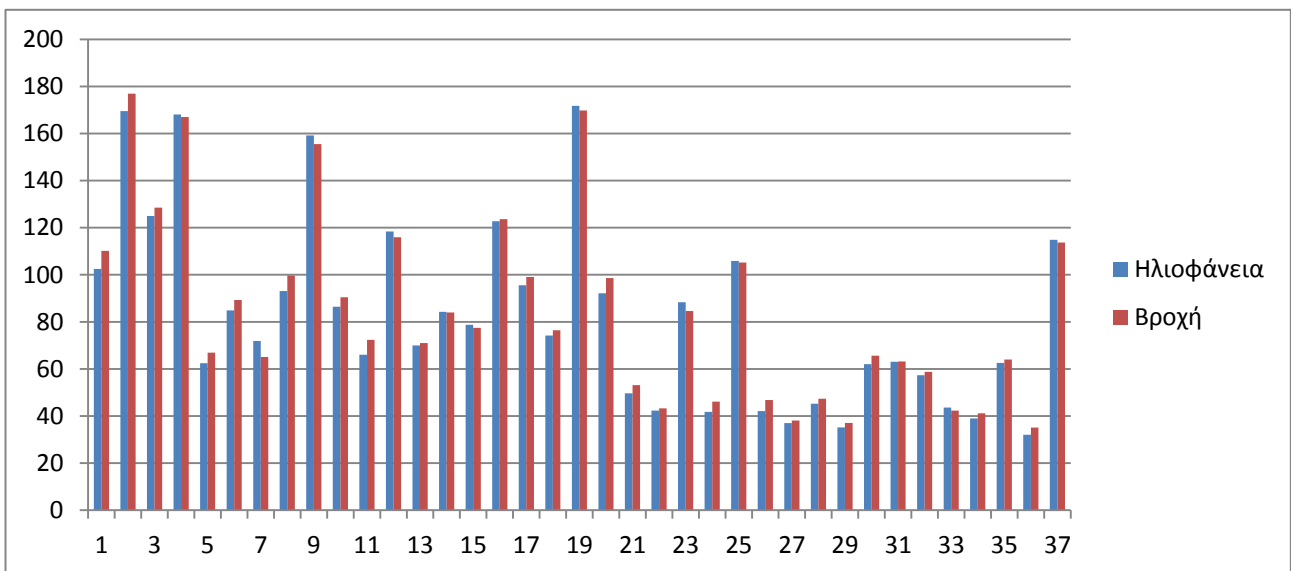


Διάγραμμα 3.3.2 Μέσοι χρόνοι κατά την επιστροφή

Στη συνέχεια παρουσιάζονται γραφικά οι χρόνοι άφιξης ανά στάση κατά τη μετάβαση και επιστροφή (Διάγραμμα 3.3.3 & 3.3.4) σε σχέση με τον καιρό, «Ηλιοφάνεια – Βροχή». Οι ημέρες των μετρήσεων με Βροχή ήταν οι 24 Ιανουαρίου, 7 και 13 Φεβρουαρίου, ενώ με Ηλιοφάνεια οι 28, 29, 30 και 31 Ιανουαρίου. Οι μέσοι χρόνοι κατά τη μετάβαση, όπως και προηγουμένως διατηρούνται σε σταθερά γενικά επίπεδα, χωρίς ιδιαίτερες διακυμάνσεις, με τις μέγιστες τιμές να εμφανίζονται ξανά στις δυο τελευταίες στάσεις. Όπως ήταν αναμενόμενο οι μέσοι απαιτούμενοι χρόνοι μεταξύ των στάσεων στις ημέρες με βροχή είναι μεγαλύτεροι από τους αντίστοιχους με ηλιοφάνεια. Συγκεκριμένα στο 71% των στάσεων εμφανίζονται μεγαλύτεροι χρόνοι όταν ο καιρός είναι βροχερός. Κατά την επιστροφή οι διακυμάνσεις του χρόνου είναι έντονες, με τις τιμές να φτάνουν σε πιο υψηλά επίπεδα, ενώ στο 70% των στάσεων εμφανίζονται μεγαλύτεροι χρόνοι όταν ο καιρός είναι βροχερός.



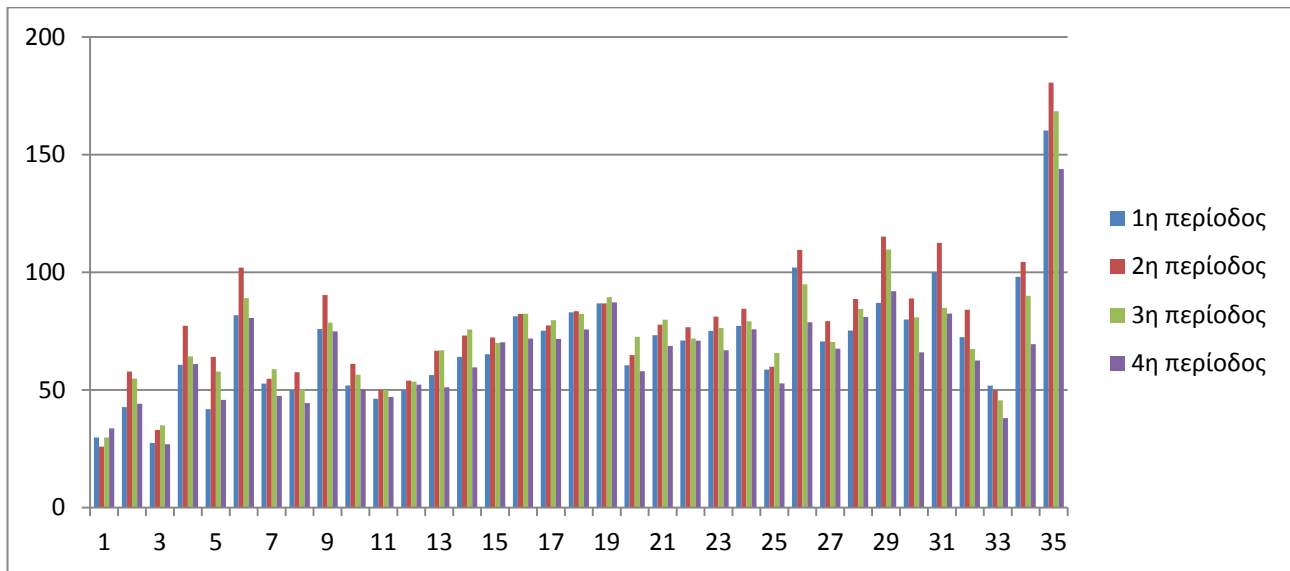
Διάγραμμα 3.3.3 Μέσοι χρόνοι σε σχέση με τον καιρό κατά τη μετάβαση



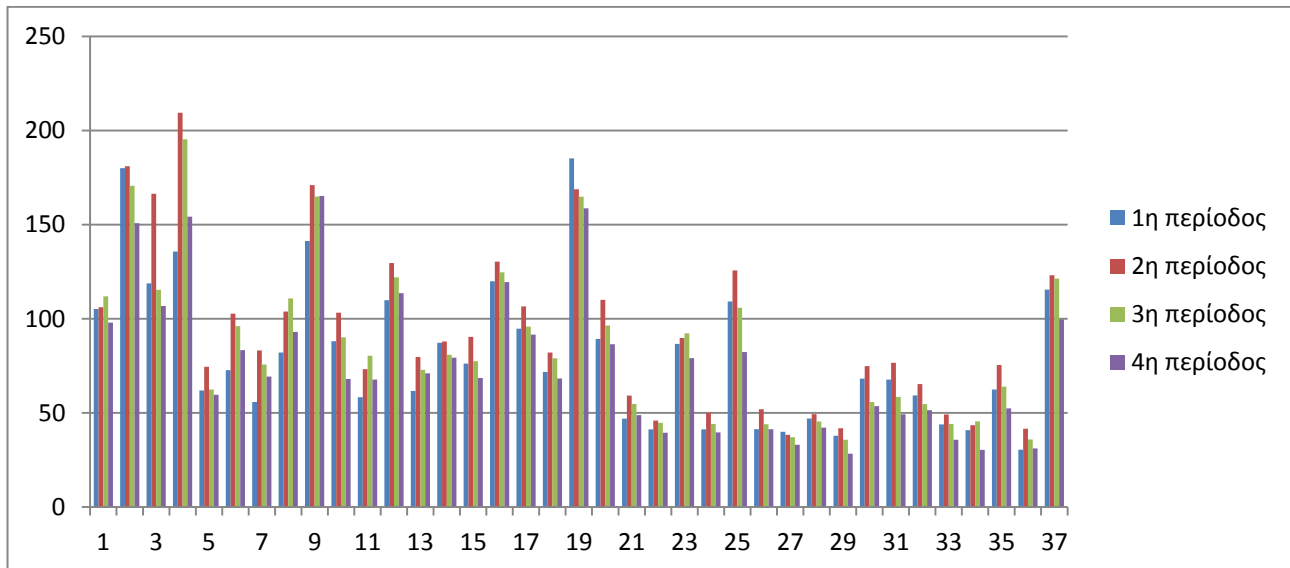
Διάγραμμα 3.3.4 Μέσοι χρόνοι σε σχέση με τον καιρό κατά την επιστροφή

Μια αντίστοιχη προσέγγιση με την παραπάνω έχει πραγματοποιηθεί και για τις περιόδους της ημέρας, με στόχο τον προσδιορισμό της μεταβολής των χρόνων άφιξης σε σχέση με την περίοδο και κατ' επέκταση τις ώρες αιχμής και τους φόρτους. Όπως προαναφέρθηκε η ημέρα έχει χωριστεί στις τέσσερις εξής περιόδους: 1^η 5:00 – 11:00/ 2^η 12:00 – 15:00/ 3^η 16:00 – 19:00/ 4^η 20:00 – 02:00. Από τα παρακάτω διαγράμματα είναι φανερό ότι και στις δυο περιπτώσεις κατά τη 2^η περίοδο της ημέρας οι χρόνοι εμφανίζουν μεγάλη αύξηση. Συγκεκριμένα κατά τη μετάβαση στο 66% των στάσεων η μέγιστη τιμή εμφανίζεται μεταξύ 12:00 – 15:00, ενώ κατά την επιστροφή το ποσοστό ανέρχεται στο 81%. Οι αυξημένες αυτές τιμές δικαιολογούνται, καθώς η 2^η περίοδος συμπίπτει με τη μεσημεριανή αιχμή.

Αντίθετα, η περίοδος στη οποία εμφανίζονται συνήθως οι μικρότεροι χρόνοι διαδρομής είναι η 4^η περίοδος, που κατά τη μετάβαση στο 77% των στάσεων η ελάχιστη τιμή βρίσκεται μεταξύ 20:00 – 02:00 και αντίστοιχα το ποσοστό κατά την επιστροφή είναι 73%, κάτι το οποίο επίσης δικαιολογείται καθώς οι φόρτοι κατά τις μεταμεσονύκτιες ώρες, που συνυπολογίζονται σε αυτή την περίοδο, μειώνονται δραματικά.



Διάγραμμα 3.3.5 Μέσοι χρόνοι ανά περίοδο κατά τη μετάβαση



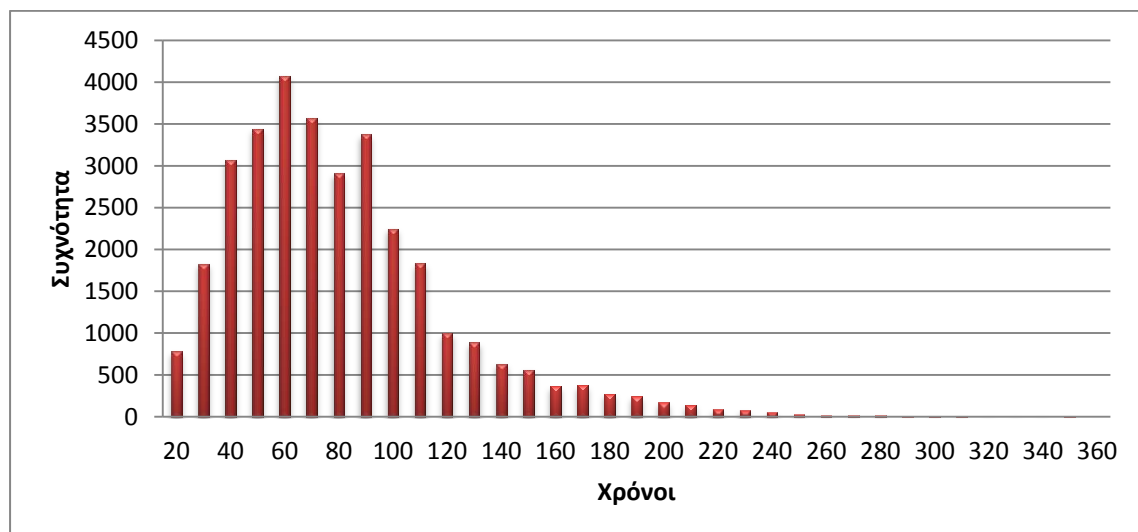
Διάγραμμα 3.3.6 Μέσοι χρόνοι ανά περίοδο κατά την επιστροφή

Στη συνέχεια, κατασκευάστηκε το διάγραμμα συχνοτήτων με σκοπό την καλύτερη κατανόηση του δείγματος. Μέσω του διαγράμματος συχνοτήτων μπορούμε να πάρουμε στοιχεία για το πως είναι κατανομημένοι οι χρόνοι, για την κύρτωση και την ασυμμετρία του συνολικού δείγματος. Όπως φαίνεται οι χρόνοι είναι συσσωρευμένοι στις χαμηλότερες τιμές. Στην συγκεκριμένη περίπτωση έχουμε θετική ασυμμετρία, καθώς οι περισσότερες παρατηρήσεις, η διάμεσος και η μέση τιμή βρίσκονται δεξιά της κορυφής και η τιμή της είναι $\gamma=1,52$. Ο συντελεστής ασυμμετρίας του δείγματος υπολογίστηκε με βάση τις ροπές από τον τύπο:

$$\gamma = \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n}}{\left\{ \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \right\}^3}$$

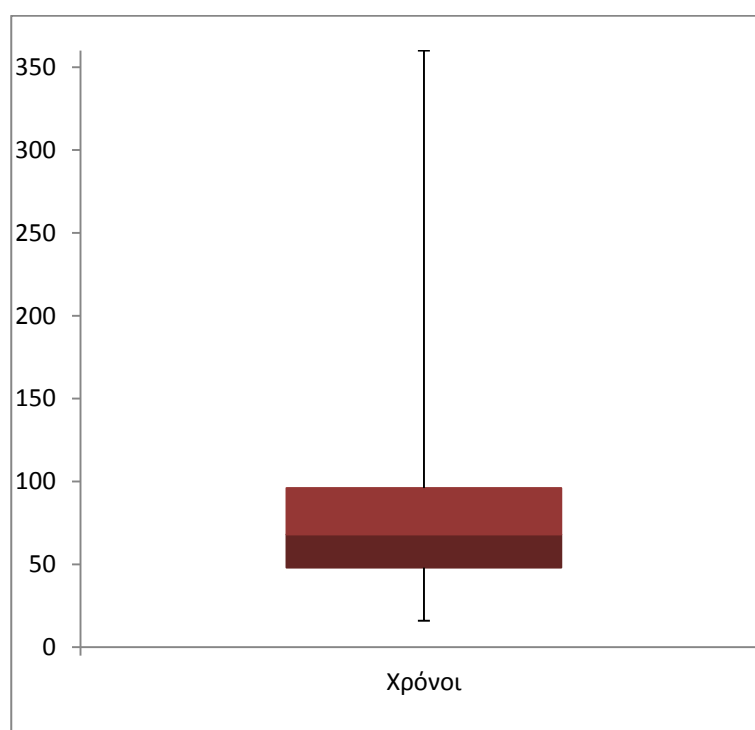
Όσον αφορά την κύρτωση, στο συγκεκριμένο δείγμα υπάρχει σχετικά μεγάλη μέγιστη συχνότητα (κορυφή) και επομένως μεγάλη συγκέντρωση τιμών γύρω από τον μέσο άρα έχουμε λεπτόκυρτη κατανομή (Leptokurtic distribution). Ένα μέτρο που εκφράζει το βαθμό κυρτότητας μιας κατανομής είναι ο συντελεστής κύρτωσης του Pearson α , ο οποίος συγκρίνεται με την τιμή $\alpha=3$ που έχουμε στην κανονική κατανομή. Στο δικό μας δείγμα προέκυψε $\alpha=3,75$ και υπολογίστηκε από τον εξής τύπο:

$$\alpha = \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n}}{\left\{ \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \right\}^4}$$



Διάγραμμα 3.3.7 Διάγραμμα Συχνοτήτων ολικού δείγματος χρόνων

Στο διάγραμμα 3.3.8 απεικονίζεται το θηκόγραμμα (box plot) της μεταβλητής του χρόνου και παρουσιάζονται έτσι κάποια από τα βασικά χαρακτηριστικά της κατανομής. Για την κατασκευή του υπολογίστηκαν οι τιμές των: Q1, πρώτο ή κατώτατο τεταρτημόριο, που συμβολίζει το 25-εκατοστιαίο σημείο, Q3 τρίτο ή ανώτερο τεταρτημόριο, η διάμεσος του δείγματος, η ελάχιστη τιμή $x_{\min}$ και η μέγιστη τιμή $x_{\max}$. Το θηκόγραμμα μας δίνει το κεντρικό διάστημα με το 50% των παρατηρήσεων και η θέση της διαμέσου και των ακραίων τιμών μας δίνουν μια εικόνα για τη συμμετρικότητα της κατανομής και της μεταβλητής μας. Συγκεκριμένα, οι τιμές που υπολογίστηκαν ήταν: $\min=16$, $\max=360$, $Q1=48$, $Q3=96$, $\text{median}=68$, $\text{whisker top}=264$, $\text{whisker bottom}=32$.



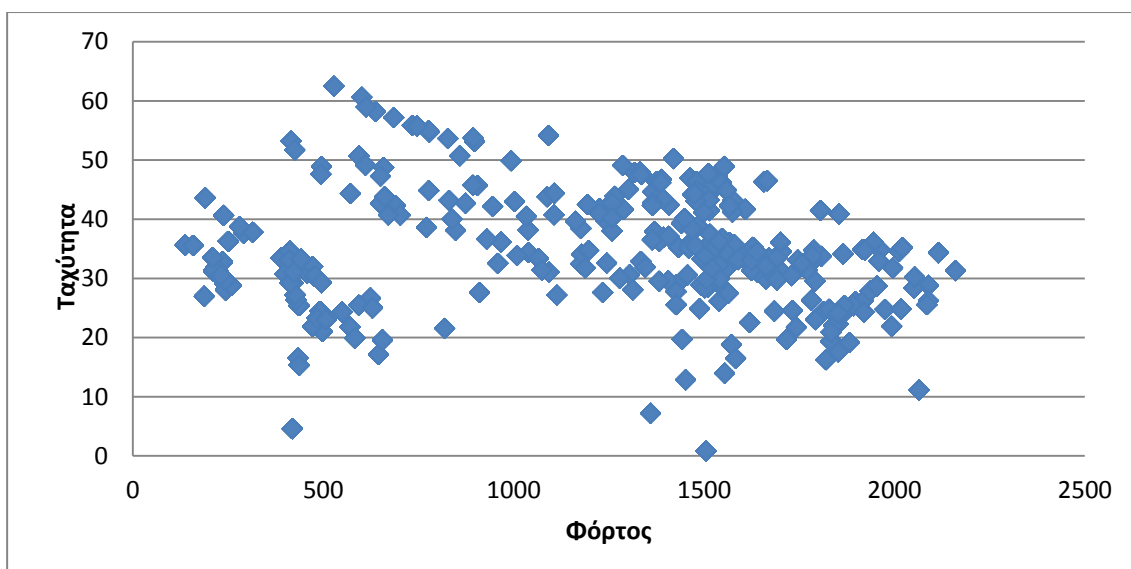
Διάγραμμα 3.3.8 Θηκόγραμμα χρόνων

3.3.2 Μελέτη διαδρομής εντός του κέντρου

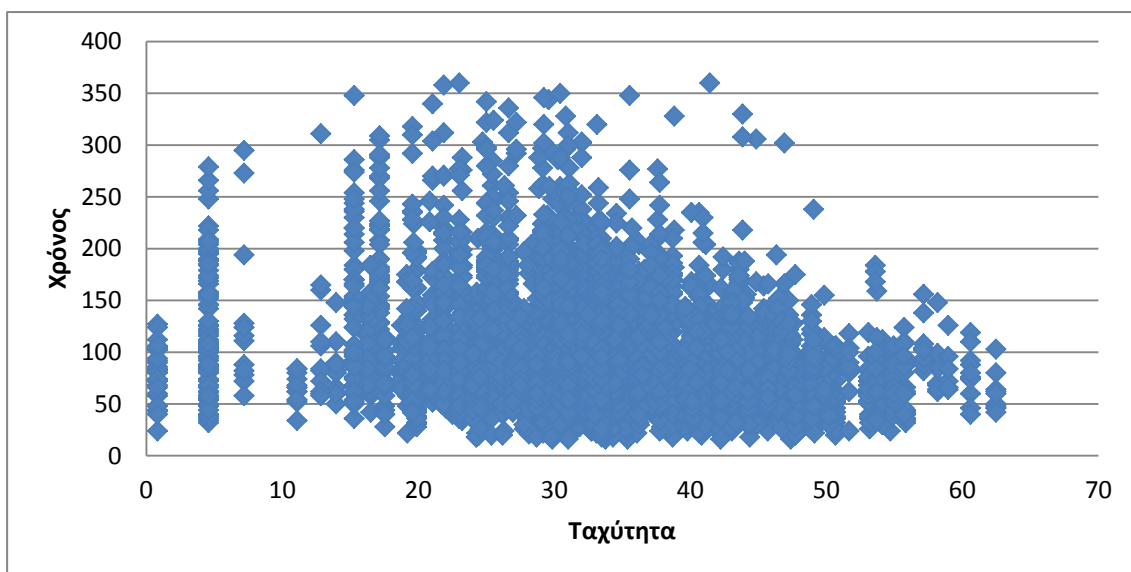
Στη συνέχεια μελετήθηκαν τα χαρακτηριστικά των μεταβλητών για τις στάσεις εντός του κέντρου. Στο κεντρικό κομμάτι όπως έχει προαναφερθεί υπάρχουν επιπλέον τα κυκλοφοριακά δεδομένα των φόρτων (οχήμ./ώρα) και των ταχυτήτων (χλμ./ώρα) στα οποία έχει δοθεί ιδιαίτερο βάρος, με σκοπό να διεξαχθούν συμπεράσματα για το πως μεταβάλλονται οι χρόνοι σε σχέση με αυτά.

Αρχικά κατασκευάστηκαν τα παρακάτω διαγράμματα διασποράς μεταξύ των Φόρτων – Ταχύτητας (Διάγραμμα 3.3.9) και Ταχύτητας – Χρόνων (Διάγραμμα 3.3.10). Από το πρώτο διάγραμμα 3.3.9 μπορούμε να διεξάγουμε κάποια γενικά συμπεράσματα για τη σχέση μεταξύ των φόρτων και της ταχύτητας, όπως για παράδειγμα ότι σε γενικές γραμμές η μεταξύ τους σχέση είναι η αναμενόμενη. Βέβαια οι μετρήσεις αυτές απέχουν πολύ από το να έχουν γραμμική σχέση. Οι τιμές των ταχυτήτων για χαμηλούς φόρτους (<700 οχήμ./ώρα) διατηρούνται σε σταθερά επίπεδα (20–40 χλμ./ώρα), ενώ για μεγαλύτερους φόρτους το εύρος τιμών των ταχυτήτων είναι αρκετά μεγαλύτερο. Επίσης συγκεντρωμένες φαίνεται να είναι οι τιμές του διαγράμματος στους υψηλούς φόρτους 1500–2000 (οχήμ./ώρα) σε αντίθεση με το αρχικό κομμάτι που οι φόρτοι έχουν χαμηλότερες τιμές.

Από το δεύτερο διάγραμμα 3.3.10 φαίνεται η σχέση μεταξύ των χρόνων και της ταχύτητας για τις στάσεις εντός του κέντρου. Αρχικά να σημειωθεί ότι οι χρόνοι είναι πολύ συσσωρευμένοι μεταξύ 30–200 δευτερολέπτων με λίγες τιμές του δείγματος να ξεφεύγουν από αυτό το εύρος. Επιπλέον οι χρόνοι μεγιστοποιούνται για τιμές της ταχύτητας 20–35 χλμ./ώρα, ενώ για μεγαλύτερες τιμές της οι χρόνοι μειώνονται ξανά, όπως ήταν αναμενόμενο, στα 30–100 δευτερόλεπτα. Σε γενικές γραμμές όμως παρατηρεί κανείς ότι οι μεταβολές των χρόνων δεν είναι μεγάλες. Αυτό οφείλεται στο γεγονός ότι στο κέντρο της Θεσσαλονίκης αφενός οι ταχύτητες, ειδικά κατά τη διάρκεια της ημέρας, δεν μπορούν να αυξηθούν ιδιαιτέρως και αφετέρου οι αποστάσεις μεταξύ των στάσεων είναι μικρές, με αποτέλεσμα οι χρόνοι να μην επηρεάζονται σε μεγάλο βαθμό από τη ταχύτητα.



Διάγραμμα 3.3.9 Διάγραμμα Φόρτων-Ταχυτήτων στο κέντρο της Θεσσαλονίκης



Διάγραμμα 3.3.10 Διάγραμμα Ταχυτήτων-Χρόνων στο κέντρο της Θεσσαλονίκης

Παρακάτω απεικονίζονται τα διαγράμματα των φόρτων και των ταχυτήτων για τις στάσεις του κέντρου, κατά τη μετάβαση και επιστροφή, για να γίνει αντιληπτή η μεταβολή τους σε σχέση με τις καιρικές συνθήκες και την περίοδο της ημέρας.

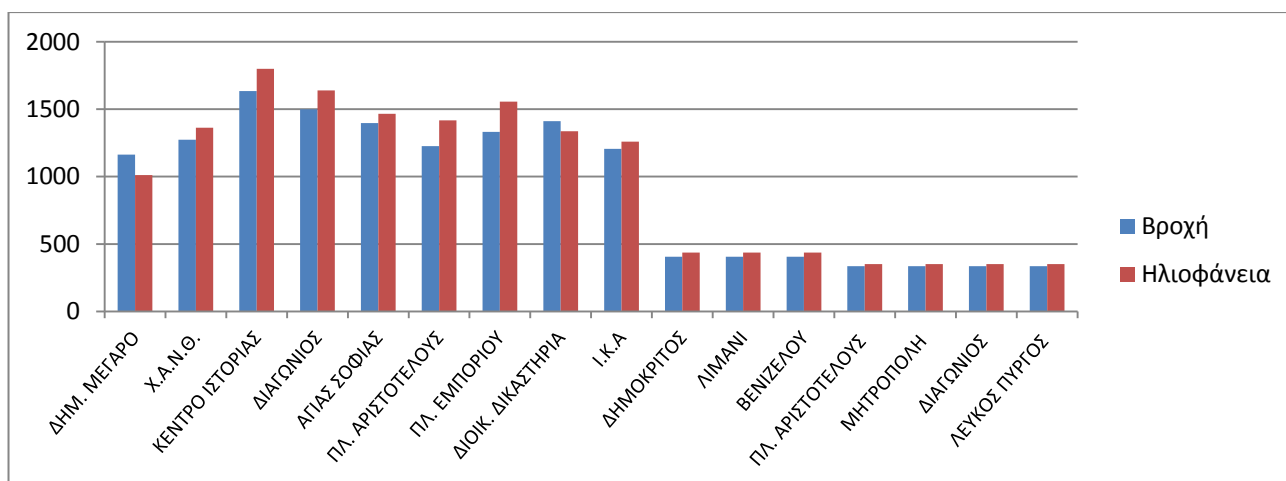
Στο πρώτο διάγραμμα 3.3.11 φαίνονται οι φόρτοι στην κάθε στάση ανάλογα με τις καιρικές συνθήκες. Στην πλειοψηφία των στάσεων οι φόρτοι κατά τη διάρκεια των ημερών με ηλιοφάνεια είναι εμφανώς αυξημένοι, κάτι που υποδηλώνει ότι ο καιρός επηρεάζει τις κυκλοφοριακές συνθήκες και τη συμφόρηση στο κέντρο της Θεσσαλονίκης. Κατά τις ημέρες με βροχή οι ταχύτητες των οχημάτων μειώνονται με αποτέλεσμα οι χρόνοι μεταξύ των στάσεων να αυξάνονται και τα οχήματα ανά ώρα που διέρχονται από τους φωρατές κυκλοφορίας να μειώνονται. Επίσης αξίζει να σημειωθεί ότι οι φόρτοι κατά τη μετάβαση του δρομολογίου (Δημαρχιακό Μέγαρο έως Ι.Κ.Α) που το λεωφορείο διασχίζει τη Τσιμισκή, είναι ιδιαίτερα αυξημένοι σε σχέση με εκείνους της επιστροφής μέσω της Δωδεκανήσου και της Μητροπόλεως.

Κάτι αντίστοιχο φαίνεται να συμβαίνει και με τις ταχύτητες στο διάγραμμα 3.3.12 όπου στο σύνολο σχεδόν των στάσεων οι ταχύτητες στις ημέρες με ηλιοφάνεια είναι μεγαλύτερες από εκείνες με βροχή. Κατ' αντιστοιχία με προηγουμένως τις ημέρες με βροχή οι χρόνοι μεταξύ των στάσεων μειώνονται και οι μέσες ταχύτητες πέφτουν. Στην περίπτωση των ταχυτήτων οι μέσες τιμές ανά στάση δεν διαφέρουν σε μεγάλο βαθμό κατά τη μετάβαση και την επιστροφή. Οι τιμές των ταχυτήτων στη μετάβαση παρουσιάζουν μια διακύμανση, ενώ κατά την επιστροφή διατηρούνται σε σταθερά γενικά επίπεδα.

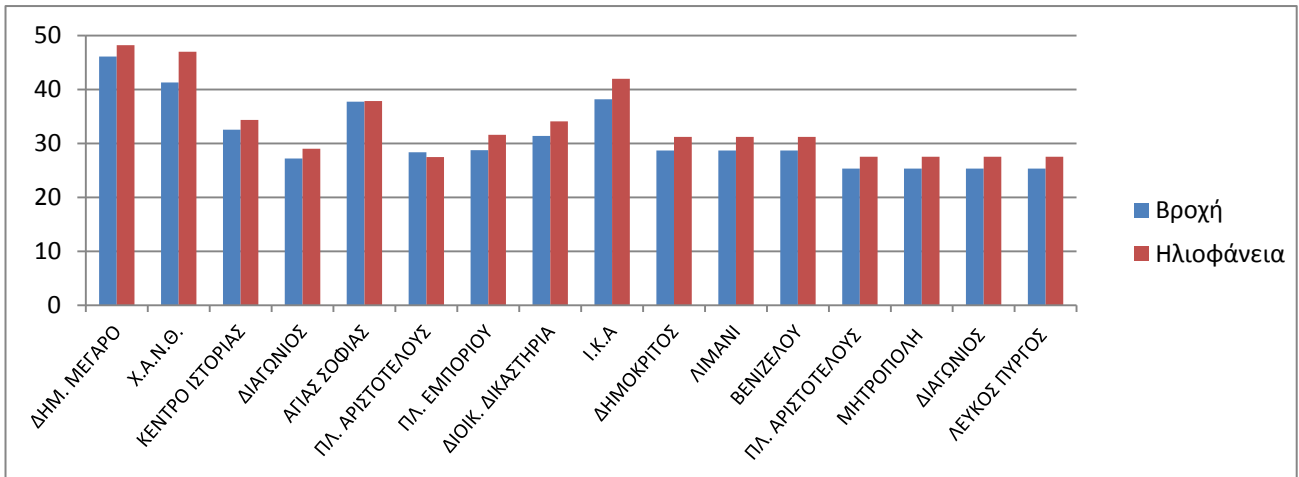
Στα επόμενα δυο διαγράμματα απεικονίζονται οι φόρτοι και οι ταχύτητες στο κέντρο της πόλης για τις τέσσερις περιόδους της ημέρας. Στο πρώτο διάγραμμα

3.3.13 φαίνεται ότι οι φόρτοι κατά την τέταρτη περίοδο είναι για το σύνολο των στάσεων οι χαμηλότεροι που εμφανίζονται. Όπως ήταν αναμενόμενο βάσει του προηγούμενου διαγράμματος των φόρτων, οι τιμές τους κατά την επιστροφή είναι σημαντικά χαμηλότερες αυτών της μετάβασης. Επιπρόσθετα, στο 50% των στάσεων οι μέγιστοι φόρτοι συναντώνται κατά τη δεύτερη περίοδο, κάτι που υποδηλώνει ότι και για του κομμάτι του κέντρου η δεύτερη περίοδος αποτελεί την αιχμή της κυκλοφοριακής συμφόρησης.

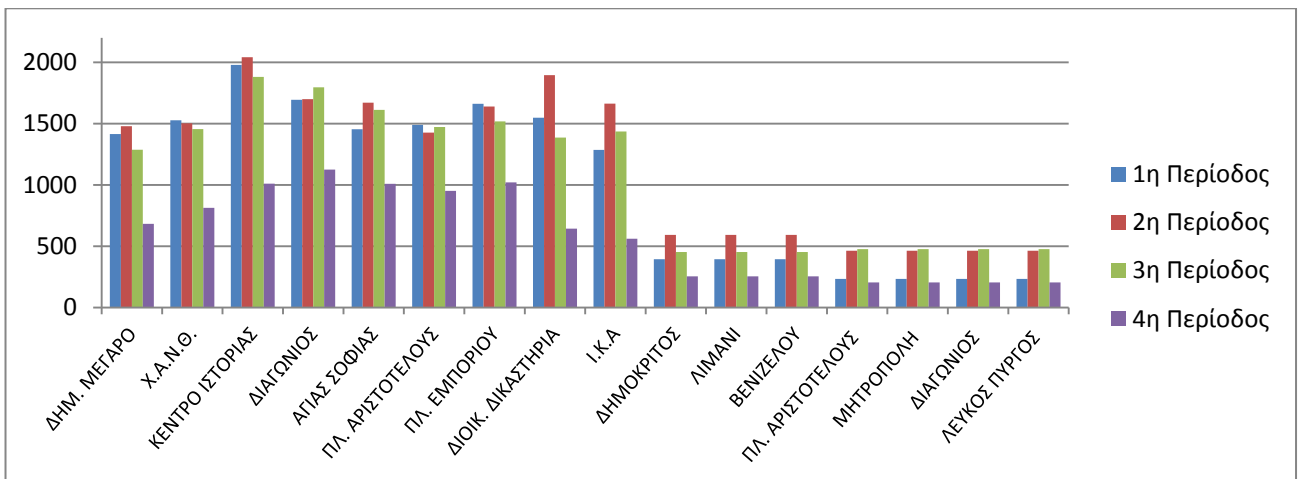
Όπως προκύπτει από το διάγραμμα των ταχυτήτων ανά περίοδο στο 94% των περιπτώσεων η μέγιστη ταχύτητα εμφανίζεται κατά την τέταρτη περίοδο, ενώ αντίθετα οι ελάχιστες ταχύτητες παρατηρούνται κατά τη δεύτερη περίοδο. Αυτό το αποτέλεσμα ήταν αναμενόμενο, καθώς κατά τη δεύτερη περίοδο μεγιστοποιούνται οι φόρτοι και κατά συνέπεια οι τιμές των χρόνων και των ταχυτήτων «βυθίζονται». Γενικότερα, παρατηρείται μια έντονη μεταβλητότητα των ταχυτήτων κατά τη μετάβαση της διαδρομής, σε αντίθεση με τις τιμές που εμφανίζονται κατά την επιστροφή όπου οι ταχύτητες διατηρούνται στα 20-30 χλμ./ώρα.



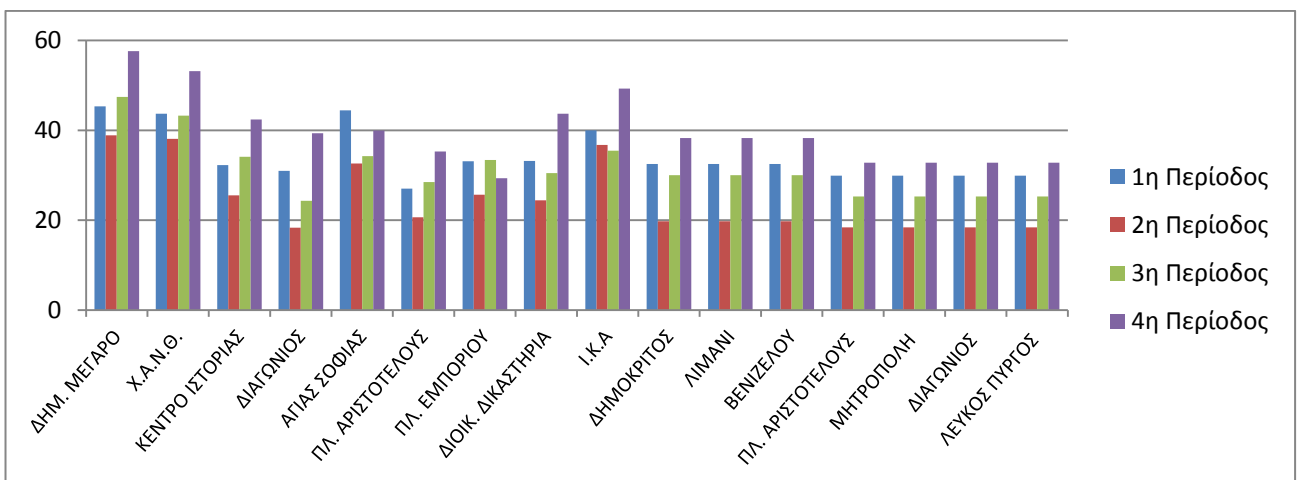
Διάγραμμα 3.3.11 Φόρτοι σε σχέση με τον καιρό στο κέντρο της Θεσσαλονίκης



Διάγραμμα 3.3.12 Ταχύτητες σε σχέση με τον καιρό στο κέντρο της Θεσσαλονίκης



Διάγραμμα 3.3.13 Φόρτοι ανά περίοδο στο κέντρο της Θεσσαλονίκης



Διάγραμμα 3.3.14 Ταχύτητες ανά περίοδο στο κέντρο της Θεσσαλονίκης

3.4 Μελέτη Κατανομής βάσης δεδομένων

Μετά την στατιστική επεξεργασία των δεδομένων και πριν την δημιουργία των μοντέλων για την τελική επεξεργασία κρίθηκε απαραίτητο να καθορισθεί η κατανομή που ακολουθεί η μεταβλητή του χρόνου που μελετάται. Αυτό πραγματοποιήθηκε με τη βοήθεια του προγράμματος BestFit, το οποίο είναι ένα εργαλείο-προσθήκη στο πρόγραμμα Excel. Το BestFit ελέγχει την κατανομή του δείγματος, συγκρίνοντάς την με 28 διαφορετικές κατανομές, και επιλέγει εκείνη που προσομοιάζει το δείγμα με το μικρότερο δυνατό σφάλμα. Μερικές από τις πιο σημαντικές κατανομές που ελέγχει το πρόγραμμα είναι οι εξής: Βήτα, Διωνυμική, χ^2 , Εκθετική, Γάμμα, Γεωμετρική, Λογιστική, Λογαριθμολογιστική, Κανονική, Λογαριθμοκανονική, Pareto, Pearson, Poisson, Student's t, Weibull. Πάντοτε τα αποτελέσματα του προγράμματος πρέπει να ελέγχονται ποιοτικά και ποσοτικά από τον μελετητή πριν χρησιμοποιηθούν.

Για τη διεξαγωγή του καλύτερου δυνατού αποτελέσματος υπολογίζεται η πιθανότητα η κατανομή των δεδομένων εισαγωγής να έχει παραχθεί από αυτήν που τελικά επιλέγεται από το πρόγραμμα. Σε πρώτο στάδιο υπολογίζονται οι παράμετροι με τη μέθοδο της μέγιστης πιθανοφάνειας. Αυτό γίνεται μεγιστοποιώντας τη συνάρτηση πιθανοφάνειας $L(\beta)$:

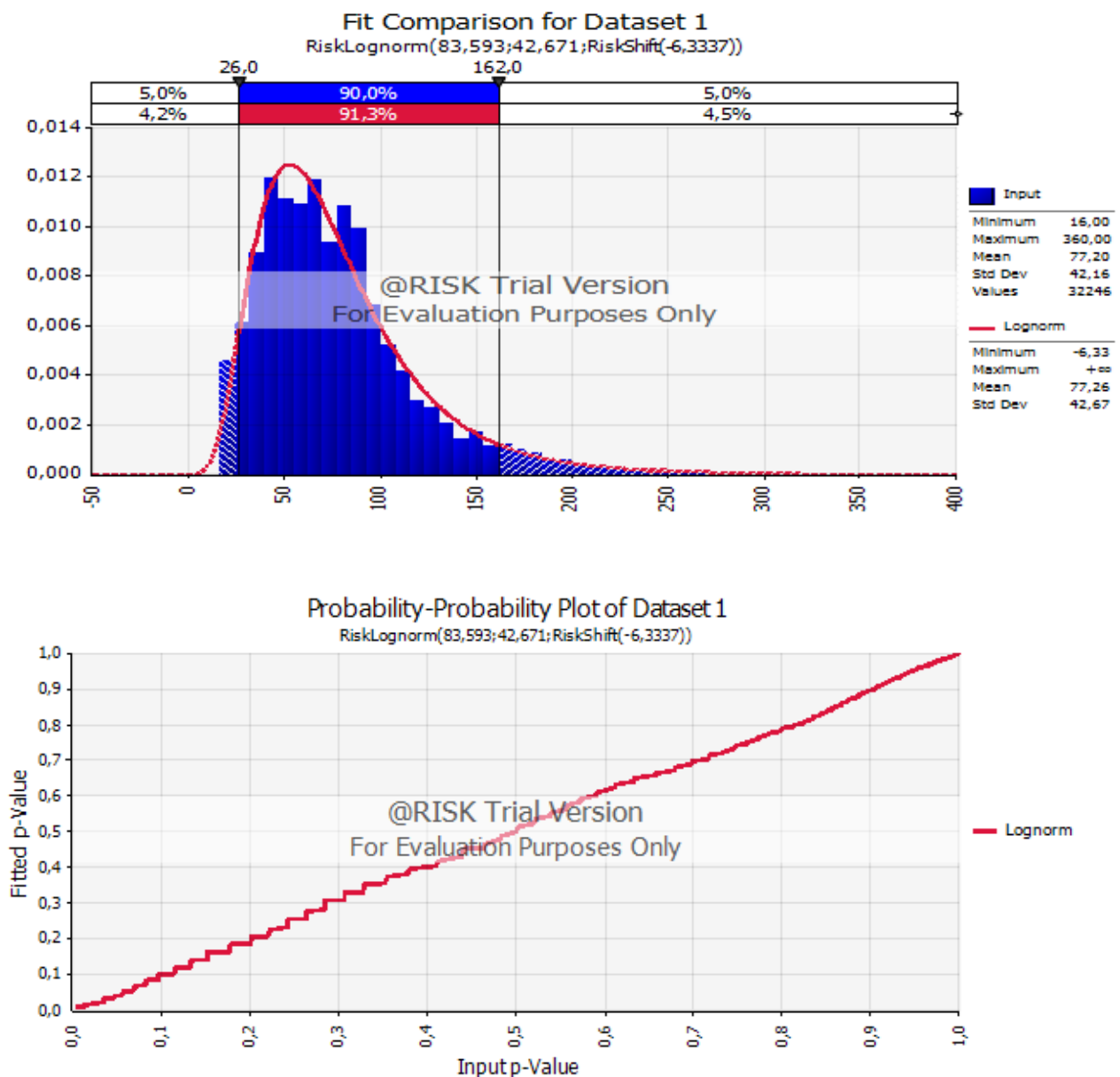
$$L(\beta) = f(x_1, x_2, \dots, x_n, \beta) = \prod_{i=1}^n f(x_i, \beta)$$

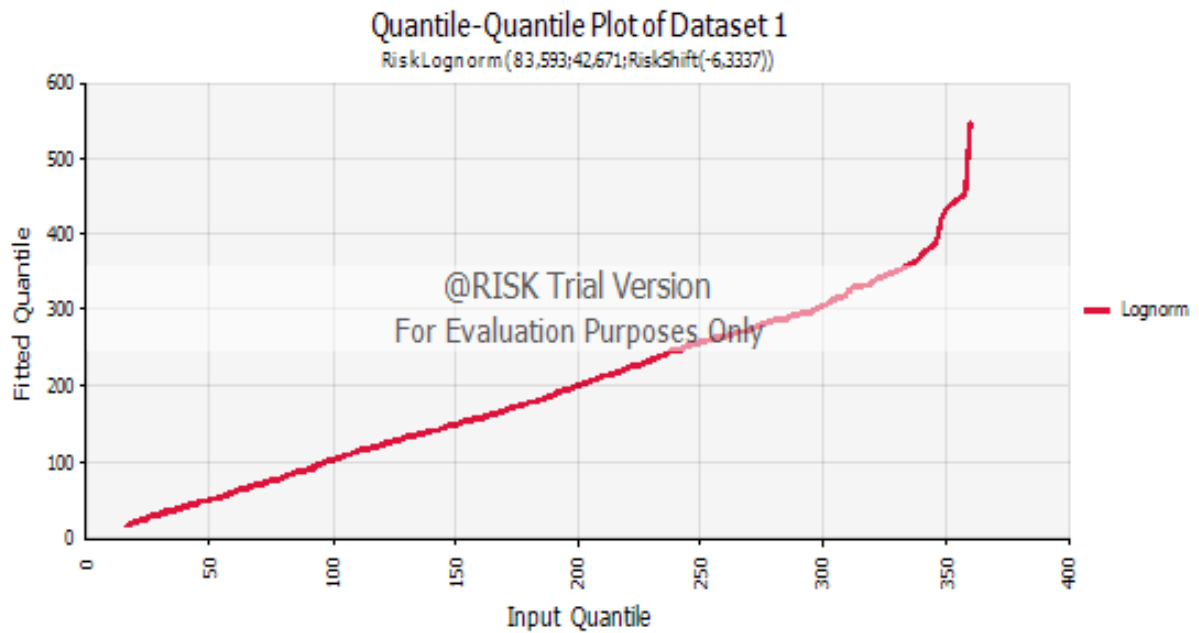
Ενώ η μέθοδος των ελαχίστων τετραγώνων χρησιμοποιείται για την εύρεση των παραμέτρων που ελαχιστοποιούν τις διαφορές μεταξύ των τιμών της καμπύλης των εισερχόμενων μεταβλητών και της θεωρητικής συνάρτησης. Τέλος οι κατανομές που είναι πιο κοντά στην πραγματική κατατάσσονται βάσει ενός ή περισσότερων στατιστικών δεικτών όπως χ^2 , Anderson-Darling και Komolgorov-Smirnov.

Αρχικά εισήχθησαν οι τιμές των χρόνων για το συνολικό μοντέλο. Τα αποτελέσματα του BestFit έδειξαν ότι η κατανομή που προσομοιάζει καλύτερα αυτές τις τιμές είναι η Λογαριθμοκανονική. Στο παρακάτω διάγραμμα 3.4.1 απεικονίζεται το διάγραμμα συχνοτήτων των χρόνων και η γραμμή της Λογαριθμοκανονικής κατανομής. Επιπλέον αναγράφονται τα στατιστικά χαρακτηριστικά των δεδομένων εισόδου (ελάχιστο, μέγιστο, μέσος όρος, τυπική απόκλιση, σύνολο τιμών), αλλά και της Λογαριθμοκανονικής κατανομής. Τέλος, μας δίνει πληροφορίες για το κατά πόσο είναι συσσωρευμένες οι τιμές των χρόνων, όπου για τα δεδομένα εισόδου το 90% των τιμών βρίσκεται μεταξύ 26-162 δευτερόλεπτα, ενώ το αναμενόμενο ποσοστό για την Λογαριθμοκανονική κατανομή είναι 91,3%. Τα επόμενα δυο διαγράμματα είναι αντίστοιχα τα Q-Q (quantile-quantile) plot P-P (probability-probability) plot. Το P-P plot είναι ένα διάγραμμα

πιθανότητας το οποίο υποδεικνύει πόσο στενά συνδεδεμένα είναι τα στοιχεία της πραγματικής κατανομής και εκείνης με την οποία έχει προσομοιαστεί. Όσο πλησιάζει η κατανομή που έχει επιλεγθεί σε αυτή του δείγματος, τόσο συσσωρεύονται οι τιμές γύρω από μια ευθεία γραμμή. Το Q-Q plot αποτελεί μια γραφική μέθοδο απεικόνισης της συνάρτησης πιθανότητας, στον άξονα των Χ τοποθετούνται οι τιμές της μεταβλητής εισόδου και στον άξονα των Υ οι αντίστοιχες της κατανομής.

Αντίστοιχα είναι και διαγράμματα κατανομής του κέντρου, στο οποίο οι χρόνοι προσομοιάζονται και πάλι με την Λογαριθμοκανονική.





Διάγραμμα 3.4.1 Κατανομή χρόνων συνολικής διαδρομής

Στο σημείο αυτό παρατίθεται ο συνολικός πίνακας των στατιστικών στοιχείων του δείγματος. Έχουν υπολογιστεί οι ελάχιστες και οι μέγιστες τιμές, καθώς επίσης και οι μέσοι όροι και τυπικές αποκλίσεις για όλες τις περιπτώσεις. Ο πίνακας αυτός έχει κατασκευασθεί για να δώσει μια γενική εικόνα στον αναγνώστη, χωρίς βέβαια να δίνει ιδιαίτερα στοιχεία για τις εναλλαγές των χρόνων κατά τη διάρκεια της ημέρας, της εβδομάδας ή των καιρικών συνθηκών. Σε γενικές γραμμές οι ελάχιστοι (16δευτ.) και οι μέγιστοι (360δευτ.) χρόνοι που παρατηρούνται είναι σταθεροί, με μόνη εξαίρεση τις ημέρες Δευτέρα και Τρίτη, όπου οι μέγιστοι χρόνοι που εμφανίζονται είναι ελαφρώς μειωμένοι. Κάτι αντίστοιχο συμβαίνει και με τις μέσες τιμές των χρόνων οι οποίες κυμαίνονται μεταξύ 70-80 δευτερόλεπτα, με μόνες τιμές που διαφέρουν να είναι της επιστροφής του συνολικού δρομολογίου, της εντός του κέντρου διαδρομής και της δεύτερης περιόδου. Όσον αφορά τώρα τις τιμές της διακύμανσης, παρατηρεί κανείς ότι είναι αρκετά μεγάλες, καθώς κυμαίνονται από 40 έως 48 δευτερόλεπτα. Αυτό επιβεβαιώνει ότι ο πίνακας δεν είναι αντιπροσωπευτικός των τιμών της βάσης δεδομένων και των διακυμάνσεων που αυτή ακολουθεί και για αυτό το λόγο η ανάλυση και ο σχολιασμός του δείγματος έχει ήδη πραγματοποιηθεί μέσω των πινάκων στο υποκεφάλαιο 3.3.

Πίνακας 3.4.1 Συγκεντρωτικά στατιστικά στοιχεία

		MIN	MAX	Μ.Ο.	Τ.Α.
Συνολικά	Συνολικά	16	360	77	42,20
Μετάβαση-Επιστροφή	Μετάβαση	16	360	71	33,76
	Επιστροφή	16	360	83	48,10
Ημέρα	Δευτέρα	16	336	76	40,45
	Τρίτη	16	338	76	40,58
	Τετάρτη	16	350	76	41,17
	Πέμπτη	16	360	79	44,04
Καιρός	Ηλιοφάνεια	16	358	76	41,18
	Βροχή	16	360	79	43,64
Εντός-Εκτός	Εκτός κέντρου	16	360	73	39,01
	Εντός κέντρου	16	360	93	48,86
Περίοδος	Περίοδος 1η	16	356	75	40,79
	Περίοδος 2η	16	358	86	46,13
	Περίοδος 3η	16	360	79	42,45
	Περίοδος 4η	16	348	70	39,72

4. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

4.1 Εισαγωγή

Στο παρόν κεφάλαιο που αφορά στο θεωρητικό υπόβαθρο, παρουσιάζεται το πλαίσιο, στο οποίο θα στηριχθεί η στατιστική ανάλυση των στοιχείων από τη δοθείσα βάση δεδομένων. Η επιλογή της μεθοδολογίας έγινε βάσει του αντικειμένου που πραγματεύεται η μελέτη και το περιεχόμενο της βάσης δεδομένων από την οποία συλλέχθηκαν τα στοιχεία. Η μελέτη πραγματοποιήθηκε με τις μεθόδους της Λογαριθμοκανονικής Παλινδρόμησης (Lognormal Regression), που είναι μια παραλλαγή της Γραμμικής Παλινδρόμησης (Linear Regression) και της Λογιστικής Παλινδρόμησης (Logistic Regression). Παρακάτω παρουσιάζονται οι μαθηματικές και στατιστικές θεωρίες όπου βασίστηκε η επιλογή των παραπάνω θεωριών.

Ως **παλινδρόμηση** ορίζεται κάθε συσχέτιση που στηρίζεται στη μέθοδο των ελάχιστων τετραγώνων. Σε κάθε πρόβλημα παλινδρόμησης διακρίνουμε δύο είδη μεταβλητών: τις **ανεξάρτητες ή ελεγχόμενες ή επεξηγηματικές** (independent, predictor, casual, input, explanatory variables) και τις **εξαρτημένες ή απόκρισης** (dependent, response variables). Ανεξάρτητη μεταβλητή X είναι εκείνη την οποία μπορούμε να ελέγξουμε, δηλαδή να καθορίσουμε τις τιμές της. Εξαρτημένη μεταβλητή Y είναι αυτή στην οποία αντανακλάται το αποτέλεσμα των μεταβολών στις ανεξάρτητες μεταβλητές. Για να πραγματοποιηθεί η παλινδρόμηση, πρέπει το σύνολο των χρονοσειρών που θα χρησιμοποιηθούν, ανεξάρτητες και εξαρτημένες, να έχουν κοινή περίοδο μετρήσεων, δηλαδή σύνολο χρονικών στιγμών για τις οποίες υπάρχουν τιμές σε όλες τις χρονοσειρές. Ο βαθμός **καταλληλότητας** και **αξιοπιστίας** της μεθόδου για τα συγκεκριμένα δεδομένα αποδίδεται από τον **συντελεστή συσχέτισης R^2** , (Hernance. J.F., 1999).

Υπενθυμίζουμε εδώ ότι οι μεταβλητές διακρίνονται σε Ποιοτικές και Ποσοτικές. Οι ποιοτικές με τη σειρά τους σε συνεχείς και διακριτές. Το είδος της Παλινδρόμησης που θα χρησιμοποιηθεί για την ανάλυση των αποτελεσμάτων της παρούσας έρευνας εξαρτάται από το είδος της μεταβλητής του εξαρτημένου μεγέθους.

- Αν η εξαρτημένη μεταβλητή είναι συνεχής χρησιμοποιείται η τεχνική της **Γραμμικής Παλινδρόμησης**.
- Αν η εξαρτημένη μεταβλητή είναι κατηγορική και παίρνει μόνο δύο τιμές χρησιμοποιείται η **Διωνυμική Λογιστική Παλινδρόμηση**, ενώ
- Αν η εξαρτημένη μεταβλητή είναι κατηγορική με περισσότερες από δύο κατηγορίες χρησιμοποιείται η **Πολυωνυμική Λογιστική Παλινδρόμηση**.

4.2 Μέθοδοι Στατιστικής Ανάλυσης

4.2.1 Γραμμική Παλινδρόμηση – Λογαριθμοκανονική Παλινδρόμηση

Ένα από τα μεγαλύτερα πεδία έρευνας του Συγκοινωνιολόγου Πολιτικού Μηχανικού αποτελεί η μελέτη και επεξήγηση δεδομένων, καθώς και η διεξαγωγή προβλέψεων βάσει ήδη υπαρχουσών συνθηκών. Αυτό πραγματοποιείται μέσω της Ανάλυσης Παλινδρόμησης (regression analysis) που εξετάζει τη σχέση μεταξύ δυο ή περισσότερων μεταβλητών με στόχο την πρόβλεψη μιας εξ αυτών. Για τον λόγο αυτό είναι απαραίτητη η δημιουργία μαθηματικών προτύπων, έτσι ώστε να αποφανθούμε για το εάν μια ή περισσότερες ανεξάρτητες μεταβλητές προκάλεσε κάποια μεταβολή στην εξαρτημένη μεταβλητή.

Μια ευρέως χρησιμοποιούμενη μέθοδος για την στατιστική επεξεργασία και ανάλυση των στοιχείων είναι η μέθοδος της Γραμμικής Παλινδρόμησης (Linear Regression method). Στη μέθοδο της Γραμμικής Παλινδρόμησης η εξαρτημένη μεταβλητή είναι συνεχής. Η απλούστερη μορφή της εν λόγω μεθόδου είναι η απλή γραμμική παλινδρόμηση που χρησιμοποιείται για να διερευνηθεί η εξάρτηση μίας μεταβλητής Y (μεταβλητή απόκρισης ή εξαρτημένη) από μία άλλη μεταβλητή X (ανεξάρτητη ή ελεγχόμενη ή επεξηγηματική).

Η εξαρτημένη τυχαία μεταβλητή Y ακολουθεί κάποια κατανομή με αθροιστική συνάρτηση κατανομής $F_Y(y|X=x)$, δεσμευμένη για κάθε τιμή x της μεταβλητής X . Υποθέτουμε ότι η εξάρτηση εκφράζεται από μια γραμμική σχέση $E(Y|X=x) = \alpha + \beta \cdot x$ και η σχέση αυτή λέγεται **γραμμική παλινδρόμηση της Y στη X** (linear regression). Θεωρούμε ότι οι τιμές της Y για κάθε τιμή $X = x$ παλινδρομούν γύρω από το σημείο $y = E(Y|X=x)$ της ευθείας $\hat{y} = \alpha + \beta \cdot x$, δηλαδή ότι οι τιμές της Y για κάθε τιμή της X βρίσκονται πάνω και κάτω από αυτήν την ευθεία. Το πρόβλημα της παλινδρόμησης είναι η εύρεση των παραμέτρων α και β που εκφράζουν καλύτερα τη γραμμική εξάρτηση της Y από τη X . Κάθε ζεύγος τιμών (α, β) καθορίζει μια διαφορετική γραμμική σχέση που εκφράζεται γεωμετρικά από ευθεία γραμμή και οι δύο παράμετροι ορίζονται ως εξής:

- Ο σταθερός όρος α είναι η τιμή του y για $x = 0$ (intercept).
- Ο συντελεστής β του x είναι η κλίση (slope) της ευθείας ή αλλιώς ο συντελεστής παλινδρόμησης (regression coefficient).

Αν θεωρήσουμε τις παρατηρήσεις $(x_1, y_1), \dots, (x_n, y_n)$ και ένα διάγραμμα διασποράς που να τις απεικονίζει σαν σημεία, μπορούμε να σχηματίσουμε άπειρες τέτοιες ευθείες που να προσεγγίζουν την υποτιθέμενη γραμμική εξάρτηση της $E(Y|X=x)$ ως προς X . Σε κάθε τιμή x_i της X αντιστοιχούν διαφορετικές τιμές y_i της Y , σύμφωνα με κάποια κατανομή πιθανότητας $F_Y(y_i|X=x_i)$, δηλαδή μπορούμε να θεωρήσουμε την y_i σαν τυχαία μεταβλητή. Η τυχαία μεταβλητή y_i για κάποια τιμή x_i της X θα δίνεται κάτω από την υπόθεση της γραμμικής παλινδρόμησης ως:

$$Y = \alpha + \beta \times x_i + \varepsilon$$

όπου ε_i είναι κι αυτή μία τυχαία μεταβλητή που ονομάζεται σφάλμα παλινδρόμησης (*regression error*) και ορίζεται ως η διαφορά της y_i από τη δεσμευμένη μέση τιμή $E(Y | X = x_i)$. Για την ανάλυση της γραμμικής παλινδρόμησης κάνουμε τις παρακάτω τρεις υποθέσεις:

- i. Η μεταβλητή X είναι ελεγχόμενη για το πρόβλημα που μελετάμε, δηλαδή γνωρίζουμε τις τιμές της χωρίς καμιά αμφιβολία.
- ii. Η εξάρτηση της Y από τη X είναι **γραμμική**.
- iii. $E(\varepsilon_i) = 0$ και $\text{Var}(\varepsilon_i) = \sigma^2$, για κάθε τιμή x_i της X , δηλαδή το σφάλμα παλινδρόμησης έχει μηδενική μέση τιμή για κάθε τιμή της X και η διασπορά του είναι σταθερή και δεν εξαρτάται από τη X . Η ιδιότητα αυτή λέγεται **ομοσκεδαστικότητα** και αντίθετα έχουμε ετεροσκεδαστικότητα όταν η διασπορά του σφάλματος ε μεταβάλλεται με τη X .

Οι παραπάνω υποθέσεις για γραμμική σχέση και σταθερή διασπορά αποτελούν χαρακτηριστικά πληθυσμών με κανονική κατανομή. Συνήθως, λοιπόν, σε προβλήματα γραμμικής παλινδρόμησης υποθέτουμε ότι η δεσμευμένη κατανομή της Y είναι κανονική:

$$Y | X = x \sim N(\alpha + \beta x, \sigma^2)$$

Τα προβλήματα γραμμικής παλινδρόμησης επιλύονται τόσο με τον προσδιορισμό του σταθερού όρου της παλινδρόμησης α και του συντελεστή της παλινδρόμησης β για να μπορέσουμε να προσδιορίσουμε την ευθεία της παλινδρόμησης, όσο και με την εύρεση της διασποράς για να γνωρίζουμε το βαθμό μεταβλητότητας, δηλαδή την απόκλιση από την ευθεία. Ο υπολογισμός των παραμέτρων α και β γίνεται με τη μέθοδο των Ελαχίστων Τετραγώνων (*method of least squares*). Η μέθοδος βρίσκει την ευθεία παλινδρόμησης με παραμέτρους α και β έτσι ώστε το άθροισμα των τετραγώνων των κατακόρυφων αποστάσεων των σημείων από την ευθεία να είναι το ελάχιστο. Οι εκτιμήσεις των α και β δίνονται από την *ελαχιστοποίηση του αθροίσματος των τετραγώνων των σφαλμάτων*:

$$\min_{\alpha, \beta} \sum (\varepsilon_i^2), \text{ για } i \text{ από } 1 \text{ έως } n \text{ ή } \min_{\alpha, \beta} \sum (y_i - \alpha - \beta \cdot x_i)^2, \text{ για } i \text{ από } 1 \text{ έως } n$$

Με τον καθορισμό της ευθείας ελαχίστων τετραγώνων, για κάθε δοθείσα τιμή x_i εκτιμούμε την τιμή $\hat{y}_i$ που γενικά είναι διαφορετική από την πραγματική τιμή y_i . Η διαφορά $e_i = y_i - \hat{y}_i$ είναι η κατακόρυφη απόσταση της πραγματικής τιμής από την ευθεία ελαχίστων τετραγώνων (εκτιμώμενη τιμή) και λέγεται **σφάλμα ελαχίστων τετραγώνων** ή απλά **υπόλοιπο** (*residual*).

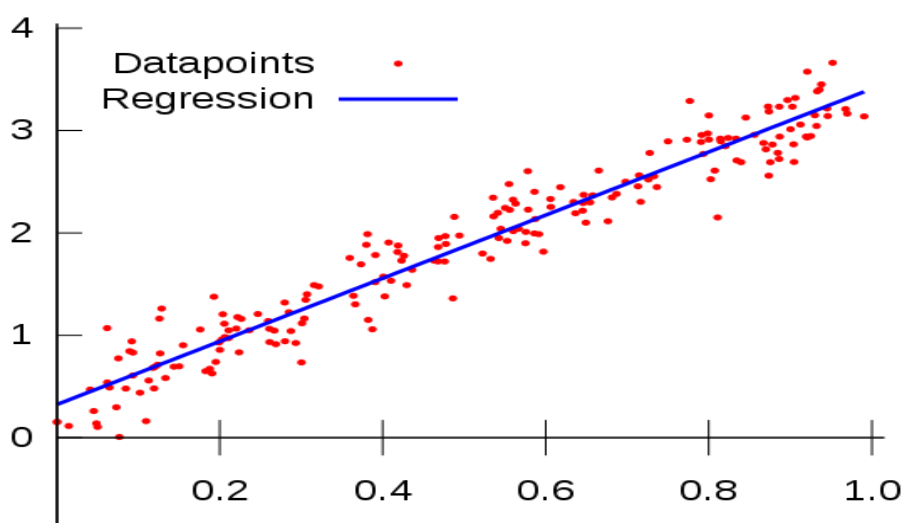
Στα αντικείμενα μελέτης του Συγκοινωνιολόγου Μηχανικού, που αφορούν σύνθετα προβλήματα, η εξαρτημένη μεταβλητή ή μεταβλητή απόκρισης Y επηρεάζεται από περισσότερες από μια ερμηνευτικές μεταβλητές X (ποσοτικές και

ποιοτικές) έστω $X_1, X_2, \dots, X_{p-1}$. Μπορούμε και πάλι να χρησιμοποιήσουμε ένα γραμμικό μοντέλο για να διερευνήσουμε την εξάρτηση της Y από τις $X_1, X_2, \dots, X_{p-1}$. Το μοντέλο αυτό, το οποίο αποτελεί γενίκευση του μοντέλου $Y = \alpha + \beta x + \varepsilon$, έχει τη μορφή:

$$Y = \alpha + \beta_1 \times x_1 + \beta_2 \times x_2 + \dots + \beta_{p-1} \times x_{p-1} + \varepsilon$$

Απόρροια του γεγονότος ότι η εξαρτημένη μεταβλητή είναι συνεχής, είναι ότι ο κάθε συντελεστής β_i δηλώνει την κατεύθυνση της επιρροής της ανεξάρτητης μεταβλητής στην εξαρτημένη. Δηλαδή, αν ο συντελεστής έχει θετικό πρόσημο, η αύξηση της τιμής της ανεξάρτητης μεταβλητής προκαλεί αύξηση στην τιμή της εξαρτημένης μεταβλητής (το αντίστροφο ισχύει για αρνητικό πρόσημο). Ακόμη πρέπει να αναφερθεί ότι η τιμή του συντελεστή β_i έχει φυσική σημασία. Δηλαδή, αν ο συντελεστής μιας ανεξάρτητης μεταβλητής έχει την τιμή +3,2 αυτό σημαίνει ότι για κάθε μία μονάδα αύξησης της τιμής της ανεξάρτητης προκαλεί στην εξαρτημένη μεταβλητή αύξηση κατά 3,2 μονάδες.

Για τη διερεύνηση της σχέσης μεταξύ της Y και των $X_1, X_2, \dots, X_{p-1}$ λαμβάνεται δείγμα μεγέθους n και για κάθε άτομο του δείγματος καταγράφονται οι τιμές των συγκεκριμένων μεταβλητών. Επί παραδείγματι, για το i -άτομο του δείγματος καταγράφονται οι τιμές $(Y_i, X_{i1}, X_{i2}, \dots, X_{i,p-1})$, $i = 1, 2, \dots, n$. Εφαρμόζουμε λοιπόν το μοντέλο $Y = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$, για $i = 1, 2, \dots, n$, όπου τα «σφάλματα» $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ θεωρούνται ανεξάρτητες τυχαίες μεταβλητές, ενώ οι ερμηνευτικές μεταβλητές $X_1, X_2, \dots, X_{p-1}$, όπως και στο απλό γραμμικό μοντέλο, δεν θεωρούνται τυχαίες. Η συγκεκριμένη μέθοδος ονομάζεται πολλαπλή γραμμική παλινδρόμηση (**multiple linear regression method**) και χρησιμοποιείται για να εξηγηθεί η μεταβλητότητα ή κύμανση της εξαρτημένης μίας μεταβλητής y βάσει των μεταβολών των ανεξάρτητων μεταβλητών x . Καλύτερη φυσική ερμηνεία της συσχέτισης δύο τυχαίων μεταβλητών επιτυγχάνεται με το r^2 που λέγεται συντελεστής προσδιορισμού (*coefficient of determination*) και εκφράζεται συνήθως σε ποσοστό, δηλαδή $100 \cdot r^2$. Ο συντελεστής προσδιορισμού δίνει το ποσοστό μεταβλητότητας των τιμών της Y που υπολογίζεται από τη X και είναι ένας χρήσιμος τρόπος να συνοψίσουμε τη σχέση μεταξύ των τυχαίων μεταβλητών του μοντέλου μας.



Σχήμα 4.1 Ευθεία Γραμμικής Παλινδρόμησης

Στην αναζήτηση μιας πολυπλοκότερης μεθόδου από τη γραμμική παλινδρόμηση, αλλά και για να πληρείται το κριτήριο της κανονικής κατανομής επιλέχθηκε η Λογαριθμοκανονική Παλινδρόμηση (Lognormal Regression). Η σχέση που συνδέει την ανεξάρτητη με τις εξαρτημένες μεταβλητές είναι όπως και προηγουμένως γραμμική με την διαφορά ότι στη Λογαριθμοκανονική Κατανομή μελετάται ο φυσικός αλγόριθμος της εξαρτημένης μεταβλητής. Η μαθηματική διατύπωση της Λογαριθμοκανονικής Κατανομής δίνεται από την παρακάτω σχέση:

$$\log y = \beta_0 + \beta_1 \times x_1 + \beta_2 \times x_2 + \dots + \beta_v \times x_v$$

όπου:

y : η εξαρτημένη μεταβλητή

$\beta_0, \beta_1, \beta_2, \beta_v$: οι συντελεστές της παλινδρόμησης.

Στο σημείο αυτό θα πρέπει να αναφερθεί ότι στα πλαίσια της παρούσας διπλωματικής εργασίας κατά τη στατιστική επεξεργασία των δεδομένων ως **εξαρτημένη μεταβλητή** τέθηκε ο χρόνος άφιξης λεωφορείου στη στάση.

4.2.2 Λογιστική Παλινδρόμηση

Στη συνέχεια έγινε μια διαφορετική προσέγγιση του δείγματος, καθώς μελετήθηκαν οι καθυστερήσεις στην κάθε στάση. Η εξαρτημένη τιμή των παίρνει τη τιμή (0) ή (1) ανάλογα με το αν το λεωφορείο φτάνει στην ώρα του ή καθυστερεί. Καθώς όμως η εξαρτημένη μεταβλητή στο σημείο αυτό της έρευνας είναι κατηγορική και όχι συνεχής, ποσοτική μεταβλητή, η μέθοδος που επιλέχθηκε είναι αυτή της Λογιστικής Παλινδρόμησης (Logistic Regression), που είναι ουσιαστικά μια γενίκευση της απλής γραμμικής παλινδρόμησης για τη στατιστική επεξεργασία των δεδομένων. Η Λογιστική Παλινδρόμηση είναι μία μέθοδος σχεδιασμένη για

ανάλυση δεδομένων που αφορούν την μελέτη και την πρόβλεψη τιμών μιας κατηγορικής εξαρτημένης μεταβλητής. Η μέθοδος αυτή χρησιμοποιεί ποσοτικές και ποιοτικές ανεξάρτητες μεταβλητές. Η μελέτη της σχέσης της κατηγορικής εξαρτημένης μεταβλητής δεν μπορεί να πραγματοποιηθεί μέσω του αλγορίθμου της Γραμμικής Παλινδρόμησης για δύο βασικούς λόγους:

1. Πρώτον, όταν προβλέπουμε τις τιμές μία κατηγορικής εξαρτημένης μεταβλητής, στην ουσία υπολογίζουμε την πιθανότητα με την οποία η εξαρτημένη μεταβλητή θα λάβει κάποια συγκεκριμένη τιμή. Η τιμή της πιθανότητας αυτής θα πρέπει, εξ ορισμού, να παίρνει τιμές μεταξύ του 0 και του 1. Με την χρήση της Γραμμικής Πολλαπλής Παλινδρόμησης μπορεί να υπολογιστούν τιμές πιθανότητας μεγαλύτερες του 1 ή μικρότερες του 0, δηλαδή άτοπο.
2. Δεύτερον, η πολλαπλή γραμμική παλινδρόμηση θα πρέπει να ικανοποιεί την υπόθεση της ισότητας των διακυμάνσεων. Ωστόσο, στην περίπτωση που η εξαρτημένη μεταβλητή είναι διχοτομική, έχει τυπική απόκλιση ($St\ dev$) $\sqrt{p(1-p)}$, όπου p είναι η μέση τιμή της μεταβλητής. Λόγω της συναρτησιακής σχέσης της τυπικής απόκλισης με την μέση τιμή, η ομοιογένεια της διακύμανσης των τιμών της εξαρτημένης μεταβλητής δεν είναι δυνατόν να ικανοποιείται.

Στο σημείο αυτό της συγκεκριμένης διπλωματικής εργασίας χρησιμοποιείται η μέθοδος της Λογιστικής Παλινδρόμησης προκειμένου μία ή περισσότερες ανεξάρτητες μεταβλητές να καθορίσουν το αποτέλεσμα το οποίο έχει μόνο δύο πιθανά εναλλακτικά ενδεχόμενα: (0) ή (1). Εκείνο που προκύπτει από τη λογιστική παλινδρόμηση δεν είναι μία ακριβής αριθμητική τιμή της ανεξάρτητης μεταβλητής, αλλά η πιθανότητα το ενδεχόμενο που εξάγεται να δηλώνει το ένα γεγονός του ή το εναλλακτικό του. Επομένως απαιτείται η κατηγοριοποίηση των εξαρτημένων μεταβλητών σε ζεύγη τιμών (0,1).

Στη μέθοδο της λογιστικής παλινδρόμησης, η εξαρτημένη μεταβλητή y εκφράζεται από τον φυσικό λογάριθμο του λόγου της πιθανότητας P_i να συμβεί το ένα γεγονός προς τη πιθανότητα $(1-P_i)$ να συμβεί το εναλλακτικό του και μαθηματικά αποδίδεται από τον εξής τύπο:

$$y = \ln[P_i/(1 - P_i)] = \alpha + b_i * \chi_i$$

όπου α σταθερά (constant) και οι b_i συντελεστές (coefficient estimate) των ανεξάρτητων μεταβλητών χ_i όπου $i=1, \dots, n$ το πλήθος των εξαρτημένων μεταβλητών. Επισημαίνεται ότι ενώ η πιθανότητα P παίρνει τιμές από 0 έως 1, ο φυσικός λογάριθμος $\log \left[\frac{P_i}{1-P_i} \right]$ κυμαίνεται μεταξύ μείον άπειρο και συν άπειρο. Οι συντελεστές των ανεξάρτητων μεταβλητών στην εξίσωση της παλινδρόμησης

εκτιμώνται βάσει της μεθόδου Μέγιστης Πιθανοφάνειας. Βάσει της μεθόδου αυτής η τιμή των συντελεστών των ανεξάρτητων μεταβλητών είναι αυτή που κάνει τις παρατηρηθείσες τιμές της εξαρτημένης μεταβλητής πιο πιθανές, με βάση το σετ των ανεξάρτητων μεταβλητών.

Με απλό μετασχηματισμό της παραπάνω σχέσης, προκύπτει:

$$[P_i/(1-P_i)] = e^{a+bi \cdot \chi_i} = e^a \cdot e^{bi \cdot \chi_i}$$

Όταν η τιμή μιας ανεξάρτητης μεταβλητής χ_i αυξάνεται κατά μία μονάδα, ενώ όλες οι υπόλοιπες παραμένουν σταθερές, τότε ο νέος λόγος πιθανοτήτων $[P_i/(1-P_i)]^*$ προκύπτει ως εξής:

$$[P_i/(1 - P_i)]^* = e^a \cdot e^{bi(\chi_i+1)} = e^a \cdot e^{bi \cdot \chi_i} \cdot e^{bi} = [P_i / (1-P_i)]^* \cdot e^{bi}$$

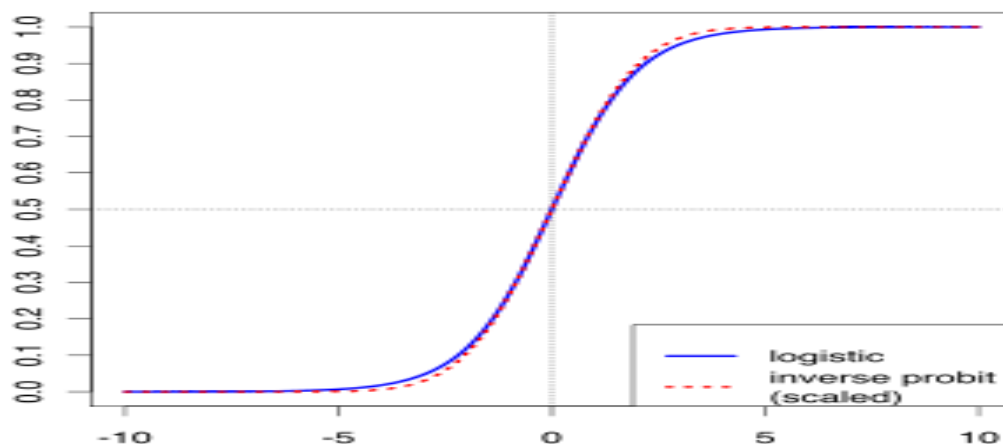
Δηλαδή όταν η τιμή μιας ανεξάρτητης μεταβλητής αυξάνεται κατά μία μονάδα, ενώ όλες οι υπόλοιπες παραμένουν σταθερές, ο νέος λόγος πιθανοτήτων είναι ίσος με τον προηγούμενο λόγο πιθανοτήτων πολλαπλασιασμένο με τον όρο e^{bi} ή $\exp(bi)$. Ο παράγοντας e^{bi} ονομάζεται λόγος πιθανοτήτων (odds ratio, OR) και λαμβάνει τιμές από το 0 έως το άπειρο.

Με την θεώρηση ότι η P_i είναι η πιθανότητα να συμβεί το γεγονός (1), ενώ $(1-P_i)$ είναι η πιθανότητα να συμβεί το γεγονός (0), από τη σχέση:

$$\left[\frac{P_i}{1-P_i}\right]^* = [P_i / (1-P_i)]^* \cdot e^{bi}$$

Προκύπτει ότι όταν το OR είναι μεγαλύτερο της μονάδας είναι περισσότερο πιθανό να συμβεί το γεγονός (1), ενώ στην αντίθετη περίπτωση είναι περισσότερο πιθανό να συμβεί το γεγονός (0). (Simon P. Washington et al, 2003)

Στο σχήμα 4.2 παρουσιάζεται γραφικά η καμπύλη του λογιστικού μοντέλου το οποίο έχει δύο ασύμπτωτα σημεία, ένα ελάχιστο στην τιμή 0 και ένα μέγιστο στην τιμή 1. Έτσι, οι εκτιμώμενες πιθανότητες του μοντέλου της λογιστικής παλινδρόμησης κυμαίνονται μεταξύ των τιμών 0 και 1, το οποίο αποτελεί ένα ρεαλιστικό επίπεδο κατανομών (Narumalani et al, 1997). Αυτή η συμπεριφορά του λογιστικού μοντέλου αποτελεί ένα σημαντικό πλεονέκτημα έναντι άλλων στατιστικών μεθόδων παλινδρόμησης όπως η γραμμική παλινδρόμηση στην οποία οι εκτιμήσεις μπορεί να κυμαίνονται έξω από το εύρος 0-1. Η μεγαλύτερη κλίση της κατανομής παρουσιάζεται στο σημείο $P=1/2$. Αυτό συνεπάγεται ότι οι μεταβολές στις ανεξάρτητες μεταβλητές στο μέσο κατανομής, θα έχουν μεγαλύτερη επιρροή στην πιθανότητα επιλογής μιας εναλλακτικής λύσης.



Σχήμα 4.2 Καμπύλη Λογιστικής Παλινδρόμησης

4.3 Βήματα Κατασκευής Μοντέλου Γραμμικής Παλινδρόμησης και Λογιστικής Παλινδρόμησης – Κριτήρια Αποδοχής

Τα βήματα κατασκευής των μοντέλων Παλινδρόμησης είναι τα εξής:

1. Προσδιορίζουμε το μέγεθος του ενδιαφέροντος (εξαρτημένη μεταβλητή) και το σύνολο των ανεξάρτητων μεταβλητών που θα συμμετέχουν στην παλινδρόμηση.
2. Διερευνούμε τα δεδομένα για τυχόν ύπαρξη ασυνήθιστων κινήσεων όπως, ακραίες τιμές, ελλείπουσες τιμές κ. λ. π.
3. Ελέγχουμε την ικανοποίηση των υποθέσεων για την σωστή εφαρμογή της Παλινδρόμησης.
4. Δημιουργούμε την εξίσωση της παλινδρόμησης.
5. Μελετάμε την επίδραση κάθε ανεξάρτητης μεταβλητής στο μοντέλο.
6. Εξετάζουμε την ικανοποίηση των υποθέσεων της Τεχνικής.

Επιπλέον, σε περίπτωση που το μοντέλο πρόκειται να χρησιμοποιηθεί για πρόβλεψη θα πρέπει να αξιολογηθεί η αποτελεσματικότητά του. Αυτό σημαίνει ότι δημιουργείται η εξίσωση σε ένα μέρος των δεδομένων και σε ένα επόμενο βήμα ελέγχεται η αποτελεσματικότητά της, στο υπόλοιπο δείγμα.

Τα **βασικά κριτήρια αποδοχής** της μεθόδου της Γραμμικής και Λογιστικής Παλινδρόμησης είναι τα εξής:

- Οι ανεξάρτητες μεταβλητές σε μια εξίσωση δεν πρέπει να σχετίζονται μεταξύ τους, δηλαδή πρέπει να είναι μεταξύ τους γραμμικώς ανεξάρτητες. Η μη πλήρωση αυτής της συνθήκης είναι μία από τις συνηθέστερες πηγές σφαλμάτων, αφού τότε είναι δύσκολο να εξακριβωθεί η επιρροή κάθε μεταβλητής στο τελικό αποτέλεσμα.

- Ένα δεύτερο κριτήριο που χρησιμοποιείται προκειμένου να ελεγχθεί η αξιοπιστία και η καταλληλότητα της εξίσωσης παλινδρόμησης αφορά στο αν τα **πρόσημα των συντελεστών της παλινδρόμησης** αντιβαίνουν τη λογική ή αν μπορούν να ερμηνευτούν με βάση τις συγκεκριμένες συνθήκες και αντιλήψεις.
- Το τρίτο και τελευταίο κριτήριο έχει σχέση με το σταθερό όρο της εξίσωσης της παλινδρόμησης. Ο σταθερός αυτός όρος, που εκφράζει το σύνολο των παραμέτρων, πρέπει να δίνει εξηγήσιμα συμπεράσματα.

Αν η τελική εξίσωση δεν πληροί μια από τις παραπάνω προϋποθέσεις, τότε δεν γίνεται αποδεκτή από τον ερευνητή.

4.4 Λειτουργία του Λογισμικού SPSS

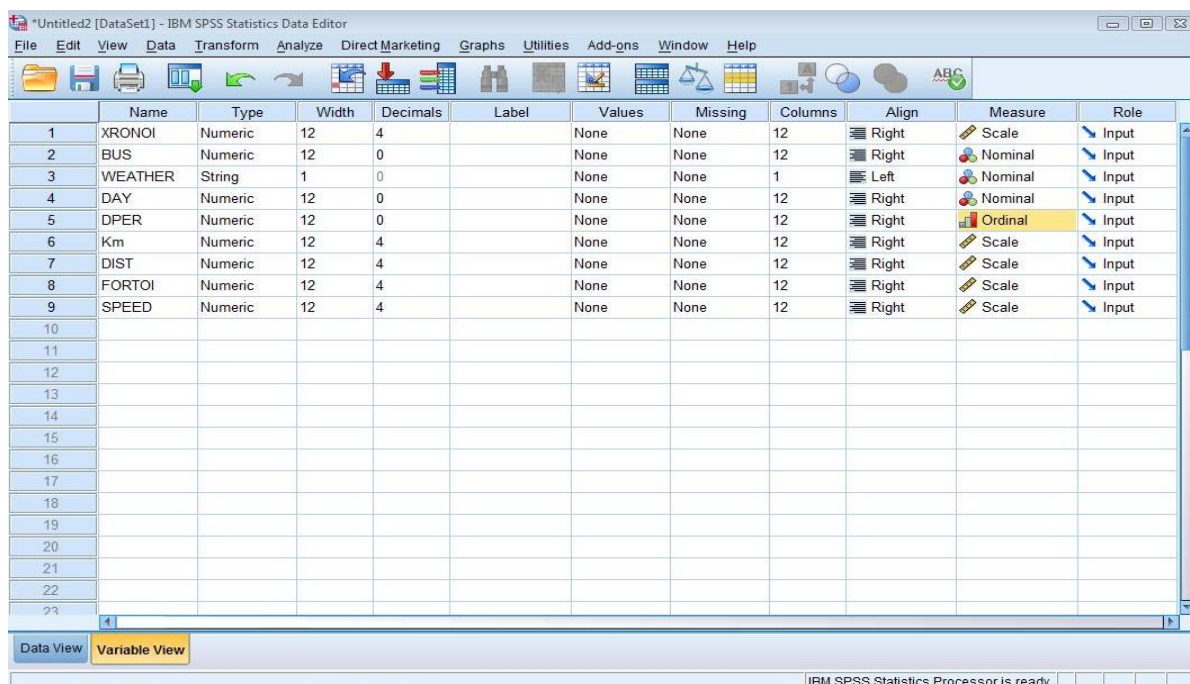
Η ανάλυση παλινδρόμησης έγινε με τη βοήθεια του στατιστικού προγράμματος SPSS (STATISTICAL PACKAGE FOR SOCIAL SCIENCE). Μετά την ολοκλήρωση της ανάλυσης στο MICROSOFT EXCEL, τα στοιχεία τοποθετήθηκαν σε αρχεία και μεταφέρθηκαν στο στατιστικό πρόγραμμα SPSS, όπου με τη διαδικασία **REGRESSION** σχηματίστηκαν οι τελικές εξισώσεις. Μέσω της διαδικασίας αυτής είναι δυνατόν να υπολογιστούν ορισμένα στατιστικά στοιχεία που αναφέρονται είτε στην εξίσωση παλινδρόμησης, είτε σε κάθε μία από τις ανεξάρτητες μεταβλητές.

4.4.1 Λειτουργία SPSS για τη Γραμμική Παλινδρόμηση

Παρακάτω παρατίθεται ο τρόπος λειτουργίας του Λογισμικού SPSS για την πρόβλεψη του μοντέλου της Γραμμικής Παλινδρόμησης, καθώς και χαρακτηριστικά παραδείγματα του τρόπου επεξεργασίας με τη βοήθεια διαδοχικών εικόνων εκτέλεσης επεξεργασίας των στοιχείων στο SPSS. Πρώτο βήμα που ακολουθείται είναι η ρύθμιση του είδους της μεταβλητής, όπως φαίνεται στην εικόνα 4.1. Στα Measures ρυθμίζουμε ανάλογα με το αν η μεταβλητή είναι ποσοτική (scale), διατεταγμένη (ordinal) ή ονομαστική (nominal).

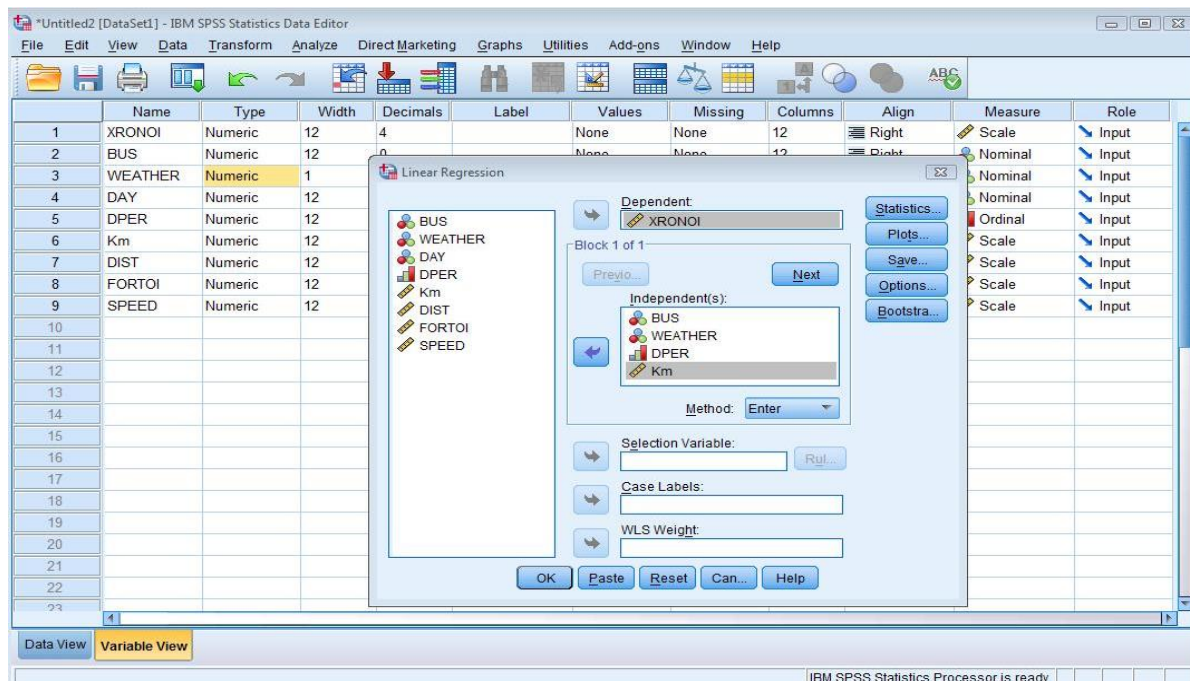
Για την Ενεργοποίηση των πλαισίων διαλόγου της Γραμμικής Παλινδρόμησης, στο SPSS, ακολουθούμε την διαδρομή:

Analyze → Regression → Linear



Εικόνα 4.1 Αρχική σελίδα εισαγωγής δεδομένων

Στη συνέχεια εισάγουμε την εξαρτημένη και τις ανεξάρτητες μεταβλητές. Στα Πεδία με την ένδειξη “Dependent” και “Independent(s)” αντίστοιχα όπως φαίνεται στην εικόνα 4.2



Εικόνα 4.2 Τρόπος εισαγωγής μεταβλητών στη Γραμμική Παλινδρόμηση

Η επιλογή των μεταβλητών που εισάγονται σε μια εξίσωση παλινδρόμησης καθορίζονται αυτόματα από το πρόγραμμα με τη βοήθεια ορισμένων κριτηρίων.

Ανάλογα με το είδος του κριτηρίου εισαγωγής μιας μεταβλητής, η λειτουργία της κάθε μεθόδου είναι:

- **Μέθοδος ENTER** : Στη μέθοδο αυτή εισάγονται διαδοχικά όλες οι μεταβλητές που παρουσιάζουν την τιμή tolerance μεγαλύτερη από 10^{-4} . Η μέθοδος αυτή παρουσιάζει τα πιο ελαστικά κριτήρια από τις υπόλοιπες μεθόδους.
- **Μέθοδος FORWARD** : Στο πρώτο βήμα της μεθόδου ελέγχεται η εισαγωγή στην εξίσωση της μεταβλητής με το μεγαλύτερο, κατά απόλυτη τιμή μερικό συντελεστή συσχέτισης.
- **Μέθοδος BACKWARD** : Αρχικά εισάγονται όλες οι ανεξάρτητες μεταβλητές στην εξίσωση παλινδρόμησης. Στη συνέχεια εξετάζεται αν η μεταβλητή με τον μικρότερο μερικό συντελεστή συσχέτισης ικανοποιεί ένα από τα κριτήρια FOUT (2,71) ή ROUT (0,10).
- **Μέθοδος STEPWISE** : Η μέθοδος αυτή αποτελεί συνδυασμό των δύο προηγούμενων μεθόδων και για το λόγο αυτό χρησιμοποιούνται πιο συχνά.

Για τη μόρφωση των εξισώσεων παλινδρόμησης στην παρούσα εργασία χρησιμοποιήθηκε η μέθοδος **ENTER** και επιλέχθηκε η μέθοδος της γραμμικής παλινδρόμησης (linear regression) από το λογισμικό SPSS, ενώ τέθηκαν ως εξαρτημένες μεταβλητές ο χρόνος άφιξης λεωφορείου στη στάση για το συνολικό δρομολόγιο 3 στην πόλη της Θεσσαλονίκης, καθώς και για τη διαδρομή εντός του κέντρου.

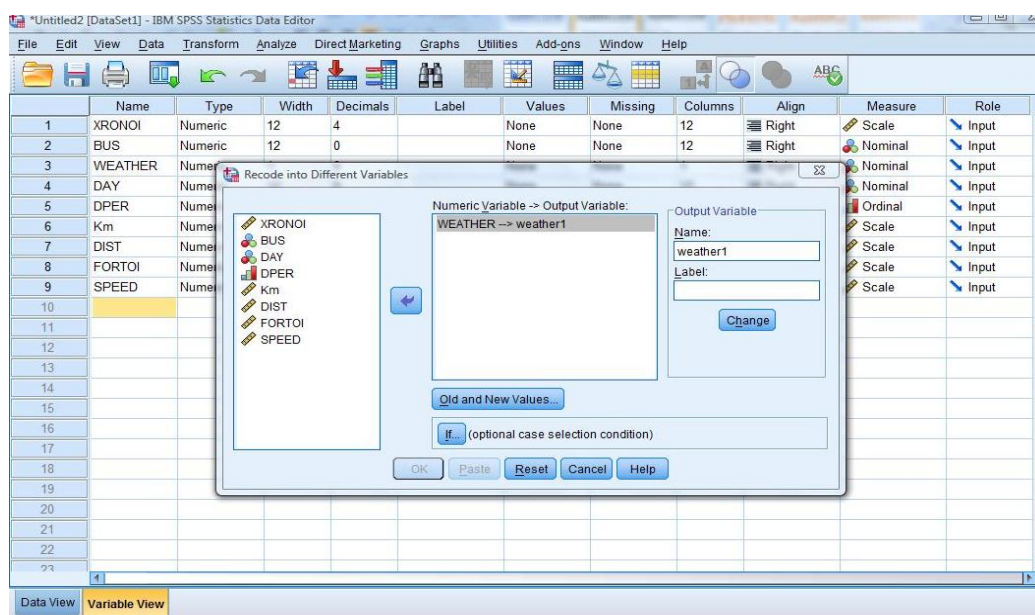
Όπως έχει αναφερθεί, η τεχνική της γραμμικής παλινδρόμησης δίνει απαντήσεις σε ερωτήματα για τη φύση της σχέσης ενός εξαρτημένου, συνεχούς μεγέθους με ένα ή περισσότερα συνεχή ανεξάρτητα μεγέθη. Ωστόσο, τις περισσότερες φορές επιθυμούμε να εξετάσουμε το εξαρτημένο μέγεθος και ως προς κάποιες διακριτές/ κατηγορικές μεταβλητές. Αφού οριστούν οι κατηγορικές μεταβλητές, προχωράμε στη διαδικασία του μετασχηματισμού τους σε ψευδομεταβλητές (dummies variables). Οι κανόνες βάσει των οποίων δημιουργούμε ψευδομεταβλητές είναι οι εξής:

- ✓ Οι ψευδομεταβλητές μπορούν να πάρουν τιμές μόνο 0 και 1.
- ✓ Για κάθε κατηγορική μεταβλητή που θα συμμετάσχει στην παλινδρόμηση και που μπορεί να πάρει (n) πιθανές τιμές, μπορούν να εισαχθούν στο μοντέλο (n-1) ψευδομεταβλητές.
- ✓ Κάθε μια ψευδομεταβλητή παίρνει τιμή 1 για μια κατηγορία ενδιαφέροντος της κατηγορικής μεταβλητής και για όλες τις υπόλοιπες παίρνει τιμή 0.

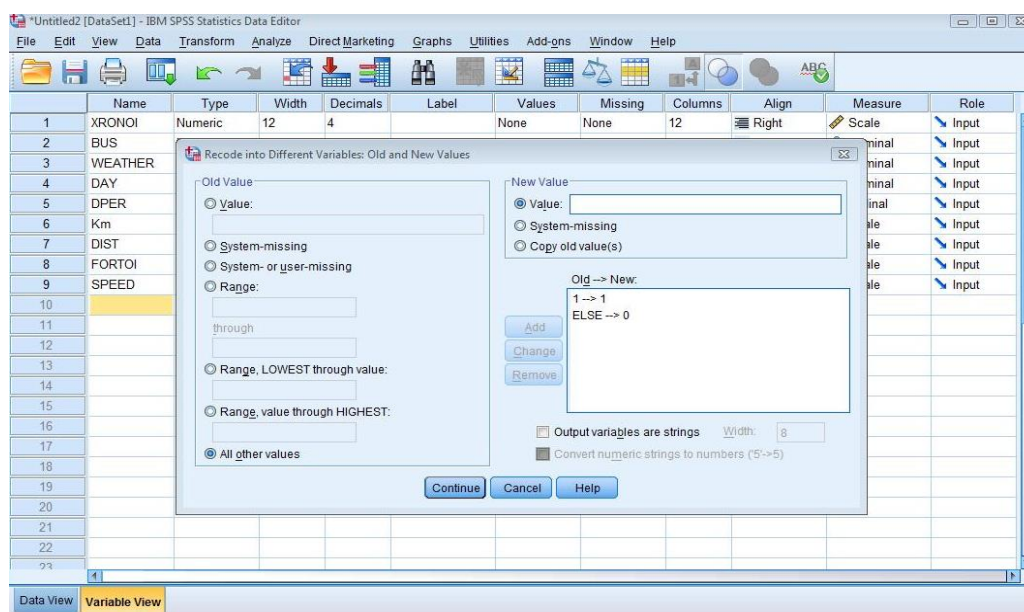
Η διαδικασία δημιουργίας ψευδομεταβλητών στο λογισμικό SPSS γίνεται ακολουθώντας τη διαδρομή:

Transform → Recode into Different Variables

Στη συνέχεια στο παράθυρο Recode into Different Variables ορίζει ο ερευνητής μια ανεξάρτητη μεταβλητή θέλει να επεξεργαστεί και στο NAME ορίζει το όνομα της καινούργιας μεταβλητής που πρόκειται να δημιουργηθεί βάση της προεπιλεγμένης (εικόνα 4.3). Με την επιλογή Old and New Values ορίζονται πλέον οι τιμές της καινούργιας μεταβλητής (εικόνα 4.4).

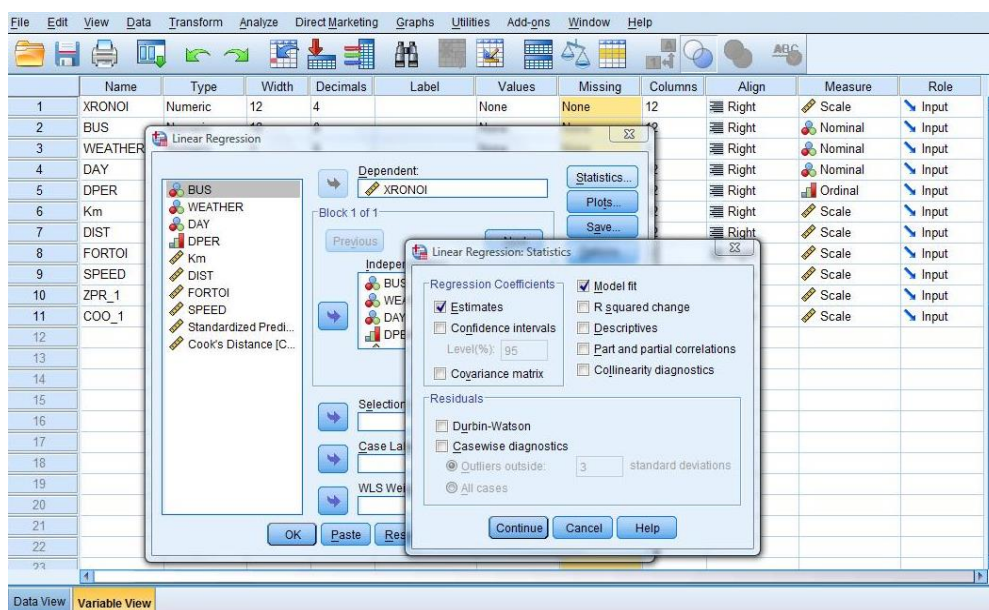


Εικόνα 4.3 Δημιουργία μεταβλητών με βάση κάποια αρχική στο SPSS



Εικόνα 4.4 Ορισμός νέων τιμών μεταβλητών στο SPSS

Μετά τη δημιουργία των ψευδομεταβλητών ξεκινά η διαδικασία της Γραμμικής Παλινδρόμησης. Κατά την εφαρμογή της τεχνικής από τα πλαίσια διαλόγου του SPSS μπορούμε να χρησιμοποιήσουμε ένα πλήθος δεικτών που σχετίζονται με το εξεταζόμενο θέμα και να αντλήσουμε από αυτούς πληροφορίες για τη στατιστική σημαντικότητα του μοντέλου που προκύπτει, για την προσαρμογή του στα εξεταζόμενα δεδομένα ή για την γενικότερη επιτυχία εφαρμογής της τεχνικής. Πιο συγκεκριμένα, επιλέγοντας το πλήκτρο **Statistics** εμφανίζεται ο παρακάτω πίνακας:



Εικόνα 4.5 Statistics στη Γραμμική Παλινδρόμηση

Οι σημαντικότερες από αυτές τις επιλογές είναι:

Estimates: Οι συντελεστές του μοντέλου παλινδρόμησης και σχετικά με αυτούς μέτρα όπως το τυπικό τους σφάλμα και ο δείκτης significant.

Confidence Interval: Ένα 95% διάστημα εμπιστοσύνης για την εκτίμηση των συντελεστών της γραμμικής παλινδρόμησης.

Model fit: Στατιστικοί δείκτες όπως το R , R^2 , ή ο προσαρμοσμένος συντελεστής R^2 .

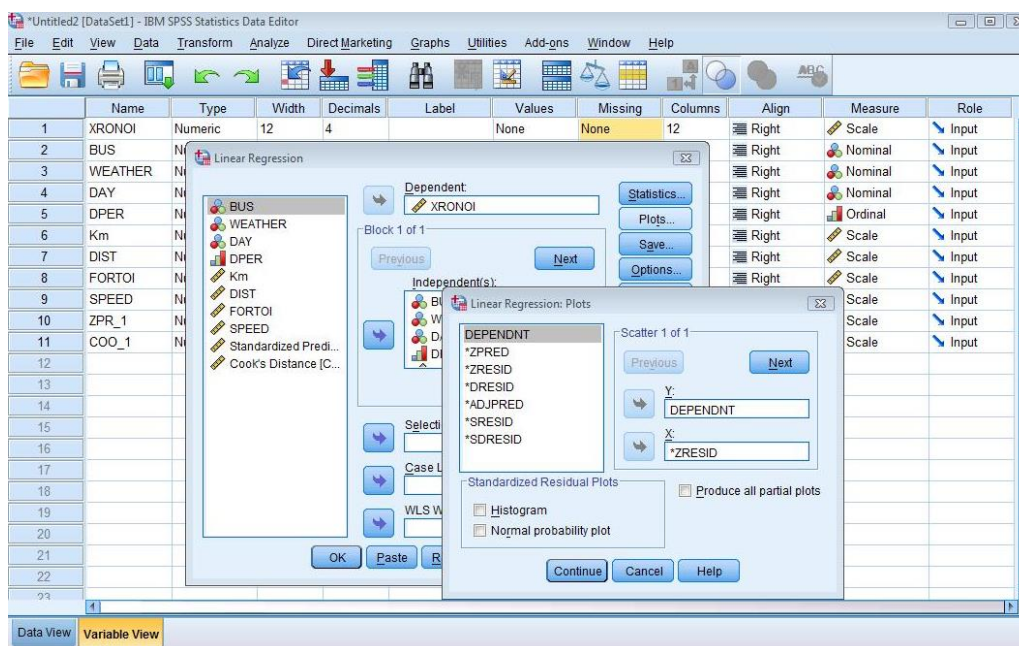
R Square Change: Επιλογή που δίνει πληροφορίες για την αλλαγή της τιμής R^2 με την προσθήκη ή τη διαγραφή μιας ανεξάρτητης μεταβλητής.

Durbin-Watson: Στατιστικός έλεγχος για τη σειριακή συσχέτιση των καταλοίπων. Με την επιλογή αυτή εμφανίζονται τα περιγραφικά στατιστικά μέτρα των καταλοίπων και των προβλεπόμενων από το μοντέλο τιμών.

Από το πλήκτρο **Plot** του πλαισίου διαλόγου της γραμμικής παλινδρόμησης ενεργοποιείται το παράθυρο που επιτρέπει τη δημιουργία γραφημάτων. Μέσω αυτού μπορούμε να δημιουργήσουμε:

Normal Probability Plots: Για τα κανονικοποιημένα κατάλοιπα, το διάγραμμα αυτό χρησιμοποιείται για τον έλεγχο της κανονικότητας των καταλοίπων.

Produce all Partial Plots: Για την κατασκευή *scatter plots* για κάθε μια από τις ανεξάρτητες με τα κατάλοιπα της εξαρτημένης.



Εικόνα 4.6 Plots στη Γραμμική Παλινδρόμηση

Στη συνέχεια, από το πλήκτρο **Save** του βασικού πλαισίου της παλινδρόμησης, αποθηκεύουμε σε μορφή μεταβλητών αποτελέσματα της ανάλυσης. Οι σημαντικότερες από τις επιλογές αναφέρονται παρακάτω:

Unstandardized Predicted Values: Για την αποθήκευση της εξαρτημένης μεταβλητής, έτσι όπως αυτή υπολογίζεται από την εξίσωση της παλινδρόμησης

Standardized Predicted Values: Για την αποθήκευση των κανονικοποιημένων προβλεπόμενων τιμών

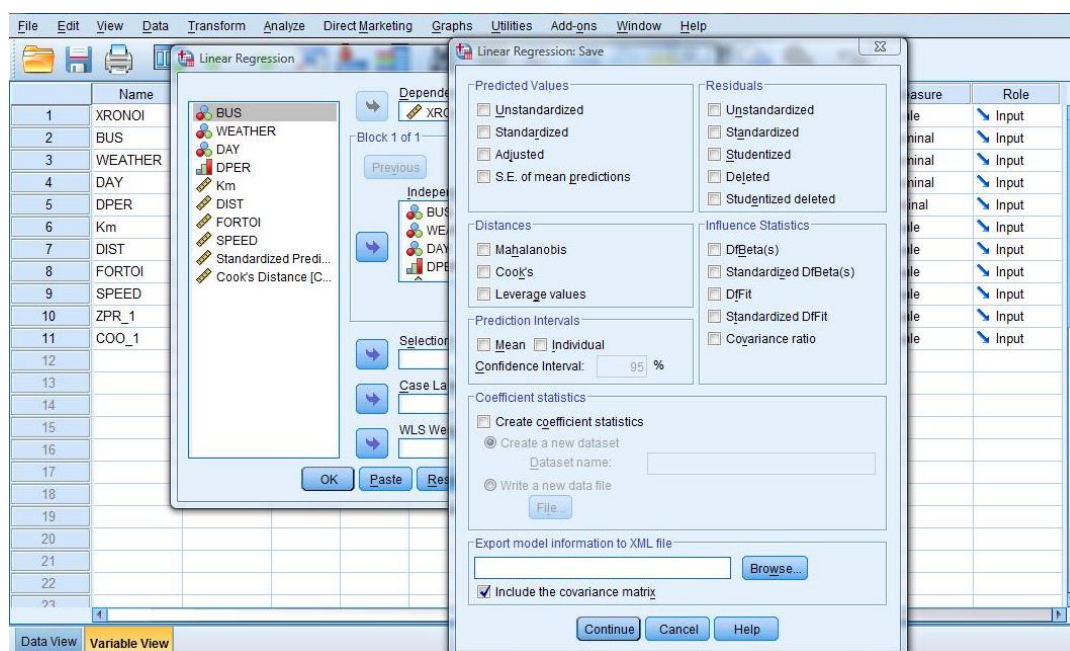
Adjusted: Για την αποθήκευση των προβλεπόμενων τιμών για κάθε εγγραφή όταν αυτή έχει εξαιρεθεί από τον υπολογισμό των συντελεστών της παλινδρόμησης

Unstandardized Residual: Για την αποθήκευση των κατάλοιπων για κάθε εγγραφή

Standardized Residual: Για την αποθήκευση των κανονικοποιημένων τιμών των καταλοίπων

Mahalanobis: Για αποθήκευση ενός μέτρου που εκφράζει τη διαφορά της εγγραφής της εξαρτημένης μεταβλητής από το μέσο όρο όλων των εγγραφών

Cook's: Για την αποθήκευση του μέτρου που εκφράζει για κάθε ξεχωριστή εγγραφή, πόσο θα μεταβληθούν τα κατάλοιπα αν η εγγραφή αυτή αφαιρεθεί από τον υπολογισμό των συντελεστών παλινδρόμησης



Εικόνα 4.7 Πλήκτρο Save στη Γραμμική Παλινδρόμηση

Τέλος, από το πλήκτρο **Options** μπορούμε να ορίσουμε το επίπεδο σημαντικότητας για την αποδοχή ή όχι του μοντέλου γραμμικής παλινδρόμησης. Επιπλέον, μπορούμε να αφαιρέσουμε το σταθερό όρο του μοντέλου. Η επιλογή *Missing Values* επιτρέπει την επέμβαση στον τρόπο χρήσης των ελλειπουσών τιμών.

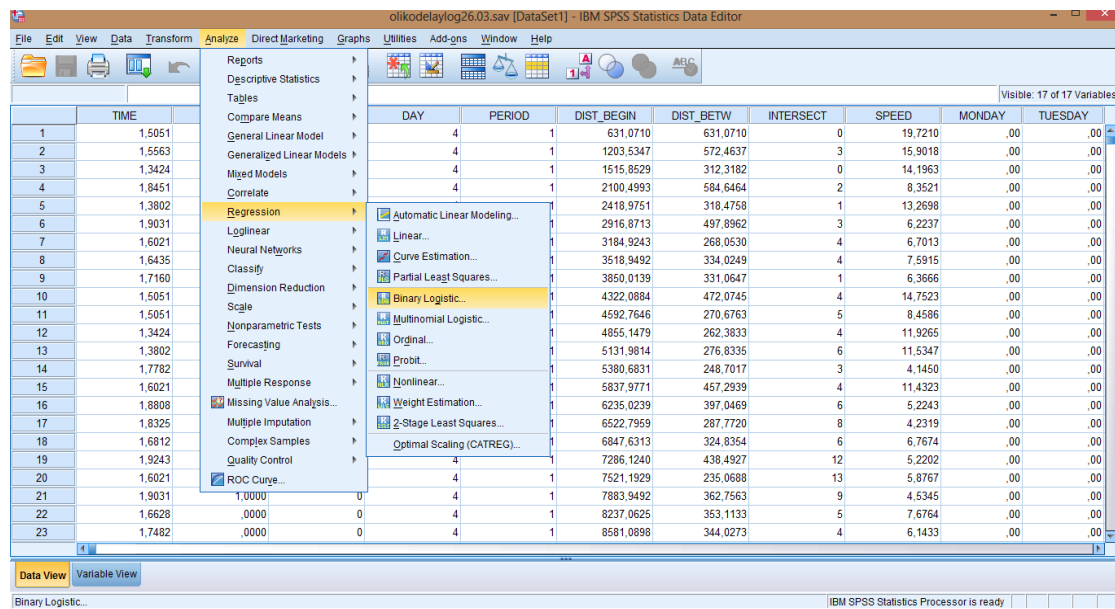
4.4.2 Λειτουργία SPSS για τη Λογιστική Παλινδρόμηση

Στη συνέχεια μελετήθηκε η καθυστέρηση του λεωφορείου με τη μέθοδο της Λογιστικής Παλινδρόμησης με το στατιστικό πρόγραμμα SPSS. Μέσω της διαδικασίας αυτής είναι δυνατόν να υπολογιστούν ορισμένα στατιστικά στοιχεία που αναφέρονται είτε στην εξίσωση παλινδρόμησης, είτε σε κάθε μία από τις ανεξάρτητες μεταβλητές.

Παρακάτω παρατίθεται ο τρόπος λειτουργίας του Λογισμικού SPSS για την πρόβλεψη του μοντέλου της Λογιστικής Παλινδρόμησης, καθώς και χαρακτηριστικά παραδείγματα του τρόπου επεξεργασίας με τη βοήθεια διαδοχικών εικόνων εκτέλεσης επεξεργασίας των στοιχείων στο SPSS.

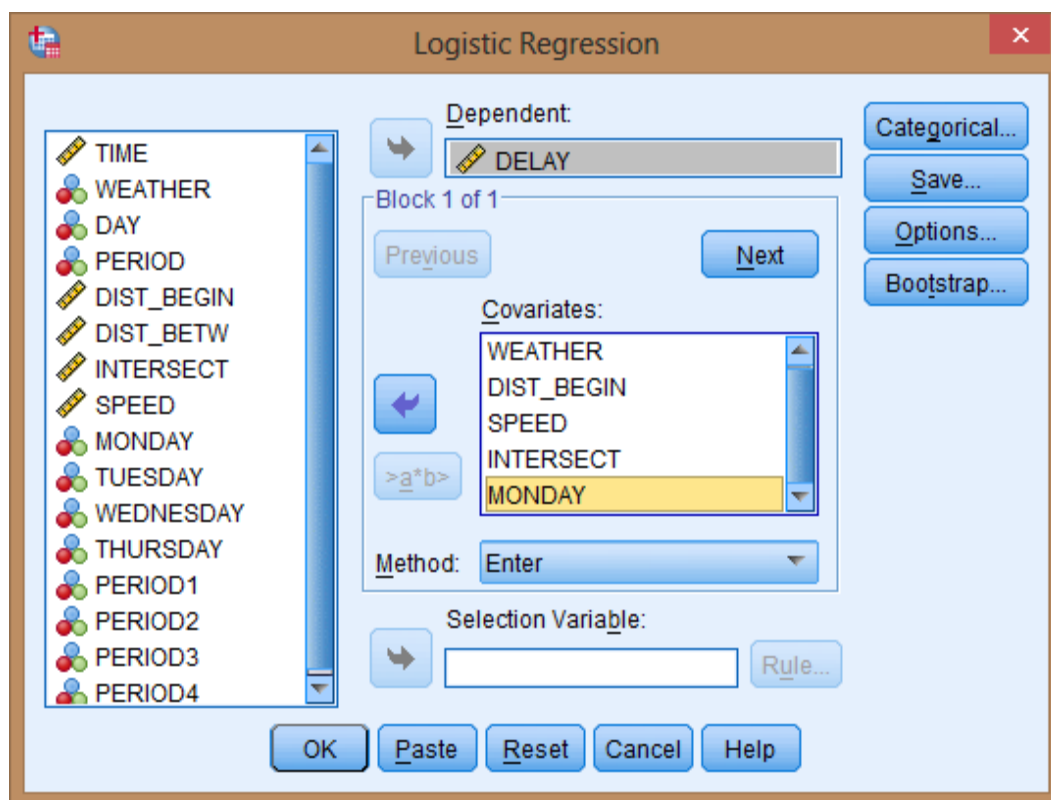
Για την Ενεργοποίηση των πλαισίων διαλόγου της Λογιστικής Παλινδρόμησης, στο SPSS, ακολουθούμε την διαδρομή:

Analyze → Regression → Binary Logistic (αν η εξαρτημένη μεταβλητή παίρνει δύο μόνο τιμές).



Εικόνα 4.8 Ενεργοποίηση λογιστικής παλινδρόμησης στο SPSS.

Στο Πεδίο με την ένδειξη **“dependent”** εισάγουμε την εξαρτημένη μεταβλητή και στο Πεδίο **“Covariates”** εισάγουμε τις ανεξάρτητες μεταβλητές.



Εικόνα 4.9 Τρόπος εισαγωγής μεταβλητών στη λογιστική παλινδρόμηση.

Για τη μόρφωση των εξισώσεων Λογιστικής Παλινδρόμησης στην παρούσα εργασία χρησιμοποιήθηκε η μέθοδος **ENTER** από το λογισμικό SPSS, ενώ τέθηκαν ως εξαρτημένες μεταβλητές η καθυστέρηση του λεωφορείου στη στάση για τη συνολική διαδρομή, καθώς και για τη διαδρομή εντός του κέντρου.

Για την εφαρμογή της η Λογιστική Παλινδρόμηση απαιτεί κατηγορική εξαρτημένη μεταβλητή και κατηγορικές ή ποσοτικές ανεξάρτητες μεταβλητές. Και οι κατηγορικές και οι ποσοτικές ανεξάρτητες μεταβλητές εισάγονται στο πλαίσιο “Covariates”. Όπως και στην γραμμική παλινδρόμηση οι ποσοτικές μεταβλητές εισάγονται στην αρχική τους μορφή.

Οι κατηγορικές μεταβλητές προκειμένου να εισαχθούν στο μοντέλο, πρέπει να μετασχηματισθούν σε ψευδομεταβλητές (dummy variables). Το πλεονέκτημα στην εφαρμογή της Λογιστικής Παλινδρόμησης, σε σχέση με την γραμμική, είναι ότι η δημιουργία των ψευδομεταβλητών μπορεί να πραγματοποιηθεί μέσα από το Menu από το πλήκτρο “**Categorical**”. Για την δημιουργία των ψευδομεταβλητών θα πρέπει μετά τον ορισμό όλων των ανεξάρτητων μεταβλητών να προσδιοριστούν εκείνες που είναι κατηγορικές και για τις οποίες θα πρέπει να δημιουργηθούν ψευδομεταβλητές. Στο αριστερό μέρος του παραθύρου “**Define Categorical Variables**” εμφανίζεται η λίστα των ανεξάρτητων μεταβλητών που ορίστηκαν στο βασικό παράθυρο της τεχνικής. Από όλες τις ανεξάρτητες μεταβλητές θα πρέπει να

μεταφερθούν το πλαίσιο “Categorical Covariates” εκείνες που είναι κατηγορικές με περισσότερες από δύο κατηγορίες.

Το SPSS, στο σημείο αυτό, παρέχει τη δυνατότητα δημιουργίας ψευδομεταβλητών με διάφορες διαδικασίες. Οι διαδικασίες κατασκευής των ψευδομεταβλητών βρίσκονται στην αναδυόμενη λίστα επιλογών που βρίσκεται δίπλα στην ένδειξη Contrast και από την οποία μπορούμε να επιλέξουμε εκείνη τη διαδικασία που ταιριάζει στην κωδικογράφηση των κατηγορικών μεταβλητών μας.

Οι καλύτερες επιλογές για την δημιουργία ψευδομεταβλητών είναι :

Indicator : Προτείνεται εξ ορισμού από το SPSS και είναι η πιο συνηθισμένη διαδικασία κατασκευής ψευδομεταβλητών. Κατά τη διαδικασία αυτή παράγονται οι ψευδομεταβλητές όπως φαίνεται στον πίνακα παρακάτω. Βάση αυτής της κωδικογράφησης μπορούμε να μελετήσουμε την επίδραση, κάποιας συγκεκριμένης κατηγορίας της κατηγορικής μεταβλητής στην εξαρτημένη μεταβλητή σε σχέση με μία άλλη κατηγορία. Κατά την διαδικασία δημιουργίας των ψευδομεταβλητών δημιουργούνται $p-1$ μεταβλητές, όπου p ο αριθμός των κατηγοριών της αρχικής μεταβλητής.

Κατηγορική Μεταβλητή	1 ^η Ψευδομεταβλητή	2 ^η Ψευδομεταβλητή
1	1	1
2	0	1
3	0	0

Deviation: Είναι ένας εναλλακτικός τρόπος δημιουργίας Ψευδομεταβλητών. Χρησιμοποιείται όταν θέλουμε να συγκρίνουμε την επίδραση κάθε κατηγορίας με την μέση επίδραση των υπόλοιπων κατηγοριών. Στην περίπτωση αυτή δημιουργούμε ξανά $p-1$ ψευδομεταβλητές με την μορφή του παρακάτω πίνακα.

Κατηγορική Μεταβλητή	1 ^η Ψευδομεταβλητή	2 ^η Ψευδομεταβλητή
1	1	0
2	0	1
3	-1	-1

Η διαδικασία κατασκευής ψευδομεταβλητών μπορεί να πραγματοποιηθεί χωρίς τη χρήση προκαθορισμένων επιλογών που προσφέρει το λογισμικό του SPSS (όπως π.χ. Indicator), αλλά από τον ίδιο τον ερευνητή με τη ίδια ακριβώς διαδικασία που περιγράφηκε παραπάνω στη Γραμμική Παλινδρόμηση:

Transform → Recode into Different Variables

Στο παράθυρο Recode into Different Variables ορίζεται πια ανεξάρτητη μεταβλητή θα επεξεργαστεί και στο NAME ορίζεται το όνομα της καινούργιας μεταβλητής που πρόκειται να δημιουργηθεί βάση της προεπιλεγμένης. Με την επιλογή Old and New Values ορίζονται πλέον οι τιμές της καινούργιας μεταβλητής.

Μετά τον ακριβή ορισμό των μεταβλητών που θέλουμε να συμμετέχουν στο μοντέλο της παλινδρόμησης θα πρέπει να προσδιορίσουμε τον αλγόριθμο εισαγωγής τους στον μοντέλο. Θα πρέπει να σημειωθεί ότι κανένας από τους αλγόριθμους δημιουργίας των μοντέλων δεν δημιουργεί το βέλτιστο μοντέλο. Κάθε μία διαδικασία οδηγεί σε διαφορετικό αποτέλεσμα το οποίο ο ερευνητής πρέπει κάθε φορά να αξιολογεί.

Από το πλήκτρο Save του βασικού πλαισίου της παλινδρόμησης, αποθηκεύουμε σε μορφή μεταβλητών αποτελέσματα της ανάλυσης. Οι επιλογές κάτω από τον τίτλο **Predicted Values** έχουν σχέση με τις προβλεπόμενες από το μοντέλο τιμές για την εξαρτημένη μεταβλητή. Συγκεκριμένα οι επιλογές είναι:

Propabilities: υπολογίζει για κάθε εγγραφή του αρχείου την πιθανότητα να συμβεί το γεγονός του ενδιαφέροντος, ενώ,

Group membeship: υπολογίζει την κατηγορία της εξαρτημένης μεταβλητής στην οποία κατατάσσεται η συγκεκριμένη εγγραφή βάση της πιθανότητας που έχει υπολογίσει το μοντέλο.

Οι επιλογές κάτω από τον τίτλο **“Residuals”** αφορούν τα κατάλοιπα. Συγκεκριμένα, οι σημαντικότερες επιλογές αναφέρονται παρακάτω:

Unstandardized: αποθηκεύει στο αρχείο των δεδομένων την διαφορά της πραγματικής τιμής, της εξαρτημένης μεταβλητής, από την προβλεπόμενη τιμή του μοντέλου.

Logit: αποθηκεύει την τιμή των καταλοίπων δια της σχέσης $p(1-p)$.

Standardized: αποθηκεύει τις κανονικοποιημένες τιμές των καταλοίπων.

Οι επιλογές κάτω από τον τίτλο **“Influence”** αφορούν επιδράσεις συγκεκριμένων τιμών στο παραγόμενο μοντέλο. Συγκεκριμένα οι επιλογές είναι:

Cooks: υπολογίζει για κάθε εγγραφή, πόσο αλλάζουν τα κατάλοιπα αν η συγκεκριμένη εγγραφή εξαιρεθεί από τον υπολογισμό των συντελεστών της παλινδρόμησης.

Leverage values: υπολογίζει την επίδραση κάθε παρατήρησης στο fit του μοντέλου.

Dfbeta(s): υπολογίζει την διαφορά που παρατηρείται στους συντελεστές της παλινδρόμησης αν δεν συμμετάσχει η συγκεκριμένη εγγραφή.

Μετά την επιλογή των μεταβλητών που θέλουμε να αποθηκεύσουμε στο αρχείο των δεδομένων επιλέγουμε από το πλήκτρο **Options** μερικά μέτρα για τον έλεγχο της στατιστικής σημαντικότητας του μοντέλου που θα δημιουργήσουμε. Συγκεκριμένα οι επιλογές που δίνονται είναι:

Classification Plots: κατασκευάζει ιστόγραμμα για τις πραγματικές και για τις προβλεπόμενες, βάση του μοντέλου, τιμές της εξαρτημένης μεταβλητής,

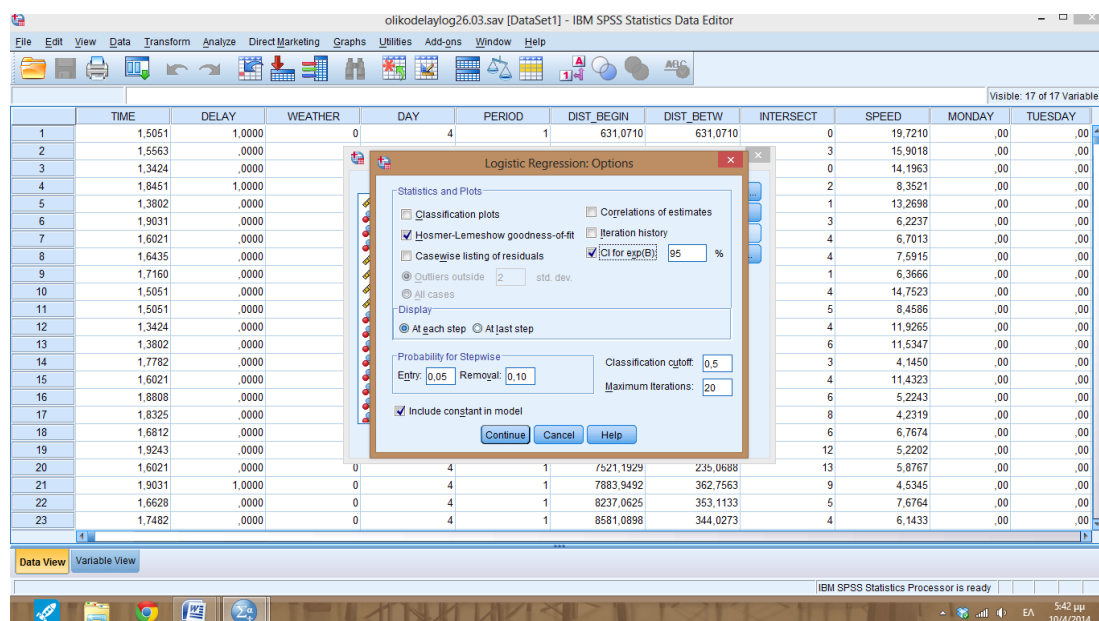
Hosmer-Lemeshow goodness of fit: υπολογίζει ένα δείκτη για την συνολική προσαρμογή του μοντέλου.

casewise listing of residuals: εμφανίζει έναν πίνακα με τις τιμές των καταλοίπων, των εκτιμημένων πιθανοτήτων, των πραγματικών τιμών της εξαρτημένης μεταβλητής και τον τιμών που προβλέφθηκαν γι' αυτή βάση του μοντέλου.

CI for exp(B): υπολογίζει ένα διάστημα εμπιστοσύνης για τις εκτιμήσεις του $\text{Exp}(B)$.

Από αυτό το παράθυρο διαλόγου, μπορούμε επίσης, να δηλώσουμε αν επιθυμούμε στο αρχείο των αποτελεσμάτων να εμφανίζονται όλα τα βήματα δημιουργίας του μοντέλου, καθώς και την ύπαρξη ή όχι σταθερού όρου στο μοντέλο της παλινδρόμησης. Επίσης, μπορούμε να επέμβουμε στον προσδιορισμό της τιμής του επιπέδου στατιστικής σημαντικότητας για την εισαγωγή και την εξαγωγή μίας μεταβλητής από το μοντέλο κατά την δημιουργία του.

Πολύ σημαντική είναι, τέλος, η επιλογή **“Classification Cutoff”** γιατί η ορισμένη τιμή της πιθανότητας είναι αυτή που προσδιορίζει σε πια κατηγορία της εξαρτημένης μεταβλητής θα καταταχθεί κάθε εγγραφή βάσης του μοντέλου.



Εικόνα 4.10 Επιλογές στη Λογιστική Παλινδρόμηση

4.5 Κριτήρια Αποδοχής Αποτελεσμάτων

Κατά την εξαγωγή αποτελεσμάτων το στατιστικό πρόγραμμα SPSS προσφέρει τη δυνατότητα στο μελετητή να ελέγξει άμεσα το αποτέλεσμα της στατιστικής επεξεργασίας, παρέχοντάς του ορισμένες πληροφορίες (output) με τη μορφή δεικτών που είναι καθοριστικοί για την επιλογή του κατάλληλου μοντέλου.

4.5.1 Κριτήρια Αποδοχής Αποτελεσμάτων στη Γραμμική Παλινδρόμηση

Κατά τη διαδικασία της Γραμμικής Παλινδρόμησης τα αποτελέσματα ελέγχονται βάσει των παρακάτω στατιστικών δεικτών, οι οποίοι συγκρίνονται για την επιλογή του τελικού μοντέλου:

- Ο συντελεστής συσχέτισης του Pearson R^2 αποτελεί ένα από τα βασικότερα μέτρα αξιολόγησης του μοντέλου. Ο συντελεστής αυτός προσδιορίζει το βαθμό συσχέτισης της εξαρτημένης με τις ανεξάρτητες μεταβλητές και εκφράζει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής που εξηγείται από το μοντέλο. Οι τιμές του κυμαίνονται από μηδέν μέχρι ένα. Όσο πλησιέστερα βρίσκεται στη μονάδα τόσο καλύτερα περιγράφει την κατάσταση του προτύπου. Με άλλα λόγια, με το συντελεστή αυτό αξιολογούμε κάθε φορά την καταλληλότητα του προτύπου.
- Οι συντελεστές που προέκυψαν από τον έλεγχο **t-test** (t-coefficients) που καταδεικνύουν πόσο στατιστικά σημαντική είναι η κάθε μια από τις ανεξάρτητες μεταβλητές x . Οι έλεγχοι t-test διεξάγονται προκειμένου να διερευνηθεί εάν η κλίση της γραμμής της παλινδρόμησης διαφέρει σημαντικά από το μηδέν. Οι συντελεστές αυτοί θα πρέπει να ικανοποιούν την εξίσωση $|t_i| \geq 1,645$, όπου 1,645 είναι η κρίσιμη τιμή t της κατανομής *Student* για επίπεδο σημαντικότητας $\alpha=0.10$ (10%), ενώ σε καμία περίπτωση δεν δεχόμαστε τιμές του t μικρότερες από 1,5. Αν πληρείται η συνθήκη $|t_i| \geq 1,645$, τότε απορρίπτεται η μηδενική υπόθεση (**Null hypothesis** ή **H₀**) ότι δηλαδή οι δύο μεταβλητές x και y είναι ασυσχέτιστες ($\beta=0$) και επαληθεύεται η εναλλακτική υπόθεση (**Alternative hypothesis** ή **H₁**), σύμφωνα με την οποία οι δύο μεταβλητές συσχετίζονται.
- Οι γραμμικοί συντελεστές παλινδρόμησης β_i (**B Coefficients** ή **Regression Coefficients**) που αποτελούν τις κλίσεις (slopes) της ευθείας της παλινδρόμησης και πρακτικά καταδεικνύουν την επίπτωση (μέγεθος, πρόσημο) που θα έχει στην τιμή της εξαρτημένης μεταβλητής y η μεταβολή της τιμής της κάθε ανεξάρτητης x μεταβλητής μεμονωμένα κατά μία μονάδα. Έτσι, θετικό πρόσημο του β_i σημαίνει θετική γραμμική σχέση μεταξύ της συγκεκριμένης μεταβλητής x και της εξαρτημένης y και συνεπώς, αύξηση του x κατά μία μονάδα θα επιφέρει αύξηση της τιμής του y κατά β_i μονάδες. Αντίστοιχα, αρνητικό πρόσημο του β_i συνεπάγεται αρνητική γραμμική σχέση

μεταξύ της συγκεκριμένης μεταβλητής x και της εξαρτημένης y , άρα αύξηση του x κατά μία μονάδα θα προκαλέσει μείωση της τιμής του y κατά β_i μονάδες.

- Τέλος, όσον αφορά τα σφάλματα, είναι απαραίτητος ο έλεγχος Durbin-Watson. Για να μην υπάρχει πρόβλημα αυτοσυσχέτισης μεταξύ διαδοχικών σφαλμάτων θα πρέπει σε γενικές γραμμές να ισχύει $1,7 \leq DW \leq 2,2$.

Αν $1,4 < DW < 1,7$, δεν μπορούμε να καταλήξουμε σε ασφαλές συμπέρασμα.

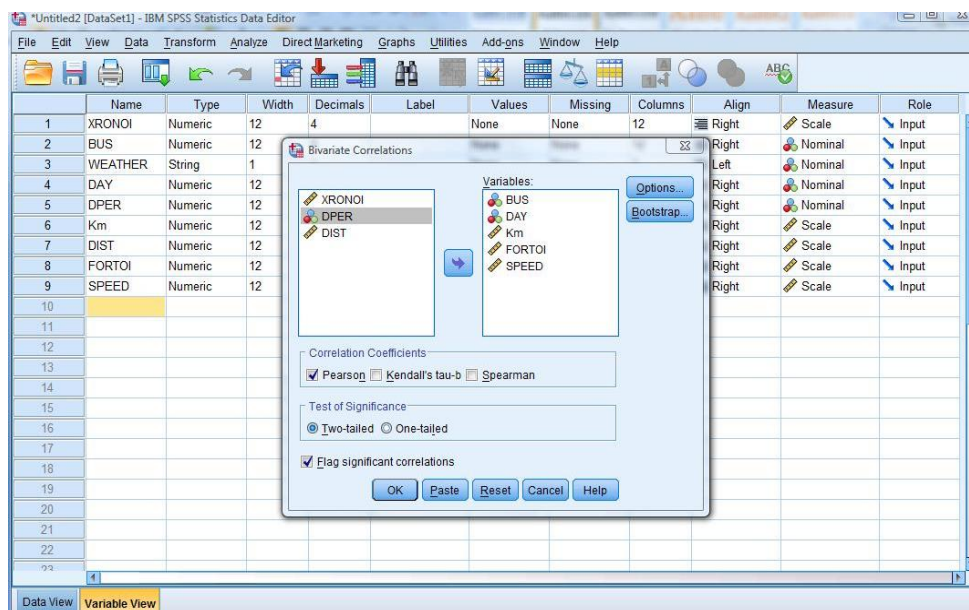
Αν $DW < 1,4$, υπάρχει πρόβλημα αυτοσυσχέτισης (autocorrelation) στο μοντέλο μας.

Επιπροσθέτως, προκειμένου να πραγματοποιηθεί σωστά η Γραμμική Παλινδρόμηση και να προκύψουν ασφαλή αποτελέσματα πρέπει να έχει πραγματοποιηθεί ο **έλεγχος συσχέτισης** των ανεξάρτητων μεταβλητών, καθώς οι ανεξάρτητες μεταβλητές που χρησιμοποιήθηκαν για τον προσδιορισμό της εξαρτημένης δεν πρέπει να σχετίζονται μεταξύ τους. Με το στατιστικό πρόγραμμα SPSS πραγματοποιήθηκε ο έλεγχος συσχέτισης μεταξύ των ανεξάρτητων μεταβλητών που χρησιμοποιήθηκαν στην στατιστική ανάλυση ως εξής:

Analyze-> Correlate->Bivariate και επιλέγω το συντελεστή συσχέτισης του Pearson.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} * \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Πρόκειται για έναν καθαρό αριθμό μεταξύ -1 και 1. Όταν $r=0$ δεν υπάρχει γραμμική συσχέτιση μεταξύ των X και Y χωρίς αυτό να αποκλείει την ύπαρξη κάποιας σχέσης άλλης μορφής. Όταν $r=+1$ υπάρχει θετική γραμμική εξάρτηση, ενώ όταν $r=-1$ υπάρχει αρνητική γραμμική εξάρτηση. Απόλυτες τιμές του συντελεστή αυτού στο $[0,0.3]$ υποδηλώνουν ασθενή γραμμική εξάρτηση, στο $(0.3,0.6]$ μεσαία, ενώ στο $(0.6,1]$ ισχυρή. Όταν η τιμή του r είναι μεγαλύτερη του 0.6 μια από τις δυο μεταβλητές δεν μπορεί να μπει στο μοντέλο.



Εικόνα 4.8 Διαδικασία συσχέτισης στο SPSS.

4.5.2 Κριτήρια Αποδοχής Αποτελεσμάτων στη Λογιστική Παλινδρόμηση

Αντίστοιχα τα σημαντικότερα στατιστικά στοιχεία που περιλαμβάνονται στα δεδομένα εξόδου του προγράμματος κατά τη Λογιστική Παλινδρόμηση είναι τα ακόλουθα:

- Ο συντελεστής συσχέτισης R^2 . Και συγκεκριμένα στη Λογιστική Παλινδρόμηση καλό είναι να κοιτάμε το R^2 Nagelkerke. Ο συντελεστής αυτός προσδιορίζει το βαθμό συσχέτισης της εξαρτημένης με τις ανεξάρτητες μεταβλητές. Όσο πλησιέστερα βρίσκεται στη μονάδα τόσο καλύτερα περιγράφει την κατάσταση του προτύπου. Στο σημείο αυτό καλό θα ήταν να αναφερθεί ότι οι τιμές του R^2 που παίρνει ένα μοντέλο Λογιστικής Παλινδρόμησης είναι σε γενικές γραμμές χαμηλότερες από αυτές που παίρνει ένα μοντέλο γραμμικής παλινδρόμησης.
- Ο δεύτερος δείκτης που χρησιμοποιείται στον έλεγχο αποτελεσμάτων είναι ο **Wald** ο οποίος είναι ανάλογος του **t-test (κριτήριο t της κατανομής student)** που χρησιμοποιείται στη γραμμική παλινδρόμηση. Μάλιστα η σχέση που τους συνδέει είναι η εξής: $t \cong \sqrt{W}$. Ο δείκτης αυτός εκφράζει αν κάθε μία από τις ανεξάρτητες μεταβλητές είναι **στατιστικά σημαντική** βάση της χ^2 κατανομής και πρέπει να συμπεριληφθεί στο τελικό πρότυπο ή όχι. Γίνεται αρχικά μια υπόθεση H_0 ότι η ανεξάρτητη μεταβλητή δεν επηρεάζει την εξαρτημένη (δηλαδή ο συντελεστής β είναι μηδενικός) και αυτή η υπόθεση αντιστοιχεί σε μία τιμή πίνακα *Wald-value ανάλογα με το επίπεδο σημαντικότητας. Αν Wald>*Wald-value τότε η μηδενική υπόθεση απορρίπτεται και αυτό σημαίνει ότι η ανεξάρτητη μεταβλητή επηρεάζει την

εξαρτημένη και συνεπώς είναι στατιστικά σημαντική. Όσο μεγαλύτερη η τιμή του Wald τόσο μεγαλύτερη είναι και η επιρροή της συγκεκριμένης μεταβλητής στο τελικό αποτέλεσμα.

- Ο έλεγχος σημαντικότητας μιας ανεξάρτητης μεταβλητής γίνεται και από το δείκτη **Sig.**, ο οποίος εκφράζει το **επίπεδο σημαντικότητας**. Για επίπεδο σημαντικότητας 95% $\rightarrow \text{Sig.}=0,05$. Σε αυτή τη περίπτωση αν $\text{Sig.}>0,05$ τότε ισχύει η μηδενική υπόθεση και η ανεξάρτητη μεταβλητή δεν είναι στατιστικώς σημαντική.
- Οι συντελεστές β_i εκφράζει τον τρόπο που επηρεάζει η κάθε ανεξάρτητη μεταβλητή την πιθανότητα να πραγματοποιηθεί το συγκεκριμένο γεγονός που έχει επιλεγεί (του έχει δώσει την τιμή 1). Το θετικό πρόσημο υποδηλώνει την θετική επιρροή της συγκεκριμένης ανεξάρτητης μεταβλητής να πραγματοποιηθεί το συγκεκριμένο γεγονός, ενώ το αρνητικό πρόσημο υποδηλώνει ότι η συγκεκριμένη ανεξάρτητη μεταβλητή δεν προωθεί την πραγματοποίηση του συγκεκριμένου γεγονότος αλλά του εναλλακτικού που του έχει δοθεί η τιμή 0. Ιδιαίτερη προσοχή πρέπει να δοθεί στην ερμηνεία των συντελεστών β_i διότι στην λογιστική παλινδρόμηση, σε αντίθεση με τη γραμμική, η εξαρτημένη με τις ανεξάρτητες μεταβλητές δεν συνδέονται γραμμικά διότι $y = \log \left[\frac{P_i}{1-P_i} \right]$.
- Ένας ακόμα σημαντικός δείκτης είναι αυτός της πιθανοφάνειας (Likelihood). Ο δείκτης **-2LL** είναι μείον δύο φορές ο λογάριθμος της πιθανοφάνειας. Το βέλτιστο μοντέλο θα έχει πιθανοφάνεια (Likelihood) ίση με 1. Άρα τιμή $-2LL=0$. Δηλαδή, όσο η τιμή του δείκτη **-2LL** πλησιάζει στο μηδέν τόσο καλύτερο είναι το μοντέλο.

Όπως και προηγουμένως για να πραγματοποιηθεί σωστά η Λογιστική Παλινδρόμηση και να προκύψουν ασφαλή αποτελέσματα πρέπει οι ανεξάρτητες μεταβλητές που χρησιμοποιήθηκαν να μην σχετίζονται μεταξύ τους. Ο **έλεγχος συσχέτισης** μεταξύ των ανεξάρτητων μεταβλητών που χρησιμοποιήθηκαν στην στατιστική ανάλυση γίνεται ως εξής:

Analyze-> Correlate->Bivariate και κλικ στην επιλογή **Spearman**.

Η διαφορά τώρα είναι ότι η συσχέτιση μεταξύ των ανεξάρτητων μεταβλητών υπολογίστηκε με τη βοήθεια του συντελεστή συσχέτισης Spearman και χρησιμοποιείται για μεταβλητές με τιμές κατάταξης και όχι καθαρά αριθμητικές. Επίσης, για την εύρεση του συντελεστή συσχέτισης Spearman δεν απαιτείται η κανονικότητα στην κατανομή των μεταβλητών. Ο τύπος υπολογισμού του είναι ο εξής:

$$r_s = 1 - \frac{6 * SUM(D^2)}{n * (n^2 - 1)}$$

όπου,

- n : ο συνολικός αριθμός παρατηρήσεων
- D : η διαφορά κατάταξης στα ζεύγη των συγκρινόμενων μεταβλητών (difference in rank of pairs x and y).

Ο συντελεστής συσχέτισης υποδεικνύει τη σχέση που μπορεί να υπάρχει μεταξύ δύο μεταβλητών και δίνει μια αίσθηση εξάρτησης ή ανεξαρτησίας ανάμεσα στα μεγέθη που εξετάζονται. Οι αποδεκτές τιμές και η σημασία τους είναι η ίδια με εκείνες της γραμμικής παλινδρόμησης.

4.6 To Likelihood Ratio Statistic Test

Το Likelihood Ratio Statistic Test είναι ένα τεστ που χρησιμοποιείται στη Λογιστική Παλινδρόμηση με το οποίο μπορεί να γίνει σύγκριση μεταξύ δύο μοντέλων έτσι ώστε να επιλεγεί αυτό που θα παρουσιάζει την καλύτερη συσχέτιση ανεξάρτητων μεταβλητών με την εξαρτημένη. Προϋπόθεση για την σύγκριση είναι το ένα μοντέλο να περιλαμβάνει όλες τις μεταβλητές του άλλου μοντέλου συν κάποιες επιπλέον.

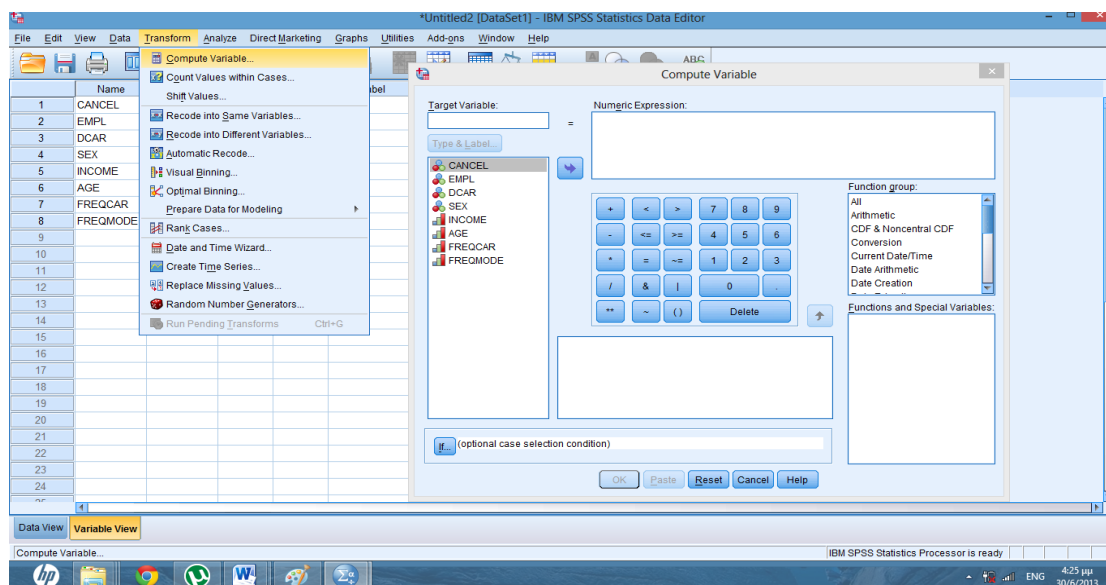
Μοντέλο 1: $y = \log[P_i/(1 - P_i)] = \beta_0 + \beta_1 X_1$

Μοντέλο 2: $y = \log[P_i/(1 - P_i)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

Μας ενδιαφέρει το κατά πόσο η επιπλέον ανεξάρτητη μεταβλητή του δεύτερου μοντέλου απαιτείται έτσι ώστε να ερμηνευτεί η εξαρτημένη μεταβλητή. Αυτό που ψάχνει ο ερευνητής είναι το αν το απλό μοντέλο είναι το ίδιο κατάλληλο με το πιο σύνθετο μοντέλο. Με άλλα λόγια εξετάζει τη μηδενική υπόθεση H_0 (null hypothesis) $\beta_2 = 0$ έναντι της εναλλακτικής υπόθεσης H_1 $\beta_2 \neq 0$.

Το Likelihood Ratio Test για να εξετάσει τη μηδενική υπόθεση βασίζεται στον δείκτη $-2LL$, ο οποίος μετράει το βαθμό ασυμφωνίας μεταξύ των παραταθέντων και των προβλεπόμενων τιμών από το μοντέλο. Είναι ένας δείκτης που επί της ουσίας εκφράζει το πόση ανεξήγητη πληροφορία υπάρχει μετά την εφαρμογή του μοντέλου. Η τιμή αυτού του δείκτη για το μικρότερο μοντέλο θα είναι μεγαλύτερη από την αντίστοιχη τιμή του δείκτη για το μεγαλύτερο μοντέλο. Η διαφορά του δείκτη ανάμεσα στο Μοντέλο 1 και στο Μοντέλο 2 ονομάζεται Likelihood Ratio (LR) test statistic και είναι αυτή που χρησιμοποιούμε για να εξετάσουμε τη μηδενική μας υπόθεση.

Η διαδικασία είναι να βρεθεί η LR test statistic και να συγκριθεί με μια τιμή πίνακα από μία chi-square κατανομή για p βαθμούς ελευθερίας. Αν η LR τιμή είναι πολύ μεγαλύτερη από τη τιμή του πίνακα τότε η μηδενική υπόθεση H_0 απορρίπτεται.



Εικόνα 4.9 Διαδικασία εφαρμογής του Likelihood Ratio Test στο SPSS.

Στο στατιστικό πρόγραμμα SPSS η σύγκριση γίνεται με τον εξής τρόπο:

Transform ->Compute Variable και στο **Numeric Expression** τοποθετούμε το LR και τους p βαθμούς ελευθερίας στη μορφή $1 - \text{CDF.CHISQ}(\text{LR}, p)$.

5. ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ – ΑΠΟΤΕΛΕΣΜΑΤΑ

5.1 Εισαγωγή

Στο παρόν κεφάλαιο της διπλωματικής εργασίας εφαρμόζεται η μεθοδολογία που έχει αναλυθεί παραπάνω, δηλαδή επεξεργάζεται στατιστικά η βάση δεδομένων, για την διεξαγωγή των τελικών προτύπων. Στόχος είναι η εύρεση των παραμέτρων που επηρεάζουν το χρόνο άφιξης του λεωφορείου στη στάση, αλλά και των παραμέτρων που επιδρούν στην καθυστέρηση.

Τη βάση δεδομένων αποτέλεσαν στοιχεία που δόθηκαν από τον Οργανισμό Αστικών Συγκοινωνιών Θεσσαλονίκης και αφορούν στη Λεωφορειακή γραμμή 3 Ανατολικός Σταθμός ΙΚΕΑ-Τερματικός Σταθμός Γ. Οι ημέρες που μελετήθηκαν ήταν οι 24, 28, 29, 30, 31 Ιανουαρίου και 7, 13 Φεβρουαρίου 2013. Τα στοιχεία που μελετήθηκαν στην παρούσα εργασία είναι ένας συνδυασμός των στοιχείων σε πραγματικό χρόνο (δελτίο κίνησης οχήματος) και των κυκλοφοριακών στοιχείων από φωρατές κυκλοφορίας τοποθετημένους στο κέντρο της πόλης.

Η προσέγγιση του αντικειμένου μελέτης έχει πραγματοποιηθεί μέσω του στατιστικού πακέτου SPSS το οποίο χρησιμοποιείται ευρέως στις μελέτες του Συγκοινωνιολόγου Πολιτικού Μηχανικού. Παρακάτω παρατίθενται αναλυτικά τα τελικά αποτελέσματα, τα μοντέλα της Παλινδρόμησης που επιλέχθηκαν, αλλά και οι παράγοντες βάσει των οποίων επιλέχθηκαν αυτά. Συγκεκριμένα, έχουν μελετηθεί δύο μοντέλα, το ολικό και του κέντρου, ως προς τους χρόνους και τις καθυστερήσεις.

Αρχικά πρέπει να γίνει μια αναφορά στις ανεξάρτητες και εξαρτημένες μεταβλητές που χρησιμοποιήθηκαν στα μοντέλα παλινδρόμησης:

Ανεξάρτητες μεταβλητές:

Με τον όρο ανεξάρτητη μεταβλητή αναφερόμαστε σε κάθε πλήρως ορισμένο στοιχείο, το οποίο εκφράζει είτε μια ιδιότητα, είτε ένα μέγεθος ή ένα συγκεκριμένο χαρακτηριστικό και το οποίο επιδρά συστηματικά, και με τρόπο που μπορεί να εκτιμηθεί, σε ένα άλλο, επίσης, καθοριζόμενο στοιχείο (εξαρτημένη μεταβλητή).

Για να επιλεγεί ένα στοιχείο και να καθορισθεί ως ανεξάρτητη μεταβλητή πρέπει να πληροί ορισμένους κανόνες που αφορούν τη φύση και τα ιδιαίτερα χαρακτηριστικά του. Συνεπώς, για να γίνει κατάλληλη αξιολόγηση των επιμέρους χαρακτηριστικών που πρόκειται να χαρακτηριστούν ως ανεξάρτητες μεταβλητές,

απαιτείται η πλήρης καταγραφή τους και η άμεση μελέτη των ιδιοτήτων που αυτά εκφράζουν. Ούτως ή άλλως, για την επιλογή ενός στοιχείου ως ανεξάρτητης μεταβλητής κατά τη διεξαγωγή μιας μελέτης αξιολόγησης, είναι απαραίτητη, αρχικά, μια αρχική εκτίμηση για το κατά πόσο επηρεάζει τη μεταβλητή και για τη χρησιμότητάς του ως εξωγενούς παράγοντα επίδρασης.

Ως ανεξάρτητες μεταβλητές στην παρούσα διπλωματική εργασία έχουν ορισθεί οι παρακάτω:

1. Ο καιρός (WEATHER)
2. Η ημέρα (DAY)
3. Η περίοδος (PERIOD)
4. Η απόσταση στάσης από την αφετηρία (DIST_BEGIN)
5. Η απόσταση μεταξύ των στάσεων (DIST_BETW)
6. Ο αριθμός των διασταυρώσεων μεταξύ των στάσεων (INTERSECT)
7. Η ταχύτητα (SPEED)
8. Ο φόρτος (VOLUME)

Στην παρούσα μελέτη αναπτύσσονται δύο μαθηματικά πρότυπα, το ενοποιημένο το οποίο αφορά στη συνολική διαδρομή του δρομολογίου 3 και εκείνο που αφορά μόνο στη διαδρομή του κέντρου. Στο ενοποιημένο μοντέλο οι ανεξάρτητες μεταβλητές είναι όλες εκτός του φόρτου, ενώ η μέση ταχύτητα έχει υπολογιστεί από τους χρόνους και τις αποστάσεις μεταξύ. Αντίθετα, στο δεύτερο πρότυπο που αναπτύσσεται και αφορά το κέντρο οι ανεξάρτητες μεταβλητές όλες, με τις ταχύτητες και τους φόρτους να έχουν μετρηθεί από τους φωρατές κυκλοφορίας.

Εξαρτημένη μεταβλητή:

Με τον όρο εξαρτημένη μεταβλητή αναφερόμαστε σε κάθε στοιχείο το οποίο αντιπροσωπεύει μια ιδιότητα ή ένα χαρακτηριστικό και τα χαρακτηριστικά του εξαρτώνται και επηρεάζονται από εξωγενείς παράγοντες που αναφέρονται ως ανεξάρτητες μεταβλητές.

Στην παρούσα διπλωματική εργασία μελετάται η εκτίμηση της επίδρασης των ανεξάρτητων μεταβλητών, οι οποίες ορίστηκαν στην προηγούμενη παράγραφο, στο χρόνο άφιξης του λεωφορείου στη κάθε στάση, καθώς επίσης και στην καθυστέρηση σε σχέση με τους αναμενόμενους χρόνους, για τη συνολική διαδρομή και για το κέντρο. Δηλαδή, στην παρούσα διπλωματική εργασία γίνεται επεξεργασία δύο εξαρτημένων μεταβλητών:

1. Χρόνος άφιξης λεωφορείου στη στάση (TIME)
2. Καθυστέρηση (DELAY)

5.2 Εισαγωγή δεδομένων στο πρόγραμμα

Μετά την αρχική επεξεργασία των στοιχείων, κατασκευάστηκαν τα συγκεντρωτικά Excel με τις μεταβλητές ομαδοποιημένες και κωδικοποιημένες ανάλογα με τα είδη των μαθηματικών προτύπων που είχαν αποφασισθεί να χρησιμοποιηθούν στο στατιστικό πρόγραμμα SPSS. Κατασκευάστηκαν λοιπόν δύο ξεχωριστά excel, συνολικό και κέντρο, με στήλες τις μεταβλητές του μοντέλου (εξαρτημένη – ανεξάρτητες) και ονομασίες που συμβαδίζουν με τις απαιτήσεις του προγράμματος. Το excel για τη συνολική γραμμή έχει μέγεθος 32246 γραμμές στοιχείων, ενώ για το κέντρο 7099 γραμμές στοιχείων.

Για τη λειτουργία του προγράμματος SPSS απαιτείται ο καθορισμός τριών αρχείων: **το αρχείο δεδομένων εισόδου** (data editor file), **το αρχείο ελέγχου** (control file) και **το αρχείο αποτελεσμάτων** (SPSS viewer). Από αυτά τα δύο πρώτα είναι αρχεία που περιέχουν στοιχεία που εισάγουμε στο πρόγραμμα, ενώ το τρίτο είναι κενό και είναι εκείνο στο οποίο θα καταχωρηθούν τα αποτελέσματα της στατιστικής ανάλυσης.

Πιο συγκεκριμένα, στο **αρχείο εισόδου** εισάγονται αρχικά τα αρχεία της βάσης δεδομένων όπως έχουν επεξεργαστεί στα αντίστοιχα φύλλα του EXCEL. Το **αρχείο ελέγχου** είναι εκείνο στο οποίο καθορίζονται από το χρήστη οι μεταβλητές που περιέχονται στο αρχείο εισόδου, δηλαδή επιλέγεται ποια από τις μεταβλητές θα είναι η εξαρτημένη και ποιες οι ανεξάρτητες.

Στη συνέχεια κατασκευάστηκαν δύο διαφορετικά Excel για τα μαθηματικά μοντέλα που επρόκειτο να αναπτυχθούν στο SPSS, ειδικότερα:

Για το χρόνο άφιξης λεωφορείου στη στάση ως εξαρτημένη μεταβλητή (TIME) δημιουργήθηκαν 2 μαθηματικά πρότυπα που αναφέρονται:

- a. Στη συνολική διαδρομή
- b. Στο κέντρο

Και αντίστοιχες ανεξάρτητες:

1. Ο καιρός (WEATHER)
2. Η ημέρα (DAY)
3. Η περίοδος (PERIOD)
4. Η απόσταση στάσης από την αφετηρία (DIST_BEGIN)
5. Η απόσταση μεταξύ των στάσεων (DIST_BETW)
6. Ο αριθμός των διασταυρώσεων μεταξύ των στάσεων (INTERSECT)
7. Η ταχύτητα (SPEED)
8. Ο φόρτος (VOLUME)

Η ανεξάρτητη τιμή της ταχύτητας (SPEED) υπολογίζεται διαφορετικά στο πρότυπο της συνολικής διαδρομής και του κέντρου, ενώ ο φόρτος (VOLUME) περιλαμβάνεται μόνο εντός του κέντρου.

Για την καθυστέρηση σε κάθε στάση σε σχέση με το αναμενόμενο ως ανεξάρτητη μεταβλητή (DELAY) δημιουργήθηκαν επίσης 2 μαθηματικά πρότυπα που αναφέρονται, για τα οποία ισχύουν ότι και προηγουμένως.

Παρακάτω παρατίθενται δύο χαρακτηριστικά αποσπάσματα πινάκων τα οποία δημιουργήθηκαν μετά την επεξεργασία των αρχικών Δελτίων Κίνησης Οχήματος και τον Κυκλοφοριακών Στοιχείων, τα οποία έχουν κωδικοποιηθεί έτσι ώστε να είναι συμβατά με τις απαιτήσεις του SPSS. Οι δύο παρακάτω πίνακες αφορούν τους χρόνους και τις καθυστερήσεις στη συνολική διαδρομή.

Ο πίνακας 5.2.1 αναφέρεται ενδεικτικά σε 24 χαρακτηριστικά στοιχεία (από το σύνολο των 32246) όπως αυτά εισήχθησαν στο SPSS και αφορούν τους χρόνους άφιξης στη συνολική γραμμή 3. Και ο πίνακας 5.2.2 αναφέρεται σε 24 χαρακτηριστικά στοιχεία όπως αυτά εισήχθησαν στο SPSS και αφορούν τις καθυστερήσεις της συνολικής γραμμής 3.

Πίνακας 5.2.1 Απόσπασμα αρχείου εισόδου χρόνων συνολικής διαδρομής στο SPSS

TIME	WEATHER	DAY	PERIOD	DIST_BEGIN	DIST_BETW	INTERSECT	SPEED
1,51	0	4	1	631,07	631,07	0	19,72
1,56	0	4	1	1203,53	572,46	3	15,90
1,34	0	4	1	1515,85	312,32	0	14,20
1,85	0	4	1	2100,50	584,65	2	8,35
1,38	0	4	1	2418,98	318,48	1	13,27
1,90	0	4	1	2916,87	497,90	3	6,22
1,60	0	4	1	3184,92	268,05	4	6,70
1,64	0	4	1	3518,95	334,02	4	7,59
1,72	0	4	1	3850,01	331,06	1	6,37
1,51	0	4	1	4322,09	472,07	4	14,75
1,51	0	4	1	4592,76	270,68	5	8,46
1,34	0	4	1	4855,15	262,38	4	11,93
1,38	0	4	1	5131,98	276,83	6	11,53
1,78	0	4	1	5380,68	248,70	3	4,15
1,60	0	4	1	5837,98	457,29	4	11,43
1,88	0	4	1	6235,02	397,05	6	5,22
1,83	0	4	1	6522,80	287,77	8	4,23
1,68	0	4	1	6847,63	324,84	6	6,77
1,92	0	4	1	7286,12	438,49	12	5,22
1,60	0	4	1	7521,19	235,07	13	5,88
1,90	0	4	1	7883,95	362,76	9	4,53
1,66	0	4	1	8237,06	353,11	5	7,68
1,75	0	4	1	8581,09	344,03	4	6,14
1,62	0	4	1	8995,68	414,59	6	9,87

Πίνακας 5.2.2 Απόσπασμα αρχείου εισόδου καθυστερήσεων συνολικής διαδρομής στο SPSS

DELAY	WEATHER	DAY	PERIOD	DIST_BEGIN	DIST_BETW	INTERSECT	SPEED
1	0	4	1	631,07	631,07	0	19,72
0	0	4	1	1203,53	572,46	3	15,90
0	0	4	1	1515,85	312,32	0	14,20
1	0	4	1	2100,50	584,65	2	8,35
0	0	4	1	2418,98	318,48	1	13,27
0	0	4	1	2916,87	497,90	3	6,22
0	0	4	1	3184,92	268,05	4	6,70
0	0	4	1	3518,95	334,02	4	7,59
0	0	4	1	3850,01	331,06	1	6,37
0	0	4	1	4322,09	472,07	4	14,75
0	0	4	1	4592,76	270,68	5	8,46
0	0	4	1	4855,15	262,38	4	11,93
0	0	4	1	5131,98	276,83	6	11,53
0	0	4	1	5380,68	248,70	3	4,15
0	0	4	1	5837,98	457,29	4	11,43
0	0	4	1	6235,02	397,05	6	5,22
0	0	4	1	6522,80	287,77	8	4,23
0	0	4	1	6847,63	324,84	6	6,77
0	0	4	1	7286,12	438,49	12	5,22
0	0	4	1	7521,19	235,07	13	5,88
1	0	4	1	7883,95	362,76	9	4,53
0	0	4	1	8237,06	353,11	5	7,68
0	0	4	1	8581,09	344,03	4	6,14
0	0	4	1	8995,68	414,59	6	9,87

Αφού εισάγουμε κατάλληλα τα δεδομένα στο πρόγραμμα για ανάλυση, ένα επόμενο βήμα είναι η δημιουργία κάποιων επιπλέον στοιχείων. Παράχθηκαν από τις ήδη υπάρχουσες μεταβλητές **ποιοτικές ψευδομεταβλητές (Dummies)** και **ποσοτικές μεταβλητές**.

Ποιοτικές Μεταβλητές (Dummies)

Οι ποιοτικές μεταβλητές εκφράζουν την επιρροή συγκεκριμένων γεγονότων που δεν είναι μετρήσιμα αριθμητικά στην εξαρτημένη μας μεταβλητή. Οι ποιοτικές μεταβλητές μπορούν να λάβουν μόνο τις τιμές 0 ή 1 (0 όταν δεν ικανοποιείται η συνθήκη ή το γεγονός και 1 όταν ικανοποιείται).

Εδώ πρέπει να τονίσουμε ότι όσον αφορά τις ποιοτικές μεταβλητές τύπου Dummies, σε καμία περίπτωση δεν επιτρέπεται να συμπεριλάβουμε στο ίδιο μοντέλο και το n αριθμό τους. Μπορούμε να συμπεριλάβουμε μέχρι $n-1$ dummy

μεταβλητές, αφού ο σταθερός όρος (intercept) θεωρείται ότι διαδραματίζει το ρόλο της νιοστής μεταβλητής. Προφανώς, η επιλογή της μεταβλητής που δεν θα συμπεριληφθεί στο μοντέλο θα γίνει με κάποια κριτήρια, όπως τη στατιστική σημαντικότητα.

Στην παρούσα διπλωματική εργασία η χρήση ποιοτικών μεταβλητών είναι περιορισμένη λόγω της φύσης των μεταβλητών που έχουν εξετασθεί. Οι μετρήσεις που έχουν πραγματοποιηθεί αφορούν κυρίως ποσοτικά στοιχεία, παρόλα αυτά υπάρχουν και μερικές ποιοτικές μεταβλητές από τις οποίες με τη βοήθεια του SPSS δημιουργήθηκαν διάφορες dummy μεταβλητές.

Στον παρακάτω πίνακα (Πίνακας 5.2.3) παρατίθενται συνοπτικά οι ανεξάρτητες και εξαρτημένες ποιοτικές μεταβλητές όλων των μαθηματικών προτύπων που αναπτύχθηκαν με τις ονομασίες εισαγωγής τους στο λογισμικό στατιστικής ανάλυσης SPSS και το εύρος τιμών που λαμβάνει η κάθε μία.

Πίνακας 5.2.3 Ποιοτικές Μεταβλητές μαθηματικών προτύπων και Μεταβλητές Dummies

DELAY (καθυστέρηση)	1 = καθυστέρηση	0 = κανονική ή νωρίτερη άφιξη		
WEATHER (καιρός)	1 = ηλιοφάνεια	0 = βροχή		
DAY (ημέρα)	1 = Δευτέρα	2 = Τρίτη	3 = Τετάρτη	4 = Πέμπτη
MONDAY	1 = Δευτέρα	0 = όλες οι άλλες ημέρες		
TUESDAY	1 = Τρίτη	0 = όλες οι άλλες ημέρες		
WEDNESDAY	1 = Τετάρτη	0 = όλες οι άλλες ημέρες		
THURSDAY	1 = Πέμπτη	0 = όλες οι άλλες ημέρες		
PERIOD (Περίοδος)	1 = 1η Περίοδος	2 = 2η Περίοδος	3 = 3η Περίοδος	4 = 4η Περίοδος
PERIOD 1	1 = 1η Περίοδος	0 = όλες οι άλλες περίοδοι		
PERIOD 2	1 = 2η Περίοδος	0 = όλες οι άλλες περίοδοι		
PERIOD 3	1 = 3η Περίοδος	0 = όλες οι άλλες περίοδοι		
PERIOD 4	1 = 4η Περίοδος	0 = όλες οι άλλες περίοδοι		

Ποσοτικές Μεταβλητές

Οι ποσοτικές μεταβλητές είναι συνεχείς τιμές οι οποίες προέκυψαν ύστερα από συνδυασμό ή διαχωρισμό σε υποομάδες ήδη υπάρχουσών ανεξάρτητων μεταβλητών. Αυτές οι μεταβλητές δημιουργήθηκαν για τα πρότυπα του κέντρου στα οποία περιλαμβάνονται οι ταχύτητες και οι φόρτοι από φωρατές. Ο λόγος της δημιουργίας τους ήταν η διεξοδικότερη μελέτη αυτών των μεταβλητών και της επιρροής τους στα μοντέλα του κέντρου.

Αρχικά από τον πολλαπλασιασμό των φόρτων με τις ταχύτητες δημιουργήθηκε μια νέα μεταβλητή.

Στη συνέχεια οι φόρτοι χωρίστηκαν σε τρεις υποκατηγορίες ανάλογα με το μέγεθος τους σε: χαμηλούς, μέσους και υψηλούς. Οι ελάχιστοι φόρτοι που παρατηρούνται είναι 138 οχήμ./ώρα, ενώ οι μέγιστοι 2160. Ως χαμηλοί θεωρήθηκαν οι φόρτοι μικρότεροι των 500 οχημ./ώρα, ως μέσοι εκείνοι με τιμές από 500 έως 1500 οχήμ./ώρα και ως υψηλοί αυτοί με τιμές μεγαλύτερες από 1500 οχήμ./ώρα.

Τέλος οι ταχύτητες χωρίστηκαν επίσης σε τρεις υποκατηγορίες ανάλογα με το μέγεθός τους. Χαμηλές ορίσθηκαν εκείνες με τιμές μικρότερες από 25 χλμ/ώρα, μέσες με τιμές μεταξύ 25 και 35 χλμ/ώρα και υψηλές με τιμές μεγαλύτερες από 35 χλμ/ώρα.

Και στις δύο παραπάνω περιπτώσεις αντίστοιχα όταν ικανοποιείται η συνθήκη η μεταβλητή διατηρεί την τιμή της, αλλιώς παίρνει την τιμή 0. Στον παρακάτω πίνακα 5.2.4 δίνεται μια συνοπτική παρουσίαση των δεδομένων.

Πίνακας 5.2.4 Δημιουργία νέων μεταβλητών

VOLUME*SPEED	Τιμή μεταβλητής	
VOLUME_L	<500: Τιμή μεταβλητής	Υπόλοιπες μεταβλητές: 0
VOLUME_M	500< x <1500: Τιμή μεταβλητής	Υπόλοιπες μεταβλητές: 0
VOLUME_H	>1500: Τιμή μεταβλητής	Υπόλοιπες μεταβλητές: 0
SPEED_L	<25: Τιμή μεταβλητής	Υπόλοιπες μεταβλητές: 0
SPEED_M	25< x <35: Τιμή μεταβλητής	Υπόλοιπες μεταβλητές: 0
SPEED_H	>35: Τιμή μεταβλητής	Υπόλοιπες μεταβλητές: 0

5.3 Κατηγοριοποίηση και Συσχέτιση Μεταβλητών

Μετά την προσθήκη των επιπλέον ανεξάρτητων μεταβλητών (dummies) που διαμορφώθηκαν με τη βοήθεια του λογισμικού SPSS, πρώτο βήμα ήταν ο διαχωρισμός των μεταβλητών σε συνεχείς (scale), ονομαστικές (nominal) και διατεταγμένες (ordinal). Αυτό πραγματοποιείται στο αρχείο ελέγχου του λογισμικού SPSS. Ως συνεχείς ορίζονται οι μεταβλητές που δύναται να πάρουν οποιαδήποτε τιμή μέσα σε ένα συγκεκριμένο διάστημα. Ως ονομαστικές ορίζονται αυτές στις οποίες η απάντηση είναι ανάμεσα σε επιλογές που δεν είναι διατεταγμένες (π.χ. περίοδος, ημέρα, καιρός). Τέλος, ως διατεταγμένες ορίζονται αυτές που έχουν μια φυσική σημασία ή μπορούν να διαταχθούν σε μία τακτική κλίμακα (δεν υπάρχουν τέτοιες μεταβλητές στο δείγμα).

Επόμενο βήμα είναι ο έλεγχος του βαθμού συσχέτισης των ανεξάρτητων μεταβλητών μεταξύ τους. Οι ανεξάρτητες μεταβλητές που χρησιμοποιούνται για το κάθε πρότυπο και το κάθε μοντέλο προτύπου πρέπει να μην σχετίζονται έτσι ώστε να πραγματοποιηθεί με επιτυχία η μέθοδος της Παλινδρόμησης. Διαφορετικά η διαδικασία θα οδηγηθεί σε σφάλματα με αποτελέσματα που δεν αντιπροσωπεύουν την πραγματικότητα. Η συσχέτιση μεταξύ των ανεξάρτητων μεταβλητών πραγματοποιήθηκε με τη βοήθεια του λογισμικού και περιγράφεται αναλυτικά για το κάθε πρότυπο παρακάτω.

Η συσχέτιση μεταξύ των ανεξάρτητων μεταβλητών υπολογίστηκε με τη βοήθεια του συντελεστή συσχέτισης Pearson για τη Γραμμική Παλινδρόμηση και το συντελεστή Spearman για τη Λογιστική Παλινδρόμηση. Έχει γίνει εκτενής αναφορά στο Κεφάλαιο 4. Όπως έχει αναφερθεί παραπάνω για τιμές του $r > 0.60$ (υψηλή συσχέτιση) μια από τις δύο ανεξάρτητες μεταβλητές δεν συμπεριλαμβάνεται στο μοντέλο παλινδρόμησης.

Στο **ΠΑΡΑΡΤΗΜΑ** παρουσιάζονται τα αποτελέσματα της συσχέτισης μεταξύ των ανεξάρτητων μεταβλητών που χρησιμοποιήθηκαν στα διάφορα πρότυπα. Διαπιστώνεται ότι μεταξύ των μεταβλητών που χρησιμοποιήθηκαν στο κάθε μοντέλο δεν υπάρχει στατιστικά σημαντική συσχέτιση. Συνεπώς είναι δυνατόν να χρησιμοποιηθούν ταυτόχρονα στην επεξεργασία οι μεταβλητές αυτές ως ανεξάρτητες.

5.4 Στατιστική Ανάλυση Δεδομένων

Στη συνέχεια, ακολουθείται η στατιστική ανάλυση των δεδομένων. Στα πλαίσια της παρούσας διπλωματικής εργασίας σκοπός της διερεύνησης είναι:

- a) Ο χρόνος άφιξης του λεωφορείου στη στάση και κατά πόσο αυτός επηρεάζεται από κάποιους παράγοντες της γραμμής, που μελετάται μέσω της Λογαριθμοκανονικής Παλινδρόμησης.
- b) Η καθυστέρηση στη κάθε στάση σε σχέση με το αναμενόμενο και η πρόβλεψη της πιθανότητας να εμφανιστεί ή όχι καθυστέρηση, που μελετάται μέσω της Λογιστικής Παλινδρόμησης.

Καθώς, δεν είναι γνωστό εκ των προτέρων ποιες από τις παραμέτρους θα απαρτίσουν τα τελικά μαθηματικά πρότυπα, πραγματοποιείται μια σειρά δοκιμών και ελέγχων έως ότου προκύψουν τα τελικά πρότυπα.

Για την αξιολόγηση των μαθηματικών προτύπων στη μέθοδο της **Λογαριθμοκανονικής Παλινδρόμησης** (a) ο δείκτης που ελέγχουμε με το πρόγραμμα SPSS είναι ο **R^2 του Pearson** ο οποίος παίρνει τιμές από 0 έως 1, όσο πιο κοντά είναι στη μονάδα τόσο καλύτερα εκφράζει τη κατάσταση του προτύπου. Ο δείκτης αυτός στην ουσία δείχνει το ποσοστό της μεταβλητότητας του δείκτη που εκφράζεται από το μοντέλο. Επίσης ελέγχεται για κάθε ανεξάρτητη μεταβλητή η τιμή του **t-test** (της κατανομής Student), μέσω του οποίου ελέγχεται το κατά πόσο είναι στατιστικά σημαντική μια μεταβλητή, με αποδεκτή τιμή $t_i > 1,645$ για επίπεδο σημαντικότητας 10%. Τέλος εξετάζεται ο συντελεστής των **Durbin-Watson** που οι αποδεκτές τιμές του είναι μεταξύ $1,7 < DW < 2,2$ και εκφράζει την αυτοσυσχέτιση μεταξύ διαδοχικών σφαλμάτων.

Στη συνέχεια, για την αξιολόγηση των προτύπων στη μέθοδο της **Λογιστικής Παλινδρόμησης** (b) οι συντελεστές που χρησιμοποιεί το λογισμικό του SPSS είναι ο R^2 των Cox & Snell και ο Nagelkerke R^2 . Θεωρητικά και οι δύο συντελεστές λαμβάνουν τιμές από 0 μέχρι 1, επειδή όμως στην πράξη ο R^2 δεν φτάνει ποτέ στην ανώτερη τιμή του, ο εναλλακτικός Nagelkerke R^2 χρησιμοποιείται ως καταλληλότερο μέτρο για την αξιολόγηση της συμπεριφοράς του εξεταζόμενου κάθε φορά προτύπου. Για τον καθορισμό των μεταβλητών που είναι στατιστικά σημαντικές χρησιμοποιείται ο δείκτης **Wald** ο οποίος είναι ανάλογος του **t-test**. Μάλιστα η σχέση που τους συνδέει είναι η εξής: $t \cong \sqrt{W}$. Η οριακή τιμή του Wald*, όπως έχει ήδη αναφερθεί, για τα διάφορα επίπεδα σημαντικότητας α είναι η τιμή πέρα από την οποία απορρίπτεται η μηδενική υπόθεση H_0 . Η μηδενική υπόθεση H_0 είναι μια αρχική υπόθεση που γίνεται ότι η ανεξάρτητη μεταβλητή δεν επηρεάζει την εξαρτημένη.

Στο τέλος της στατιστικής ανάλυσης πραγματοποιείται ο έλεγχος των τελικών, κάθε φορά, προτύπων για να διαπιστωθεί αν η τάση που χαρακτηρίζει τα αποτελέσματα είναι λογική, ανεξάρτητα από το αν οι στατιστικοί δείκτες έχουν ικανοποιητικές τιμές. Ύστερα από τη πραγματοποίηση πολλών δοκιμών, επιλέγονται τα πρότυπα, τα οποία θα παρουσιαστούν παρακάτω, και ικανοποιούν όλους τους παραπάνω ελέγχους στον καλύτερο δυνατό βαθμό.

5.5 Ανάπτυξη Μαθηματικών Προτύπων

5.5.1 Πρότυπα σχετικά με τους Χρόνους Άφιξης

Ανάλυση συνολικού δείγματος

Στην ανάλυση αυτή διερευνάται ποιες ανεξάρτητες μεταβλητές επηρεάζουν την εξαρτημένη μεταβλητή. Η εξαρτημένη μεταβλητή ορίζεται ως ο χρόνος άφιξης του λεωφορείου στη στάση. Το πρότυπο αυτό περιλαμβάνει το συνολικό δείγμα της παρούσας έρευνας. Το δείγμα που επεξεργάζεται το συγκεκριμένο πρότυπο αποτελείται 32246 μετρήσεις.

Στον πίνακα 5.5.1. παρουσιάζεται ο δείκτης R^2 που φανερώνει την καταλληλότητα του προτύπου, δηλαδή κατά πόσο οι τιμές που έχουν προκύψει από τις μετρήσεις μπορούν να εκφραστούν από τις υπολογιζόμενες από το πρότυπο τιμές. Στο σημείο αυτό να αναφέρουμε ότι παρότι το επιθυμητό είναι ο δείκτης R^2 να πλησιάζει στη μονάδα, όταν πρόκειται για πραγματικές μετρήσεις μικρότερες τιμές του είναι αποδεκτές. Στη συγκεκριμένη περίπτωση βέβαια το R^2 είναι υψηλό με τιμή 0,838. Από τη σχέση αυτή διαφαίνεται πως η συσχέτιση των παραπάνω μεταβλητών αφορά στο 84% των χρόνων. Επιπλέον φαίνεται ο δείκτης Durbin-Watson με τιμή 1,899 που υποδεικνύει ότι δεν υπάρχει πρόβλημα αυτοσυσχέτισης μεταξύ των διαδοχικών σφαλμάτων.

Πίνακας 5.5.1 Δείκτες R² και DW

Model	R	R Square	Durbin-Watson
1	,915 ^a	,838	1,899

Πίνακας 5.5.2 Συντελεστές B και t

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	1,818	0,002		789,495	0,00
DIST_BEGIN	-7,15E-06	0,00	-0,113	-50,308	0,00
DIST_BETW	0,001	0,00	0,632	259,811	0,00
INTERSECT	0,004	0,00	0,051	21,274	0,00
SPEED	-0,047	0,00	-0,85	-335,644	0,00
WEDNESDAY	-0,002	0,001	-0,005	-2,13	0,033
PERIOD1	0,007	0,001	0,015	4,811	0,00
PERIOD2	0,028	0,002	0,05	17,115	0,00
PERIOD3	0,014	0,002	0,024	8,364	0,00
WEATHER	-0,005	0,001	-0,01	-4,639	0,00

Μόρφωση μαθηματικής σχέσης:

$$\text{Logy} = 1,818 - 7,14E-06 \cdot \text{DISTBEG} + ,001 \cdot \text{DISTBETW} + ,004 \cdot \text{INTER} - ,047 \cdot \text{SPEED} - ,002 \cdot \text{WED} + ,07 \cdot \text{PER1} + ,028 \cdot \text{PER2} + ,014 \cdot \text{PER3} - ,005 \cdot \text{WEATH}$$

Μετά την απολογαρίθμηση:

$$y = 10^{(1,818 - 7,14E-06 \cdot \text{DISTBEG} + ,001 \cdot \text{DISTBETW} + ,004 \cdot \text{INTER} - ,047 \cdot \text{SPEED} - ,002 \cdot \text{WED} + ,07 \cdot \text{PER1} + ,028 \cdot \text{PER2} + ,014 \cdot \text{PER3} - ,005 \cdot \text{WEATH})}$$

Σχολιασμός παραπάνω μαθηματικού προτύπου:

Το παραπάνω μοντέλο επιλέχθηκε ανάμεσα σε μια σειρά μοντέλων μετά από πολλές δοκιμές, παρακάτω εξηγούνται τα πρόσημα και η σημασία της κάθε μεταβλητής.

- Το αρνητικό πρόσημο μπροστά από τη μεταβλητή **DIST_BEGIN** υποδηλώνει ότι όσο η απόσταση από την αφετηρία αυξάνεται, τόσο μειώνονται οι χρόνοι. Αυτό μπορεί να ερμηνευθεί ως εξής, είτε κατά τη μετάβαση είτε κατά την επιστροφή προς το τέλος της διαδρομής το λεωφορείο βρίσκεται εκτός του κεντρικού τομέα της πόλης, οι ταχύτητες αυξάνονται και κατ' επέκταση οι

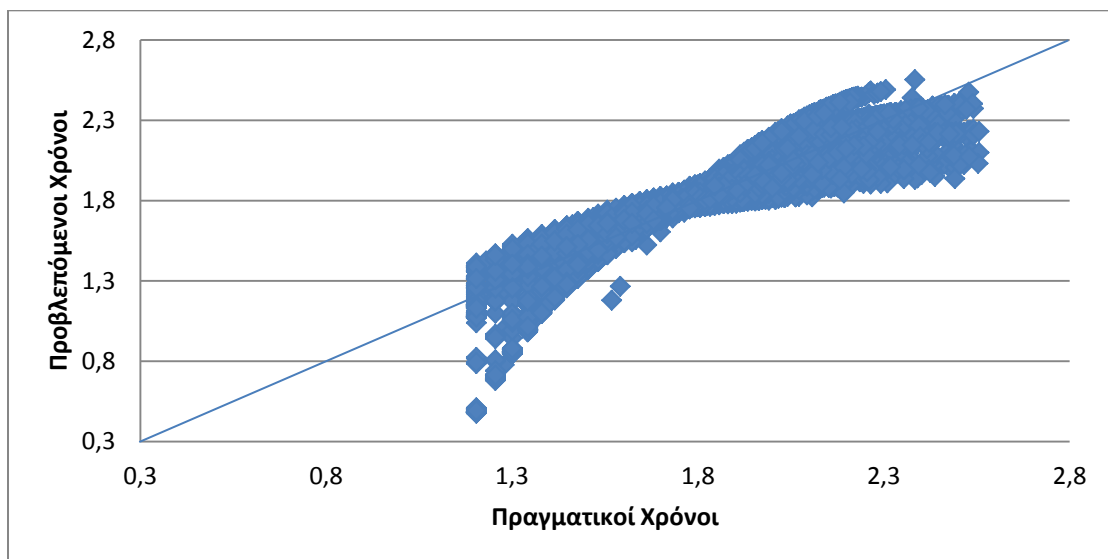
- χρόνοι μειώνονται. Η μείωση των χρόνων, βέβαια, είναι αμελητέα καθώς είναι της τάξεως του 10^{-6} .
- Το θετικό πρόσημο μπροστά από τη μεταβλητή **DIST_BETW** δείχνει κάτι αναμενόμενο, ότι όσο η απόσταση μεταξύ των στάσεων αυξάνεται, τόσο αυξάνονται και οι χρόνοι από στάση σε στάση. Αν αναλογιστεί κανείς ότι η μέσες ταχύτητες των αστικών λεωφορείων διατηρούνται σε σταθερά γενικά επίπεδα, η αύξηση των αποστάσεων οδηγεί σε αύξηση των χρόνων.
 - Το θετικό πρόσημο μπροστά από τη μεταβλητή **INTERSECT** σημαίνει ότι όσο ο αριθμός των διασταυρώσεων που παρεμβάλλονται μεταξύ των στάσεων αυξάνεται τόσο αυξάνονται και οι χρόνοι άφιξης του λεωφορείου στην κάθε στάση. Το αποτέλεσμα είναι αναμενόμενο, καθώς ένας κόμβος, σηματοδοτούμενος ή μη, προκαλεί καθυστερήσεις άρα και αύξηση των χρόνων.
 - Το αρνητικό πρόσημο μπροστά από τη μεταβλητή **SPEED** υποδηλώνει ότι τυχόν αύξηση της ταχύτητας προκαλεί μείωση των χρόνων και κατ'αντιστοιχία πιθανή μείωση ταχύτητας σε αύξηση των χρόνων άφιξης του λεωφορείου στη στάση. Το αποτέλεσμα αυτό, επίσης, προκύπτει λογικά και επιβεβαιώνεται από το μαθηματικό πρότυπο.
 - Το αρνητικό πρόσημο μπροστά από τη μεταβλητή **WEDNESDAY** δείχνει ότι οι χρόνοι αντιμετωπίζουν μια μείωση κατά την ημέρα της εβδομάδας Τετάρτη. Κάτι τέτοιο πιθανόν να οφείλεται στο γεγονός ότι κατά την Τετάρτη το ωράριο των καταστημάτων είναι έως τις μεσημεριανές ώρες και πιθανόν η συμφόρηση του οδικού δικτύου είναι μειωμένη σε σχέση τις υπόλοιπες ημέρες.
 - Το θετικό πρόσημο μπροστά από τη μεταβλητή **PERIOD1** δείχνει ότι οι χρόνοι άφιξης του λεωφορείου στην κάθε στάση επηρεάζονται από την περίοδο και μάλιστα θετικά. Ομοίως θετικά είναι τα πρόσημα μπροστά από τις μεταβλητές **PERIOD2** και **PERIOD3**, που υποδηλώνει ότι η περίοδος της ημέρας οδηγεί σε αύξηση των χρόνων διαδρομής. Όπως φαίνεται από τους συντελεστές β_i την μεγαλύτερη θετική επιρροή στο μοντέλο έχει η δεύτερη περίοδος, στη συνέχεια η τρίτη περίοδος και τέλος η πρώτη. Το γεγονός αυτό υποδηλώνει ότι οι χρόνοι αυξάνονται περισσότερο κατά τη δεύτερη περίοδο, κάτι που πιθανόν οφείλεται στο γεγονός ότι η δεύτερη περίοδος αποτελεί τη μεσημεριανή αιχμή.
 - Το αρνητικό πρόσημο μπροστά από τη μεταβλητή **WEATHER** δείχνει ότι οι χρόνοι μειώνονται όσο η τιμή της μεταβλητής του καιρού αυξάνεται, δηλαδή από την τιμή 0 στην τιμή 1. Η τιμή 0 αφορά τις ημέρες με βροχή, ενώ η τιμή 1

εκείνες με ηλιοφάνεια. Συνεπώς το αρνητικό πρόσημο αντιπροσωπεύει την πραγματικότητα, καθώς σε ημέρες με βροχή η συμφόρηση και οι καθυστερήσεις αυξάνονται με αντίκτυπο την αύξηση των προβλεπόμενων χρόνων ανάμεσα στις στάσεις.

- Τέλος η τιμή της **σταθεράς (constant)** προέκυψε θετική 1,818 και είναι η τιμή της Y για $x=0$.

Από όλες της μεταβλητές του μαθηματικού προτύπου αυτή που αυξάνει περισσότερο τους χρόνους άφιξης είναι η PERIOD2 με τιμή 0,028, ενώ αυτή που συντελεί περισσότερο στη μείωση των χρόνων είναι η SPEED με τιμή -0,047.

Παρακάτω παρατίθεται το διάγραμμα διασποράς (scatter plot) μεταξύ των πραγματικών χρόνων και των προβλεπόμενων από το πρόγραμμα. Και οι δύο τιμές είναι σε μορφή λογάριθμου. Όπως φαίνεται οι περισσότερες τιμές είναι συγκεντρωμένες κοντά στη ευθεία $Y=X$, ενώ λίγες απέχουν σημαντικά από την ευθεία.



Διάγραμμα 5.5.1 Διάγραμμα διασποράς Πραγματικοί –Προβλεπόμενοι Χρόνοι

Στη συνέχεια πραγματοποιείται ανάπτυξη προτύπων με την ίδια εξαρτημένη μεταβλητή (χρόνοι άφιξης λεωφορείου στη στάση) με τη διαφορά ότι τώρα η μελέτη αφορά το κομμάτι της διαδρομής που είναι εντός του κέντρου.

Ανάλυση δείγματος εντός κέντρου

Στην ανάλυση αυτή διερευνάται ποιες ανεξάρτητες μεταβλητές επηρεάζουν την εξαρτημένη μεταβλητή, το χρόνο άφιξης του λεωφορείου στη στάση. Το πρότυπο αυτό περιλαμβάνει μόνο τη διαδρομή εντός του κέντρου και το μέγεθος του δείγματος είναι 7099 μετρήσεις.

Παρακάτω παρουσιάζονται όπως προηγουμένως ο δείκτης R^2 που δείχνει την καταλληλότητα του προτύπου. Στη συγκεκριμένη περίπτωση το R^2 είναι χαμηλότερο με τιμή 0,314, το οποίο μπορεί να γίνει αποδεκτό. Επιπλέον φαίνεται ο δείκτης Durbin-Watson με τιμή 1,753.

Πίνακας 5.5.3 Δείκτες R^2 και DW

Model	R	R Square	Durbin-Watson
1	,560 ^a	0,314	1,753

Πίνακας 5.5.4 Συντελεστές B και t

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	1,899	0,016		117,462	0
MONDAY	-0,015	0,007	-0,025	-2,319	0,02
TUESDAY	-0,014	0,006	-0,024	-2,175	0,03
WEDNESDAY	-0,023	0,005	-0,049	-4,479	0
PERIOD1	0,012	0,007	0,028	1,841	0,066
PERIOD2	0,088	0,009	0,171	9,99	0
PERIOD3	0,049	0,008	0,095	6,479	0
VOLUME_M	2,31E-05	0	0,058	5,337	0
DIST_BEG	-1,60E-05	0	-0,317	-24,577	0
INTERSECT	0,039	0,001	0,386	38,213	0
SPEED	-0,002	0	-0,075	-4,537	0
SPEED_L	-0,001	0	-0,033	-2,615	0,009
VOLSPEED	-2,96E-07	0	-0,032	-1,892	0,059

Μόρφωση μαθηματικής σχέσης:

$$\begin{aligned} \text{Logy} = & 1,899 - ,0015 \cdot \text{MON} - ,014 \cdot \text{TUES} - ,023 \cdot \text{WED} + ,012 \cdot \text{PER1} + ,088 \cdot \text{PER2} + ,049 \cdot \text{PER3} \\ & + 2,31E - 05 \cdot \text{VOLM} - 1,6E - 05 \cdot \text{DISTBEG} + ,039 \cdot \text{INTER} - ,002 \cdot \text{SPEED} - ,001 \\ & \cdot \text{SPEEDL} - 2,96E - 07 \cdot \text{VOLSP} \end{aligned}$$

Μετά την απολογαρίθμηση:

$$y = 10^{(1,899 - ,0015 \cdot \text{MON} - ,014 \cdot \text{TUES} - ,023 \cdot \text{WED} + ,012 \cdot \text{PER1} + ,088 \cdot \text{PER2} + ,049 \cdot \text{PER3} + 2,31E - 05 \cdot \text{VOLM} - 1,6E - 05 \cdot \text{DISTBEG} + ,039 \cdot \text{INTER} - ,002 \cdot \text{SPEED} - ,001 \cdot \text{SPEEDL} - 2,96E - 07 \cdot \text{VOLSP})}$$

Σχολιασμός παραπάνω μαθηματικού προτύπου:

Το παραπάνω μοντέλο επιλέχθηκε ανάμεσα σε μια σειρά μοντέλων μετά από πολλές δοκιμές, παρακάτω επεξηγούνται τα πρόσημα και η σημασία της κάθε μεταβλητής.

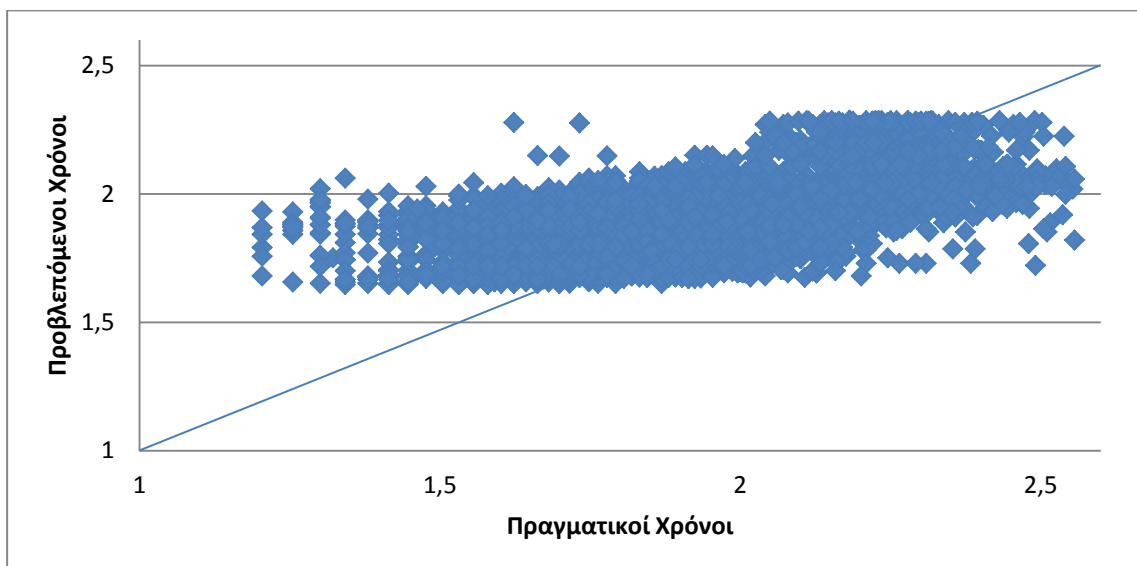
- Το αρνητικό πρόσημο μπροστά από τη μεταβλητή **MONDAY** δείχνει ότι οι χρόνοι υπόκεινται μείωση κατά την ημέρα αυτή. Επιπλέον αρνητικές είναι και οι τιμές των μεταβλητών **TUESDAY** και **WEDNESDAY**. Ειδικότερα κατά την Τετάρτη προκαλείται η μεγαλύτερη μείωση των χρόνων με τιμή της μεταβλητής 0,023, ακολουθεί η Δευτέρα με 0,015 και τέλος η Τρίτη με τιμή 0,014.
- Το θετικό πρόσημο μπροστά από τη μεταβλητή **PERIOD1** δείχνει ότι οι χρόνοι κατά την πρώτη περίοδο αυξάνονται. Το ίδιο ισχύει και για τη δεύτερη και τρίτη περίοδο. Όπως φαίνεται από τον πίνακα η μεταβλητή **PERIOD2** προκαλεί την μεγαλύτερη αύξηση των χρόνων, κάτι που πιθανόν οφείλεται στη μεσημεριανή αιχμή. Στη συνέχεια, έρχεται η μεταβλητή **PERIOD3** και τέλος η **PERIOD1**.
- Το θετικό πρόσημο μπροστά από τη μεταβλητή **VOLUME_M** υποδηλώνει ότι οι χρόνοι αυξάνονται όσο οι μέσοι φόρτοι του κυκλοφοριακού δικτύου αυξάνονται. Το αποτέλεσμα αυτό ήταν αναμενόμενο, καθώς η αύξηση των φόρτων συνεπάγεται μεγαλύτερη συμφόρηση στο κέντρο και κατ' επέκταση μεγαλύτερους χρόνους διαδρομής από τη μια στάση στην επόμενη.
- Το θετικό πρόσημο μπροστά από τη μεταβλητή **DIST_BEG** δείχνει ότι για το κέντρο όσο αυξάνονται οι αποστάσεις από την αρχή, οι χρόνοι παρουσιάζουν αύξηση. Το πρόσημο αυτής της μεταβλητής διαφοροποιείται στο πρότυπο του κέντρου σε σχέση με αυτό της συνολικής διαδρομής. Αυτό συμβαίνει γιατί στο κέντρο που οι απαιτούμενοι χρόνοι, λόγω της κυκλοφοριακής συμφόρησης, είναι μεγαλύτεροι οι καθυστερήσεις από στάση σε στάση συσσωρεύονται με αποτέλεσμα προς τις τελευταίες στάσεις οι χρόνοι να αυξάνονται αρκετά.
- Το θετικό πρόσημο μπροστά από τη μεταβλητή **INTERSECT** σημαίνει ότι όσο ο αριθμός των διασταυρώσεων που παρεμβάλλονται μεταξύ των στάσεων αυξάνεται τόσο αυξάνονται και οι χρόνοι άφιξης του λεωφορείου στην κάθε στάση. Ένας κόμβος, σηματοδοτούμενος ή μη, προκαλεί καθυστερήσεις άρα και αύξηση των χρόνων. Όπως φαίνεται από την τιμή του B αν μεταξύ δυο στάσεων υπάρχουν 10 διασταυρώσεις ο χρόνος αυξάνεται κατά 0,39 μονάδες.
- Το αρνητικό πρόσημο μπροστά από τη μεταβλητή **SPEED** υποδηλώνει ότι τυχόν αύξηση της ταχύτητας προκαλεί μείωση των χρόνων λεωφορείου στη στάση. Το αποτέλεσμα αυτό, επίσης, προκύπτει λογικά και επιβεβαιώνεται από το μαθηματικό πρότυπο. Μια πιθανή αύξηση της ταχύτητας κατά 10χλμ/ ώρα

προκαλεί μείωση των χρόνων κατά 0,02 μονάδες. Αντίστοιχη επιρροή στους χρόνους ασκεί και η μεταβλητή **SPEED_L**, με συντελεστή 0,001.

- Το αρνητικό πρόσημο της μεταβλητής **VOLSPEED**, που αποτελεί το γινόμενο φόρτου-ταχύτητας, υποδηλώνει ότι μια πιθανή αύξηση της προκαλεί μείωση των χρόνων. Από αυτό το αποτέλεσμα μπορεί κανείς να συμπεράνει ότι η επιρροή των φόρτων πάνω στους χρόνους είναι μικρότερη από αυτή της ταχύτητας.
- Τέλος η τιμή **της σταθεράς (constant)** προέκυψε θετική 1,899 και είναι η τιμή της Y για $x=0$.

Από το σύνολο των μεταβλητών που έχουν εισαχθεί στο μοντέλο τη μεγαλύτερη θετική επιρροή έχει η **PERIOD2** με τιμή 0,088, ενώ τη μεγαλύτερη μείωση των χρόνων προκαλεί η μεταβλητή **WEDNESDAY** με τιμή 0,023.

Παρακάτω παρατίθεται το διάγραμμα διασποράς (scatter plot) μεταξύ των πραγματικών χρόνων και των προβλεπόμενων από το πρόγραμμα για το κέντρο. Και οι δύο τιμές είναι σε μορφή λογάριθμου. Όπως φαίνεται αρκετές είναι οι τιμές που απέχουν από την ευθεία $Y=X$, κάτι που αναμένονταν, καθώς το R^2 στην περίπτωση του κέντρου δεν ήταν τόσο υψηλό. Παρ' όλα αυτά είναι αποδεκτό.



Διάγραμμα 5.5.2 Διάγραμμα διασποράς Πραγματικοί-Προβλεπόμενοι Χρόνοι

Συγκεντρωτικός πίνακας αποτελεσμάτων:

Παρακάτω παρατίθεται ένας συνολικός πίνακας που απεικονίζει τα αποτελέσματα από τα δυο παραπάνω μοντέλα σχετικά με τους χρόνους άφιξης στην ολική διαδρομή και στο κέντρο.

Πίνακας 5.5.5 Συνολικός πίνακας αποτελεσμάτων για όλα τα μοντέλα των Χρόνων

	Συνολική διαδρομή		Διαδρομή εντός κέντρου	
	B	t	B	t
(constant)	1,818	789,495	1,899	117,462
WEATHER (καιρός)	-,005	-4,639		
DAY				
PERIOD				
DIST_BEG	-7,147E-06	-50,308	-1,60E-05	-24,577
DIST_BETW	,001	259,811		
INTERSECT	,004	21,274	0,039	38,213
VOLUME				
VOLUME_L				
VOLUME_M			2,31E-05	5,337
VOLUME_H				
SPEED	-,047	-335,644	-0,002	-4,537
SPEED_L			-0,001	-2,615
SPEED_M				
SPEED_H				
VOL*SPEED			-2,96E-07	-1,892
MONDAY			-0,015	-2,319
TUESDAY			-0,014	-2,175
WEDNESDAY	-,002	-2,130	-0,023	-4,479
THURSDAY				
PERIOD 1	,007	4,811	0,012	1,841
PERIOD 2	,028	17,115	0,088	9,99
PERIOD 3	,014	8,364	0,049	6,479
PERIOD 4				
R Square	0,838		0,314	
Durbin - Watson	1,899		1,753	

5.5.2 Πρότυπα σχετικά με τις καθυστερήσεις

Ανάλυση συνολικού δείγματος

Στην ανάλυση αυτή διερευνάται ποιες ανεξάρτητες μεταβλητές επηρεάζουν την εξαρτημένη μεταβλητή. Η εξαρτημένη μεταβλητή ορίζεται ως η πιθανότητα το λεωφορείο να καθυστερήσει σε σχέση με τον προγραμματισμένο χρόνο. Η ανάλυση του προτύπου εδώ γίνεται με τη μέθοδο της Λογιστικής Παλινδρόμησης. Αρχικά μελετάται το πρότυπο του συνολικού δείγματος, το οποίο αφορά στην ολική διαδρομή της γραμμής 3 και έχει μέγεθος 32246 μετρήσεις.

Στον πίνακα 5.5.5 παρουσιάζεται ο δείκτης R^2 Nagelkerke που φανερώνει την καταλληλότητα του προτύπου, δηλαδή κατά πόσο οι υπολογιζόμενες από το πρότυπο τιμές με αυτές που έχουν προκύψει από την έρευνα του πεδίου μπορούν να συσχετιστούν. Στο σημείο αυτό να αναφέρουμε ότι στη λογιστική παλινδρόμηση οι τιμές που λαμβάνει ο δείκτης R^2 είναι σημαντικά μικρότερες από αυτές που λαμβάνει στη γραμμική παλινδρόμηση. Όπως βλέπουμε και στο πίνακα, η τιμή του δείκτη R^2 είναι 0,279. Η τιμή αυτή δεν είναι η καλύτερη για ένα πρότυπο λογιστικής παλινδρόμησης, αλλά μπορεί να γίνει αποδεκτή.

Πίνακας 5.5.6 Δείκτες R^2 Nagelkerke

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	36973,753 ^a	0,209	0,279

Στον πίνακα 5.5.6 εμφανίζονται τα ποσοστά επιτυχίας του προτύπου για την καθυστέρηση ή μη στην κάθε στάση. Το ποσοστό σωστής πρόβλεψης του μοντέλου είναι 72,1%, το οποίο είναι αρκετά ικανοποιητικό (>50%).

Πίνακας 5.5.7 Ποσοστά πρόβλεψης προτύπου

Παρατηρούμενα		Προβλεπόμενα		
		DELAY		Ποσοστό επιτυχίας
		,0000	1,0000	
Step 1	DELAY ,0000	11934	5345	69,1
	DELAY 1,0000	3649	11318	75,6
	Συνολικό ποσοστό			72,1

Για το καθορισμό των μεταβλητών που είναι στατιστικά αποδεκτές λαμβάνεται τιμή του επιπέδου σημαντικότητας $\alpha=0,10$. Ελάχιστη αποδεκτή τιμή για

τον δείκτη t είναι η $Wald^* \cong 2,706$ λόγω της σχέσης $t = \sqrt{w}$. Στον παρακάτω πίνακα 5.5.8 παρουσιάζονται τις τιμές του Wald για τις επιλεγμένες μεταβλητές που περιγράφουν το συγκεκριμένο πρότυπο. Όπως προκύπτει, λοιπόν, από τον πίνακα το Wald των μεταβλητών είναι μεγαλύτερο της κρίσιμης τιμής W^* . Αυτό υποδεικνύει ότι οι συγκεκριμένες ανεξάρτητες μεταβλητές είναι στατιστικά σημαντικές. Όσο μεγαλύτερη η τιμή του Wald τόσο μεγαλύτερη είναι και η επιρροή της συγκεκριμένης μεταβλητής στο τελικό αποτέλεσμα. Από τον παρακάτω πίνακα φαίνεται ότι η μεταβλητή ταχύτητα έχει τη μεγαλύτερη τιμή Wald και συνεπώς έχει τη μεγαλύτερη επιρροή στο τελικό αποτέλεσμα.

Πίνακας 5.5.8 Τιμές συντελεστή B και δείκτη Wald

Μεταβλητές της εξίσωσης			
	B	Wald	Sig.
DIST_BEGIN	0	71,876	0
DIST_BETW	0,003	1065,896	0
MONDAY	-0,068	3,242	0,072
WEDNESDAY	0,082	8,469	0,004
PERIOD1	-0,083	5,343	0,021
PERIOD2	-0,396	100,725	0
PERIOD3	-0,284	52,221	0
SPEED	-0,435	4123,569	0
Constant	1,186	510,027	0

Μόρφωση μαθηματικής σχέσης:

$$\log \left[\frac{P1}{1 - P1} \right] = 1,186 + ,000 \cdot DISTBEG + ,003 \cdot DISTBETW - ,068 \cdot MON + ,082 \cdot WED - ,083 \cdot PER1 - ,396 \cdot PER2 - ,284 \cdot PER3 - ,435 \cdot SPEED$$

Σχολιασμός παραπάνω μαθηματικού προτύπου:

Το $P1$ είναι η πιθανότητα καθυστέρησης της άφιξης του λεωφορείου στη στάση και το $1 - P1$ είναι το λεωφορείο να φθάσει εντός του προγραμματισμένου χρόνου ή νωρίτερα.

- Το θετικό πρόσημο μπροστά από τη μεταβλητή **DIST_BEGIN** σημαίνει ότι η πιθανότητα να εμφανιστούν καθυστερήσεις αυξάνεται όσο η απόσταση από την αρχή μεγαλώνει. Αυτό μπορεί να ερμηνευθεί ως εξής, κατά μήκος της διαδρομής λόγω διαφόρων κυκλοφοριακών συνθηκών δημιουργούνται

καθυστερήσεις οι οποίες συσσωρεύονται στη διάρκεια του δρομολογίου. Η τιμή του συντελεστή B δεν είναι 0, είναι απλά πολύ μικρή.

- Το θετικό πρόσημο μπροστά από τη μεταβλητή **DIST_BETW** σημαίνει ότι οι αποστάσεις μεταξύ των στάσεων έχουν θετική επιρροή πάνω στις καθυστερήσεις. Δηλαδή για στάσεις που απέχουν πολύ μεταξύ τους αυξάνεται την πιθανότητα εμφάνισης καθυστερήσεων. Αυτό είναι λογικό καθώς όσο αυξάνεται η απόσταση μεταξύ των στάσεων τόσο περισσότερες μεταβολές στις κυκλοφοριακές συνθήκες μπορούν να συμβούν, που μπορεί να οδηγήσουν σε αύξηση των χρόνων και των καθυστερήσεων.
- Το αρνητικό πρόσημο μπροστά από τη μεταβλητή **MONDAY** σημαίνει ότι κατά τη διάρκεια της Δευτέρας μειώνεται η πιθανότητα καθυστερήσεων. Αυτό μπορεί να εξηγηθεί από το γεγονός ότι τη Δευτέρα που τα καταστήματα είναι κλειστά κατά τις απογευματινές ώρες, οι φόρτοι είναι πιθανό να μειώνονται και κατ' επέκταση μειώνονται οι καθυστερήσεις. Αντίθετα, το θετικό πρόσημο μπροστά από τη μεταβλητή **WEDNESDAY** υποδηλώνει ότι κατά την Τετάρτη αυξάνεται η πιθανότητα εμφάνισης καθυστερήσεων.
- Το αρνητικό πρόσημο μπροστά από τη μεταβλητή **PERIOD1** δείχνει ότι κατά την πρώτη περίοδο η πιθανότητα εμφάνισης καθυστερήσεων μεταξύ των στάσεων μειώνεται. Το ίδιο ισχύει και για τις μεταβλητές **PERIOD2** και **PERIOD3**, οι οποίες έχουν αρνητικό πρόσημο. Στο σημείο αυτό παρατηρείται ότι ο συντελεστής της μεταβλητής PERIOD2 είναι ο μεγαλύτερος από τους τρεις, στη συνέχεια έρχεται αυτός της PERIOD3 και τέλος της PERIOD1. Το γεγονός αυτό υποδηλώνει ότι είναι λιγότερο πιθανό να εμφανιστούν καθυστερήσεις κατά τη δεύτερη περίοδο.
- Το αρνητικό πρόσημο μπροστά από τη μεταβλητή **SPEED** δείχνει ότι η αύξηση των ταχυτήτων οδηγεί σε μείωση της πιθανότητας εμφάνισης καθυστερήσεων. Αυτό το αποτέλεσμα είναι λογικό, καθώς αύξηση της ταχύτητας σημαίνει μείωση των χρόνων και κατ' επέκταση των καθυστερήσεων.
- Το θετικό πρόσημο της **σταθεράς** υποδηλώνει ότι γενικά κατά τη διάρκεια μιας διαδρομής είναι πιο πιθανή η εμφάνιση καθυστερήσεων, σε σχέση με τους αναμενόμενους χρόνους, μεταξύ των στάσεων.

Από όλες της μεταβλητές του μαθηματικού προτύπου αυτή που αυξάνει περισσότερο την πιθανότητα εμφάνισης καθυστερήσεων είναι η WEDNESDAY, δηλαδή κατά τη διάρκεια της Τετάρτης με τιμή 0,082. Και αυτή που αυξάνει

περισσότερο την πιθανότητα το λεωφορείο να φθάσει εντός του προγραμματισμένου χρόνου είναι η μεταβλητή SPEED με απόλυτη τιμή 0,435.

Στη συνέχεια πραγματοποιείται ανάπτυξη προτύπων με την ίδια εξαρτημένη μεταβλητή (καθυστέρηση ή άφιξη εντός του αναμενόμενου) με τη διαφορά ότι τώρα πρόκειται να μελετηθεί η διαδρομή εντός του κέντρου της πόλης.

Ανάλυση δείγματος εντός κέντρου

Στην ανάλυση αυτή μελετώνται οι ανεξάρτητες μεταβλητές που έχουν επιρροή πάνω στην εξαρτημένη, στην καθυστέρηση της άφιξης λεωφορείου στη στάση. Συγκεκριμένα μελετάται η διαδρομή εντός του κέντρου με μέγεθος δείγματος 7099 μετρήσεις.

Στον πίνακα 5.5.9 παρουσιάζεται ο δείκτης R^2 Nagelkerke που φανερώνει την καταλληλότητα του προτύπου, δηλαδή κατά πόσο οι υπολογιζόμενες από το πρότυπο τιμές με αυτές που έχουν προκύψει από την έρευνα του πεδίου μπορούν να συσχετιστούν. Όπως βλέπουμε και στο πίνακα, η τιμή που παίρνει ο δείκτης R^2 είναι ,061. Η τιμή αυτή δεν είναι ικανοποιητική για ένα πρότυπο λογιστικής παλινδρόμησης, αλλά είναι η καλύτερη δυνατή για το συγκεκριμένο πρότυπο.

Πίνακας 5.5.9 Δείκτης R^2 Nagelkerke

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	19755,6	0,042	0,061

Στον πίνακα 5.5.10 παρουσιάζονται τα ποσοστά επιτυχίας του προτύπου για την καθυστέρηση ή μη στην κάθε στάση. Το συνολικό ποσοστό σωστής πρόβλεψης του προτύπου ανέρχεται στο 55%. Το ποσοστό αυτό είναι ικανοποιητικό διότι είναι μεγαλύτερο από το 50%.

Πίνακας 5.5.10 Ποσοστά πρόβλεψης προτύπου

Παρατηρούμενα		Προβλεπόμενα		
		DELAY		Ποσοστό επιτυχίας
		0	1	
0		3513	349	91
DELAY 1		2846	391	12,1
Step 1	Συνολικό ποσοστό			55

Στον παρακάτω πίνακα 5.5.11 παρουσιάζονται τις τιμές του Wald για τις επιλεγμένες μεταβλητές που περιγράφουν το συγκεκριμένο πρότυπο. Όπως

προκύπτει, λοιπόν, από τον πίνακα το Wald των μεταβλητών είναι μεγαλύτερο της κρίσιμης τιμής W^* . Αυτό υποδεικνύει ότι οι συγκεκριμένες ανεξάρτητες μεταβλητές είναι στατιστικά σημαντικές. Από τον παρακάτω πίνακα φαίνεται ότι η μεταβλητή των διασταυρώσεων (INTERSECT) έχει τη μεγαλύτερη τιμή Wald και συνεπώς έχει τη μεγαλύτερη επιρροή στο τελικό αποτέλεσμα.

Πίνακας 5.5.11 Τιμές συντελεστή B και δείκτη Wald

Μεταβλητές της εξίσωσης			
	B	Wald	Sig.
DIST_BEG	0,000	3,618	0,057
INTERSECT	0,037	10,204	0,001
MONDAY	0,271	13,29	0,000
TUESDAY	0,207	8,451	0,004
WEDNESDAY	0,171	8,846	0,003
VOLSPEED	0,000	2,579	0,108
Constant	-0,479	40,287	0,000

Μόρφωση μαθηματικής σχέσης:

$$\log \left[\frac{P1}{1 - P1} \right] = -0,479 + ,000 \cdot DISTBEG + ,0037 \cdot INTER + 0,271 \cdot MON + 0,207 \cdot TUES \\ + ,0171 \cdot WED + ,000 \cdot VOLSPE$$

Σχολιασμός παραπάνω μαθηματικού προτύπου:

Το παραπάνω μοντέλο επιλέχθηκε ανάμεσα σε μια σειρά μοντέλων μετά από πολλές δοκιμές, παρακάτω επεξηγούνται τα πρόσημα και η σημασία της κάθε μεταβλητής.

- Το θετικό πρόσημο μπροστά από τη μεταβλητή **DIST_BEG** σημαίνει ότι η πιθανότητα να υπάρξει καθυστέρηση σε μια στάση αυξάνεται όσο αυξάνεται η απόσταση από την αφετηρία. Αυτό μπορεί να ερμηνευθεί ως εξής, κατά μήκος της διαδρομής δημιουργούνται καθυστερήσεις οι οποίες συσσωρεύονται στη διάρκεια του δρομολογίου. Η τιμή του συντελεστή B είναι πολύ μικρή, κοντά στο 0.

- Το θετικό πρόσημο μπροστά από τη μεταβλητή **INTERSECT** υποδηλώνει ότι αυξάνεται η πιθανότητα εμφάνισης καθυστερήσεων, όταν ο αριθμός των διασταυρώσεων που παρεμβάλλονται μεταξύ των στάσεων μεγαλώνει. Η τιμή της μεταβλητής είναι 0,037.
- Τα θετικά πρόσημα μπροστά από τις μεταβλητές **MONDAY**, **TUESDAY** και **WEDNESDAY** δείχνουν ότι κατά τη διάρκεια και τριών ημερών αυξάνεται η πιθανότητα εμφάνισης καθυστερήσεων. Μεγαλύτερη επιρροή στο πρότυπο ασκεί η μεταβλητή MONDAY με τιμή 0,271, ακολουθεί η TUESDAY με 0,207 και τέλος η μεταβλητή WEDNESDAY με 0,171.
- Το θετικό πρόσημο μπροστά από τη μεταβλητή **VOLSPEED** δείχνει ότι πιθανή αύξηση του γινομένου του φόρτου με την ταχύτητα οδηγεί σε αύξηση της πιθανότητας εμφάνισης καθυστερήσεων. Το αποτέλεσμα αυτό δείχνει ότι η επιρροή των φόρτων στις καθυστερήσεις είναι μεγαλύτερη από αυτή των ταχυτήτων στο συγκεκριμένο πρότυπο.
- Τέλος η σταθερά (Constant) είναι η μόνη τιμή η οποία έχει αρνητικό πρόσημο. Αυτό υποδηλώνει ότι γενικά κατά τη διάρκεια μιας διαδρομής είναι πιο πιθανό να μην υπάρξουν καθυστερήσεις, αλλά να φτάσει εντός του προγραμματισμένου χρόνου το λεωφορείο.

Από όλες τις μεταβλητές του μαθηματικού προτύπου αυτή που αυξάνει περισσότερο την πιθανότητα εμφάνισης καθυστερήσεων είναι η μεταβλητή της ημέρας MONDAY, καθώς έχει το μεγαλύτερο συντελεστή 0,271. Παρατηρείται επίσης ότι όλες οι μεταβλητές, εκτός της σταθεράς, έχουν θετικό πρόσημο, άρα όλες αυξάνουν την πιθανότητα εμφάνισης καθυστερήσεων.

Συγκεντρωτικός πίνακας αποτελεσμάτων:

Παρακάτω παρατίθεται ένας συνολικός πίνακας που απεικονίζει τα αποτελέσματα από τα δυο παραπάνω μοντέλα σχετικά με τους χρόνους άφιξης στην ολική διαδρομή και στο κέντρο.

Πίνακας 5.5.12 Συνολικός πίνακας αποτελεσμάτων για όλα τα μοντέλα των καθυστερήσεων

	Συνολική διαδρομή		Διαδρομή εντός κέντρου	
	B	t	B	t
(constant)	1,186	510,027	-0,479	40,287
WEATHER (καιρός)				
DAY				
PERIOD				
DIST_BEG	0,000	71,876	0,000	3,618
DIST_BETW	0,003	1065,896		
INTERSECT			0,037	10,204
VOLUME				
VOLUME_L				
VOLUME_M				
VOLUME_H				
SPEED	-0,435	4123,569		
SPEED_L				
SPEED_M				
SPEED_H				
VOL*SPEED			0,000	2,579
MONDAY	-0,068	3,242	0,271	13,29
TUESDAY			0,207	8,451
WEDNESDAY	0,082	8,469	0,171	8,846
THURSDAY				
PERIOD 1	-0,083	5,343		
PERIOD 2	-0,396	100,725		
PERIOD 3	-0,284	52,221		
PERIOD 4				
R Square	0,279		0,061	

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

6.1 Εισαγωγή

Στην παρούσα διπλωματική εργασία γίνεται μια προσπάθεια καταγραφής των παραγόντων που επηρεάζουν τους χρόνους άφιξης και τις καθυστερήσεις από στάση σε στάση για την Λεωφορειακή Γραμμή 3 στην πόλη της Θεσσαλονίκης. Τη βάση δεδομένων αποτέλεσαν στοιχεία μετρήσεων επτά συνολικά ημερών. Η στατιστική επεξεργασία, της οποίας η περιγραφή έγινε στο κεφάλαιο 5, πραγματοποιήθηκε με το πρόγραμμα SPSS και πιο συγκεκριμένα με τις μεθόδους της Λογαριθμοκανονικής Παλινδρόμησης και της Λογιστικής Παλινδρόμησης. Σκοπός της επεξεργασίας ήταν να διερευνηθούν και να προτυποποιηθούν ο χρόνος άφιξης και οι καθυστερήσεις στις στάσεις, καθώς και να αξιολογηθεί ο βαθμός επιρροής των παραμέτρων που βρέθηκε ότι τα επηρεάζουν. Στο παρόν κεφάλαιο παρουσιάζονται αφενός τα αποτελέσματα της ανάλυσης των δεδομένων για τον προσδιορισμό των μεταβλητών που ασκούν επιρροή στο χρόνο διαδρομής και την καθυστέρηση και αφετέρου παρατίθενται γενικά συμπεράσματα και προτάσεις για την περαιτέρω έρευνα.

6.2 Σύνοψη Αποτελεσμάτων

Αποτελέσματα επεξεργασίας βάσης δεδομένων

Από την επεξεργασία της βάσης δεδομένων προέκυψαν πολλά χρήσιμα συμπεράσματα για τα χαρακτηριστικά του δείγματος. Αρχικά να αναφερθεί ότι τη βάση δεδομένων αποτελούν το ολικό δείγμα που αφορά τη συνολική διαδρομή και έχει μέγεθος 32246 γραμμές στοιχείων και δείγμα που αφορά τη διαδρομή εντός του κέντρου με μέγεθος 7099 γραμμές στοιχείων. Από τις γραφικές απεικονίσεις των χρόνων του συνολικού δείγματος συμπεράναμε ότι κατά τη μετάβαση του δρομολογίου οι χρόνοι εμφανίζουν μια αυξητική τάση, ενώ αντίθετα κατά την επιστροφή εμφανίζουν μια πτωτική τάση με περισσότερες διακυμάνσεις.

Από την μελέτη των χρόνων άφιξης σε σχέση με τις καιρικές συνθήκες παρατηρείται ότι εμφανίζουν, σε μετάβαση και επιστροφή, αυξημένες τιμές κατά τη διάρκεια ημερών με βροχόπτωση παρά εκείνων με ηλιοφάνεια. Κάτι τέτοιο αναμένονταν καθώς η κυκλοφοριακή συμφόρηση αυξάνεται τις ημέρες με βροχή και κατ' επέκταση οι χρόνοι αυξάνονται. Μάλιστα το ποσοστό των στάσεων με τιμές

χρόνων μεγαλύτερες στη βροχόπτωση φτάνει το 70% και στις δύο περιπτώσεις. Επιπλέον οι τιμές των χρόνων στη μετάβαση διατηρούνται γενικά σε σταθερά επίπεδα, ενώ οι τιμές κατά την επιστροφή είναι σε υψηλότερα επίπεδα με αρκετές διακυμάνσεις.

Η επεξεργασία των χρόνων με βάση την περίοδο της ημέρας επίσης έδωσε χρήσιμα αποτελέσματα. Το κυριότερο συμπέρασμα είναι ότι οι χρόνοι τόσο κατά τη μετάβαση, όσο και κατά την επιστροφή είναι ιδιαίτερα αυξημένοι στη δεύτερη περίοδο. Η δεύτερη περίοδος συμπίπτει με την μεσημεριανή αιχμή, καθώς είναι από τις 12:00 - 15:00, άρα το αποτέλεσμα αυτό ήταν αναμενόμενο. Συγκεκριμένα, σε ποσοστό μεγαλύτερο από 60% των στάσεων στη μετάβαση οι μέγιστοι χρόνοι εμφανίστηκαν κατά τη διάρκεια της δεύτερης περιόδου, ενώ στην επιστροφή το ποσοστό αυτό ξεπέρασε το 80%, εμφανίζοντας παράλληλα περισσότερες διακυμάνσεις στους χρόνους.

Από την κατασκευή του ραβδογράμματος συχνοτήτων των χρόνων συμπεραίνουμε ότι το δείγμα παρουσιάζει θετική ασυμμετρία εφόσον οι χρόνοι είναι συσσωρευμένοι στις χαμηλές τιμές και όσον αφορά την κύρτωση παρατηρείται λεπτόκυρτη κατανομή. Επίσης, όπως υπολογίστηκε από το πρόγραμμα Bestfit, η κατανομή που ακολουθούν οι χρόνοι προσομοιάζεται καλύτερα με την Λογαριθμοκανονική Κατατομή.

Στη συνέχεια, από την επεξεργασία των στοιχείων της βάσης που αφορούν στο κέντρο προέκυψαν χρήσιμα συμπεράσματα για τις μεταβλητές του φόρτου και της ταχύτητας. Αρχικά, από το διάγραμμα διασποράς των φόρτων με τις ταχύτητες παρατηρεί κανείς ότι η μεταξύ τους σχέση είναι σε γενικές γραμμές η αναμενόμενη, με τις ταχύτητες να μειώνονται με την αύξηση των φόρτων. Βέβαια παρατηρείται ότι για φόρτους μικρότερους από 700 οχήμ./ώρα οι ταχύτητες διατηρούνται σε σχετικά σταθερά επίπεδα, ενώ το εύρος τους αυξάνεται για τιμές των φόρτων μεγαλύτερες των 700 οχήμ./ώρα.

Από την μελέτη της σχέσης μεταξύ χρόνων και ταχυτήτων εντός του κέντρου φαίνεται ότι σε γενικές γραμμές οι χρόνοι είναι συσσωρευμένοι μεταξύ 30-200 δευτερολέπτων. Παρατηρείται ότι οι χρόνοι εμφανίζουν μια αύξηση για τιμές της ταχύτητας 20-35 χλμ/ώρα, ενώ για μεγαλύτερες τιμές της οι χρόνοι μειώνονται ξανά, όπως ήταν αναμενόμενο, στα 30-100 δευτερόλεπτα. Σε γενικές γραμμές όμως παρατηρεί κανείς ότι οι μεταβολές των χρόνων δεν είναι μεγάλες.

Από τη γραφική απεικόνιση των φόρτων σε σχέση με τον καιρό εντός του κέντρου φαίνεται ότι οι φόρτοι είναι εμφανώς αυξημένοι τις ημέρες με ηλιοφάνεια, αυτό υποδηλώνει ότι η συμφόρηση στο κέντρο μειώνεται και η ροή των οχημάτων είναι μεγαλύτερη. Επιπλέον κατά τη μετάβαση οι φόρτοι είναι αυξημένοι σε σχέση με την επιστροφή, αυτό συμβαίνει γιατί στη μετάβαση το λεωφορείο διασχίζει τη

Τσιμισκή που αποτελεί μια από τις κεντρικότερες οδούς της πόλης. Κάτι παρόμοιο φαίνεται να συμβαίνει και με τις ταχύτητες, όπου τις ημέρες με ηλιοφάνεια είναι μεγαλύτερες από εκείνες με βροχή. Κατ' αντιστοιχία με προηγούμενως τις ημέρες με βροχή οι χρόνοι μεταξύ των στάσεων μειώνονται και οι μέσες ταχύτητες πέφτουν. Στην περίπτωση των ταχυτήτων οι μέσες τιμές ανά στάση δεν διαφέρουν σε μεγάλο βαθμό κατά τη μετάβαση και την επιστροφή. Οι τιμές των ταχυτήτων στη μετάβαση παρουσιάζουν μια διακύμανση, ενώ κατά την επιστροφή διατηρούνται σε σταθερά γενικά επίπεδα από εκείνες με βροχή.

Από τη γραφική απεικόνιση των φόρτων φαίνεται ότι κατά την τέταρτη περίοδο εμφανίζονται οι χαμηλότεροι φόρτοι στο σύνολο της διαδρομής εντός του κέντρου. Οι τιμές τους κατά την επιστροφή είναι σημαντικά χαμηλότερες από αυτών της μετάβασης. Επιπρόσθετα, στο 50% των στάσεων οι μέγιστοι φόρτοι συναντώνται στη δεύτερη περίοδο, κάτι που υποδηλώνει ότι η δεύτερη περίοδος αποτελεί την αιχμή της κυκλοφοριακής συμφόρησης.

Από το διάγραμμα των ταχυτήτων ανά περίοδο εντός κέντρου συμπεραίνουμε ότι οι μεγαλύτερες ταχύτητες εμφανίζονται κατά την τέταρτη περίοδο, συγκεκριμένα στο 94% των περιπτώσεων. Αντίθετα οι ελάχιστες ταχύτητες παρατηρούνται κατά τη δεύτερη περίοδο. Αυτό το αποτέλεσμα ήταν αναμενόμενο, καθώς κατά την τέταρτη περίοδο ελαχιστοποιούνται οι φόρτοι και κατά συνέπεια οι τιμές των ταχυτήτων αυξάνονται ιδιαίτερα. Γενικότερα παρατηρείται μια έντονη μεταβλητότητα των ταχυτήτων κατά τη μετάβαση της διαδρομής, σε αντίθεση με τις τιμές που εμφανίζονται κατά την επιστροφή.

Αποτελέσματα επεξεργασίας μαθηματικών μοντέλων

Στο τελικό στάδιο της παρούσας διπλωματικής εργασίας πραγματοποιήθηκε στατιστική επεξεργασία της βάσης δεδομένων και ανάπτυξη μαθηματικών προτύπων για τους παράγοντες που επηρεάζουν τους χρόνους άφιξης και για την πρόβλεψη των καθυστερήσεων μεταξύ των στάσεων.

Παρακάτω παρατίθενται τα αποτελέσματα των μαθηματικών προτύπων σχετικά με τους χρόνους άφιξης του λεωφορείου στη στάση, για τη συνολική διαδρομή και για τη διαδρομή εντός του κέντρου.

1. Τα αποτελέσματα των χρόνων άφιξης των λεωφορείων σε κάθε στάση για τη συνολική διαδρομή οδηγούν στα εξής συμπεράσματα:

Οι μεταβλητές που επηρεάζουν περισσότερο την αύξηση των χρόνων είναι εκείνες της δεύτερης και της τρίτης περιόδου, ενώ η πρώτη περίοδος ασκεί μικρότερη θετική επιρροή. Δηλαδή κατά τη διάρκεια της δεύτερης και της τρίτης περιόδου οι απαιτούμενοι χρόνοι διαδρομής μεταξύ των στάσεων αυξάνονται ιδιαίτερα. Αυτό μπορεί να ερμηνευτεί από τους ιδιαίτερα αυξημένους φόρτους της

μεσημεριανής αιχμής. Επίσης συμπεραίνουμε ότι θετική αλλά μικρότερη επιρροή στους χρόνους ασκεί η μεταβλητή που αφορά τις αποστάσεις μεταξύ των στάσεων και η μεταβλητή των διασταυρώσεων. Το πρόσημο των αποστάσεων μεταξύ προκύπτει λογικά και επιβεβαιώνεται από το πρότυπο. Παράλληλα το θετικό πρόσημο μπροστά από τις διασταυρώσεις σημαίνει ότι ένας κόμβος, σηματοδοτούμενος ή μη, προκαλεί καθυστερήσεις άρα και αύξηση των απαιτούμενων χρόνων.

Αντίθετα, οι σημαντικότεροι παράγοντες που προκαλούν μείωση των χρόνων στο συγκεκριμένο πρότυπο είναι ο καιρός και η ταχύτητα. Συμπεραίνουμε, λοιπόν, ότι η μεταβολή του καιρού από βροχή σε ηλιοφάνεια συντελεί αρκετά στη μείωση των απαιτούμενων χρόνων, ενώ παράλληλα μια πιθανή αύξηση της ταχύτητας του αστικού λεωφορείου οδηγεί στη μείωση των χρόνων. Τα δυο παραπάνω συμπεράσματα προκύπτουν λογικά και επιβεβαιώνονται από το πρότυπο. Ακόμη, μείωση των χρόνων αλλά με μικρότερο βάρος παρατηρείται κατά τη διάρκεια της Τετάρτης όπου το ωράριο των καταστημάτων είναι έως τις μεσημεριανές ώρες και η συμφόρηση του οδικού δικτύου είναι μειωμένη σε σχέση τις υπόλοιπες ημέρες. Καθώς επίσης και με την αύξηση των αποστάσεων από την αρχή, κάτι που οφείλεται στο γεγονός ότι σε απομακρυσμένες από την αφετηρία στάσεις το λεωφορείο βρίσκεται εκτός του κεντρικού τομέα της πόλης, οι ταχύτητες αυξάνονται και κατ' επέκταση οι χρόνοι μειώνονται.

2. Τα αποτελέσματα των χρόνων άφιξης του λεωφορείου σε κάθε στάση για τη διαδρομή εντός του κέντρου οδηγούν στα εξής συμπεράσματα:

Οι μεταβλητές που επηρεάζουν περισσότερο την αύξηση των χρόνων σε αυτό το πρότυπο είναι της δεύτερης και της τρίτης περιόδου, κάτι που μπορεί να οφείλεται στους ιδιαίτερα αυξημένους φόρτους της μεσημεριανής αιχμής. Επίσης η μεταβλητή των διασταυρώσεων επιδρά σημαντικά στην αύξηση των χρόνων, κάτι που οδηγεί στο συμπέρασμα ότι εντός του κέντρου που οι φόρτοι είναι μεγαλύτεροι η ύπαρξη διασταυρώσεων επιδεινώνει το πρόβλημα των υψηλών χρόνων άφιξης. Επιπλέον, από το πρότυπο που προέκυψε συμπεραίνουμε ότι η μεταβλητή της πρώτης περιόδου, των μέσων φόρτων και των αποστάσεων από την αρχή συντελούν στην αύξηση των χρόνων απλά σε μικρότερο βαθμό. Το αποτέλεσμα για τους φόρτους ήταν αναμενόμενο, καθώς η αύξηση των φόρτων συνεπάγεται μεγαλύτερη συμφόρηση στο κέντρο και κατ' επέκταση μεγαλύτερους χρόνους διαδρομής από τη μια στάση στην επόμενη. Αντίθετα, σε ότι αφορά τις αποστάσεις από την αρχή το αποτέλεσμα διαφοροποιείται σε σχέση με του συνολικού προτύπου, αυτό συμβαίνει γιατί στο κέντρο δημιουργούνται καθυστερήσεις λόγω της κυκλοφοριακής συμφόρησης με αποτέλεσμα προς τις τελευταίες στάσεις του κέντρου οι χρόνοι να αυξάνονται αρκετά.

Αντίθετα, τη μεγαλύτερη μείωση των χρόνων προκαλούν οι μεταβλητές που αφορούν τις ημέρες Δευτέρα, Τρίτη και Τετάρτη, κάτι το οποίο δείχνει ότι για τη διαδρομή εντός του κέντρου η ημέρα παίζει σημαντικό ρόλο. Παράλληλα, μείωση

των χρόνων προκαλούν και οι μεταβλητές της ταχύτητας, της χαμηλής ταχύτητας και του γινομένου φόρτου – ταχύτητας, αλλά με μικρότερο συντελεστή. Όσον αφορά την ταχύτητα και τη χαμηλή ταχύτητα το αποτέλεσμα ήταν αυτό που περιμέναμε, ενώ από το πρόσημο του γινομένου φόρτου – ταχύτητας εξάγεται το συμπέρασμα ότι η επιρροή των ταχυτήτων στους χρόνους είναι μεγαλύτερη από αυτή των φόρτων.

Παρακάτω παρατίθενται τα αποτελέσματα των μαθηματικών προτύπων σχετικά με τις καθυστερήσεις του λεωφορείου στη στάση, για τη συνολική διαδρομή και για τη διαδρομή εντός του κέντρου.

3. Τα αποτελέσματα των καθυστερήσεων του λεωφορείου σε κάθε στάση για τη συνολική διαδρομή οδηγούν στα εξής συμπεράσματα:

Η πιθανότητα εμφάνισης καθυστερήσεων στη συνολική διαδρομή είναι μεγαλύτερη κατά τη διάρκεια της Τετάρτης. Επιπλέον, την πιθανότητα εμφάνισης καθυστερήσεων αυξάνουν οι μεταβλητές της απόστασης από την αρχή και της απόστασης μεταξύ των στάσεων. Από τα παραπάνω μπορεί να εξαχθεί το συμπέρασμα ότι κατά μήκος της διαδρομής λόγω διαφόρων κυκλοφοριακών συνθηκών δημιουργούνται καθυστερήσεις οι οποίες συσσωρεύονται στη διάρκεια του δρομολογίου, καθώς και ότι όσο μεγαλώνει η απόσταση μεταξύ των στάσεων, τόσο περισσότερες μεταβολές στις κυκλοφοριακές συνθήκες μπορούν να συμβούν και να οδηγήσουν σε αύξηση των χρόνων και των καθυστερήσεων.

Αντίθετα, την πιθανότητα εμφάνισης καθυστερήσεων μειώνουν σημαντικά οι μεταβλητές της ταχύτητας, της δεύτερης περιόδου και της τρίτης περιόδου. Το αποτέλεσμα αυτό για την ταχύτητα είναι λογικό, εφόσον πιθανή αύξησή της οδηγεί σε μείωση των χρόνων και κατ' επέκταση των καθυστερήσεων. Από την άλλη για τις άλλες δύο μεταβλητές μπορεί να εξαχθεί το συμπέρασμα ότι επειδή οι χρόνοι άφιξης είναι αυξημένοι σε εκείνες τις περιόδους, όπως προέκυψε από την στατιστική ανάλυση, η πιθανότητα το λεωφορείο να υπερβεί τους αναμενόμενους χρόνους μειώνεται. Από το παραπάνω πρότυπο προέκυψε επιπρόσθετα ότι κατά τη Δευτέρα και κατά την πρώτη περίοδο μειώνεται η πιθανότητα εμφάνισης καθυστερήσεων, αλλά με μικρότερη επιρροή. Το γεγονός ότι κατά τη Δευτέρα μειώνεται η πιθανότητα εμφάνισης καθυστερήσεων πιθανόν οφείλεται στους μειωμένους φόρτους την ημέρα αυτή, ενώ αντίστοιχα το πρόσημο της μεταβλητής της πρώτης περιόδου μπορεί να ερμηνευθεί από την ύπαρξη σχετικά υψηλών χρόνων στην πρώτη περίοδο και την μειωμένη πιθανότητα να ξεπερασθούν.

4. Τα αποτελέσματα των καθυστερήσεων του λεωφορείου σε κάθε στάση για τη διαδρομή εντός του κέντρου οδηγούν στα εξής συμπεράσματα:

Στο συγκεκριμένο πρότυπο όλες οι μεταβλητές εκτός της σταθεράς έχουν θετικό πρόσημο, αυτό μας οδηγεί στο συμπέρασμα ότι γενικά οι καθυστερήσεις εντός του κέντρου τείνουν να αυξάνονται. Ειδικότερα, η πιθανότητα εμφάνισης

καθυστερήσεων αυξάνεται σημαντικά κατά τη διάρκεια της Δευτέρας και της Τρίτης στο κέντρο. Επίσης οι υπόλοιπες μεταβλητές που επιδρούν στην αύξηση των καθυστερήσεων, αλλά σε μικρότερο βαθμό, είναι η μεταβλητή της ημέρας Τετάρτης, των διασταυρώσεων, της απόστασης από την αρχή και του γινομένου του φόρτου – ταχύτητας. Από τα παραπάνω συμπεραίνουμε ότι η αύξηση του αριθμού των διασταυρώσεων που παρεμβάλλονται μεταξύ των στάσεων επιδρά στις καθυστερήσεις, αυτό οφείλεται στο γεγονός ότι ένας κόμβος σηματοδοτούμενος ή μη προκαλεί αλλαγές στην κυκλοφορία, αυξήσεις στους χρόνους διαδρομής και αύξηση των καθυστερήσεων. Από το γεγονός ότι η αύξηση των αποστάσεων από την αρχή μεγαλώνει την πιθανότητα εμφάνισης καθυστερήσεων, συμπεραίνουμε ότι κατά μήκος της διαδρομής δημιουργούνται καθυστερήσεις οι οποίες συσσωρεύονται στη διάρκεια του δρομολογίου. Τέλος, από το θετικό πρόσημο της μεταβλητής φόρτου – ταχύτητας εξάγεται το συμπέρασμα ότι στη διαδρομή εντός κέντρου η επιρροή των φόρτων στις καθυστερήσεις είναι μεγαλύτερη από αυτή των ταχυτήτων.

6.3 Προτάσεις για περαιτέρω έρευνα

Στο παρόν τεύχος παρουσιάζεται η έρευνα που έγινε στα πλαίσια της διπλωματικής εργασίας, η οποία είχε στόχο τη μελέτη και προτυποποίηση των χρόνων άφιξης και των καθυστερήσεων σε μια συγκεκριμένη λεωφορειακή γραμμή. Η αξιοπιστία των δρομολογίων αστικών συγκοινωνιών και η ακριβής πρόβλεψη των χρόνων άφιξης των λεωφορείων είναι ένα θέμα που απασχολεί το μεγαλύτερο μέρος του επιβατικού κοινού και χρήζει περαιτέρω μελέτης.

Χρήσιμο θα ήταν να εκπονηθεί μια έρευνα που να αφορά μετρήσεις περισσότερων ημερών και μηνών για να δοθεί μια πιο ολοκληρωμένη εικόνα της λεωφορειακής γραμμής, να υπάρχει η δυνατότητα μελέτης περισσότερων χαρακτηριστικών της χρονοσειράς και το δείγμα να μην επηρεάζεται από τυχαία γεγονότα και μεμονωμένα περιστατικά. Επιπλέον, μια πολύ πιο ολοκληρωμένη προσέγγιση του θέματος θα μπορούσε να πραγματοποιηθεί με μια μελέτη μεγαλύτερης κλίμακας που να αφορά περισσότερα δρομολόγια λεωφορείων τα οποία να καλύπτουν την ευρύτερη περιοχή της Θεσσαλονίκης και τις κεντρικές αρτηρίες της πόλης.

Επιπλέον, οι κυκλοφοριακές συνθήκες που επικρατούν στο σύνολο του δικτύου μπορούν να βοηθήσουν ιδιαίτερα στην διεξαγωγή πιο ολοκληρωμένης έρευνας. Στην πόλη της Θεσσαλονίκης οι φωρατές κυκλοφορίας που είναι εγκατεστημένοι περιορίζονται στο κέντρο της πόλης. Για το λόγο αυτό δημιουργήθηκαν τα δύο ξεχωριστά πρότυπα, της συνολικής διαδρομής και του κέντρου, όπου οι μέσοι φόρτοι και ταχύτητες αφορούσαν τις μετρήσεις του κέντρου. Με τη χρήση κατάλληλων μηχανημάτων μέτρησης των κυκλοφοριακών συνθηκών, δηλαδή με τη χρήση **Intelligent Transportation Systems (ITS)**, στο

σύνολο του δικτύου θα δίνονταν η δυνατότητα διεξαγωγής μιας πιο εμπεριστατωμένης μελέτης.

Μία ακόμα έρευνα που θα μπορούσε να πραγματοποιηθεί είναι η εκτίμηση των χρόνων διαδρομής και των καθυστερήσεων σε συνδυασμό με τους χρόνους μετακίνησης των χρηστών των αστικών συγκοινωνιών και των χρηστών Ι.Χ. Η συλλογή της βάσης δεδομένων για τους χρόνους μετακίνησης των μετακινούμενων πολιτών θα μπορούσε να πραγματοποιηθεί με τη χρήση ερωτηματολογίων και η στατιστική τους επεξεργασία θα μπορούσε να γίνει με τη χρήση κάποιου είδους παλινδρόμησης. Ενώ, η συλλογή της βάσης δεδομένων για τους χρόνους μετακίνησης των Ι.Χ. θα μπορούσε να πραγματοποιηθεί με τη χρήση αυτοματοποιημένων μηχανημάτων καταγραφής κυκλοφοριακών φόρτων και ταχυτήτων, που θα τοποθετούνταν σε διάφορα σημεία σε επιλεγμένες αρτηρίες και η επεξεργασία των στοιχείων αυτών θα μπορούσε να γίνει με χρήση **Intelligent Transportation Systems (ITS)**.

Ταυτόχρονα προτείνεται η διεξαγωγή αντίστοιχων ερευνών και σε άλλα αστικά κέντρα της Ελλάδας (Αθήνα, Πάτρα). Ειδικότερα στην Αθήνα όπου το δίκτυο εξυπηρετεί πολύ μεγαλύτερο επιβατικό κοινό και είναι αρκετά πολύπλοκο μια τέτοια έρευνα θα είχε ιδιαίτερο ενδιαφέρον. Θα μπορούσε, επιπλέον, να γίνει και μία σύγκριση αποτελεσμάτων μεταξύ των μεγαλουπόλεων.

Τέλος, σκόπιμη θεωρείται η επανάληψη της παρούσας έρευνας με τη χρήση άλλης μεθοδολογίας. Το αντικείμενο των χρόνων άφιξης και της εμφάνισης καθυστερήσεων θα μπορούσε να προσεγγισθεί με μοντέλα Χρονικής Διάρκειας (Duration Models). Τα Duration Models έχουν εφαρμογή στον τομέα των συγκοινωνιών σε μελέτες που αφορούν τη χρονική διάρκεια μετακινήσεων και μπορούν να οδηγήσουν στον εντοπισμό πιθανών σφαλμάτων και παραλείψεων της παρούσας διπλωματικής εργασίας με αποτέλεσμα να διορθωθούν και να οδηγήσουν σε καλύτερα αποτελέσματα.

Βιβλιογραφία

Ξένη Βιβλιογραφία

1. Washington, Simon P., Karlaftis Matthew G., Mannering, Fred L., (2003), "Statistical and Econometrical Methods for Transportation Data Analysis", Chapman & Hall/CRC Press Publishing.
2. Mei Chen, Xiaobo Liu, Jingxin Xia, Steven I. Chien, (2004), "A Dynamic Bus – Arrival Time Prediction Model Based on APC Data", Computer-Aided Civil and Infrastructure Engineering Journal, pp. 364 – 376.
3. Field A., (2005), Discovering Statistics Using S.P.S.S., Logistic Regression, pp.218 - 239
4. Jing-nan Wang, Xu-mei Chen, Shu xia Guo, (2009), "Bus travel time prediction model with v-support vector regression", Beijing JiaoTong University.
5. YU Bo, LU Jing, YU Bin, YANG Zhongzhen, (2009), "An Adaptive Bus Arrival Time Prediction Model", Dalian Maritime University.
6. Bin Yu, William H.K. Lam, Mei Lam Tam, (2011), "Bus arrival time prediction at bus stop with multiple routes", Hong Kong Polytechnic University.
7. Coffey, C., Pozdnoukhov, A., Calabrese, F., (2011), "Time of arrival predictability horizons for public bus routes", National University of Ireland Maynooth.
8. Li, F., Yu, Y., Lin, H., Min, W., (2011), "Public bus arrival time prediction based on traffic information management system", IBM Research.
9. Sun, Y.-Y., Liu, Y., Zhou, X.-Y., Sun, L.-M., (2012), Ruan Jian Xue Bao/Journal of Software, pp. 87 -99.
10. Mazloumi E., Moridpour S., Currie G., Rose G., (2012), "Exploring the value of traffic flow data in bus travel time prediction", Journal of Transportation Engineering, pp. 436-446.
11. Mazloumi E., Rose G., Currie G. & Sarvi M., (2012), "An Integrated Framework to predict Bus Travel Time and its Variability Using Traffic Flow Data", Journal of Intelligent Transportation Systems, pp. 75-90.

12. Sinn, M., Ji Won Yoon, Calabrese, F., Bouillet, E., (2012), “Predicting arrival times of buses using real time GPS measurements”, 15th International IEEE Conference, pp. 1227 – 1232
13. Tongyu Zhu, Jian Dong, Jian Huang, Songsong Pang, (2012), “The bus arrival time service based on dynamic traffic information”, 6th International Conference of AICT, pp. 1 – 6.
14. Feng Li, HongBin Lin, Yu Yuan, Changjie Guo, Wanli Min, XinYu Zhao, (2012), “Bus arrival prediction and trip planning for better user experience and services”, Journal Service science, management and engineering, pp. 189 – 221.

Ελληνική Βιβλιογραφία

15. Αδαμόπουλος Λ., Δαμιανού Χ., Σβέρκος Α., (1999), “Μαθηματικά και Στοιχεία Στατιστικής, Γ΄ Ενιαίου Λυκείου”, ΟΕΔΒ, Αθήνα.
16. Κοκολάκης Γ. και Σπηλιώτης Ι., (1999) “Εισαγωγή στη Θεωρία Πιθανοτήτων και Στατιστικής”, διδακτικό εγχειρίδιο, Αθήνα.
17. Σταθόπουλος Α. και Καρλαύτης Μ., (2008), “Σχεδιασμός Μεταφορικών Συστημάτων”, εκδόσεις Παπασωτηρίου, Αθήνα.

ΠΑΡΑΡΤΗΜΑ

**ΑΠΟΤΕΛΕΣΜΑΤΑ ΣΥΣΧΕΤΙΣΗΣ ΜΕΤΑΞΥ ΤΩΝ ΑΝΕΞΑΡΤΗΤΩΝ ΜΕΤΑΒΛΗΤΩΝ ΠΟΥ
ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΑΝ ΣΤΑ ΔΙΑΦΟΡΑ ΠΡΟΤΥΠΑ**

Συντελεστής συσχέτισης μεταξύ ανεξάρτητων μεταβλητών στο πρότυπο των
χρόνων για τη συνολική διαδρομή.

Correlations

		DIST_BEGIN	DIST_BETW	INTERSECT	SPEED	WEDNESDAY	PERIOD1	PERIOD2	PERIOD3	WEATHER
DIST_BEGIN	Pearson Correlation	1	,029**	,086**	-,013	,000	-,026**	-,003	,011	-,001
	Sig. (2-tailed)		,000	,000	,015	,995	,000	,553	,049	,824
	N	32246	32246	32246	32246	32246	32246	32246	32246	32246
DIST_BETW	Pearson Correlation	,029**	1	,125**	,312**	-,002	-,006	-,001	-,001	,001
	Sig. (2-tailed)	,000		,000	,000	,774	,311	,926	,885	,794
	N	32246	32246	32246	32246	32246	32246	32246	32246	32246
INTERSECT	Pearson Correlation	,086**	,125**	1	-,284**	,000	,011	,004	,000	,004
	Sig. (2-tailed)	,000	,000		,000	,974	,047	,500	,939	,481
	N	32246	32246	32246	32246	32246	32246	32246	32246	32246
SPEED	Pearson Correlation	-,013	,312**	-,284**	1	,007	,038**	-,082**	-,041**	,008
	Sig. (2-tailed)	,015	,000	,000		,196	,000	,000	,000	,132
	N	32246	32246	32246	32246	32246	32246	32246	32246	32246
WEDNESDAY	Pearson Correlation	,000	-,002	,000	,007	1	,016**	-,020**	-,012	-,083**
	Sig. (2-tailed)	,995	,774	,974	,196		,005	,000	,031	,000
	N	32246	32246	32246	32246	32246	32246	32246	32246	32246
PERIOD1	Pearson Correlation	-,026**	-,006	,011	,038**	,016**	1	-,402**	-,403**	-,026**
	Sig. (2-tailed)	,000	,311	,047	,000	,005		,000	,000	,000
	N	32246	32246	32246	32246	32246	32246	32246	32246	32246
PERIOD2	Pearson Correlation	-,003	-,001	,004	-,082**	-,020**	-,402**	1	-,285**	-,009
	Sig. (2-tailed)	,553	,926	,500	,000	,000	,000		,000	,116
	N	32246	32246	32246	32246	32246	32246	32246	32246	32246
PERIOD3	Pearson Correlation	,011	-,001	,000	-,041**	-,012	-,403**	-,285**	1	,007
	Sig. (2-tailed)	,049	,885	,939	,000	,031	,000	,000		,233
	N	32246	32246	32246	32246	32246	32246	32246	32246	32246
WEATHER	Pearson Correlation	-,001	,001	,004	,008	-,083**	-,026**	-,009	,007	1
	Sig. (2-tailed)	,824	,794	,481	,132	,000	,000	,116	,233	
	N	32246	32246	32246	32246	32246	32246	32246	32246	32246

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

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

Συντελεστής συσχέτισης ανεξάρτητων μεταβλητών στο πρότυπο των χρόνων για
τη διαδρομή εντός του κέντρου.

Correlations

		MONDAY	TUESDAY	WEDNESDAY	PERIOD1	PERIOD2	PERIOD3	VOLUME_M	DIST_BEG	INTERSECT	SPEED	SPEED_L	VOLSPEED
MONDAY	Pearson Correlation	1	-,173**	-,264**	-,037**	-,004	-,001	,004	,000	,000	,067**	-,003	,039**
	Sig. (2-tailed)		,000	,000	,002	,712	,939	,736	1,000	,995	,000	,789	,001
	N	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099
TUESDAY	Pearson Correlation	-,173**	1	-,283**	-,022	,013	-,001	-,008	,003	,002	,047**	-,114**	,042**
	Sig. (2-tailed)	,000		,000	,062	,257	,907	,490	,820	,867	,000	,000	,000
	N	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099
WEDNESDAY	Pearson Correlation	-,264**	-,283**	1	,001	-,021	,012	-,037**	,005	,000	,006	-,079**	,010
	Sig. (2-tailed)	,000	,000		,943	,081	,298	,002	,649	,990	,635	,000	,421
	N	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099
PERIOD1	Pearson Correlation	-,037**	-,022	,001	1	-,408**	-,411**	-,016	,026	,002	,136**	-,281**	,126**
	Sig. (2-tailed)	,002	,062	,943		,000	,000	,192	,027	,888	,000	,000	,000
	N	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099
PERIOD2	Pearson Correlation	-,004	,013	-,021	-,408**	1	-,281**	-,108**	,001	-,003	-,423**	,419**	-,015
	Sig. (2-tailed)	,712	,257	,081	,000		,000	,000	,947	,792	,000	,000	,199
	N	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099
PERIOD3	Pearson Correlation	-,001	-,001	,012	-,411**	-,281**	1	,034**	,001	-,004	-,045**	,096**	,037**
	Sig. (2-tailed)	,939	,907	,298	,000	,000		,004	,910	,767	,000	,000	,002
	N	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099
VOLUME_M	Pearson Correlation	,004	-,008	-,037**	-,016	-,108**	,034**	1	,266**	-,090**	,334**	-,082**	,340**
	Sig. (2-tailed)	,736	,490	,002	,192	,000	,004		,000	,000	,000	,000	,000
	N	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099
DIST_BEG	Pearson Correlation	,000	,003	,005	,026	,001	,001	,266**	1	-,092**	,184**	-,100**	,584**
	Sig. (2-tailed)	1,000	,820	,649	,027	,947	,910	,000		,000	,000	,000	,000
	N	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099
INTERSECT	Pearson Correlation	,000	,002	,000	,002	-,003	-,004	-,090**	-,092**	1	-,017	-,048**	,045**
	Sig. (2-tailed)	,995	,867	,990	,888	,792	,767	,000	,000		,153	,000	,000
	N	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099
SPEED	Pearson Correlation	,067**	,047**	,006	,136**	-,423**	-,045**	,334**	,184**	-,017	1	-,531**	,518**
	Sig. (2-tailed)	,000	,000	,635	,000	,000	,000	,000	,000	,153		,000	,000
	N	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099
SPEED_L	Pearson Correlation	-,003	-,114**	-,079**	-,281**	,419**	,096**	-,082**	-,100**	-,048**	-,531**	1	-,211**
	Sig. (2-tailed)	,789	,000	,000	,000	,000	,000	,000	,000	,000	,000		,000
	N	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099
VOLSPEED	Pearson Correlation	,039**	,042**	,010	,126**	-,015	,037**	,340**	,584**	,045**	,518**	-,211**	1
	Sig. (2-tailed)	,001	,000	,421	,000	,199	,002	,000	,000	,000	,000	,000	
	N	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099	7099

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

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

Συντελεστής συσχέτισης ανεξάρτητων μεταβλητών στο πρότυπο των
καθυστερήσεων για τη συνολική διαδρομή

Correlations

		DIST_BEGIN	DIST_BETW	MONDAY	WEDNESDAY	PERIOD1	PERIOD2	PERIOD3	SPEED
DIST_BEGIN	Pearson Correlation	1	,029**	-,001	,000	-,026**	-,003	,011*	-,013*
	Sig. (2-tailed)		,000	,913	,995	,000	,553	,049	,015
	N	32246	32246	32246	32246	32246	32246	32246	32246
DIST_BETW	Pearson Correlation	,029**	1	,001	-,002	-,006	-,001	-,001	,312**
	Sig. (2-tailed)	,000		,850	,774	,311	,926	,885	,000
	N	32246	32246	32246	32246	32246	32246	32246	32246
MONDAY	Pearson Correlation	-,001	,001	1	-,260**	-,036**	-,007	,004	,000
	Sig. (2-tailed)	,913	,850		,000	,000	,237	,473	,937
	N	32246	32246	32246	32246	32246	32246	32246	32246
WEDNESDAY	Pearson Correlation	,000	-,002	-,260**	1	,016**	-,020**	-,012*	,007
	Sig. (2-tailed)	,995	,774	,000		,005	,000	,031	,196
	N	32246	32246	32246	32246	32246	32246	32246	32246
PERIOD1	Pearson Correlation	-,026**	-,006	-,036**	,016**	1	-,402**	-,403**	,038**
	Sig. (2-tailed)	,000	,311	,000	,005		,000	,000	,000
	N	32246	32246	32246	32246	32246	32246	32246	32246
PERIOD2	Pearson Correlation	-,003	-,001	-,007	-,020**	-,402**	1	-,285**	-,082**
	Sig. (2-tailed)	,553	,926	,237	,000	,000		,000	,000
	N	32246	32246	32246	32246	32246	32246	32246	32246
PERIOD3	Pearson Correlation	,011*	-,001	,004	-,012*	-,403**	-,285**	1	-,041**
	Sig. (2-tailed)	,049	,885	,473	,031	,000	,000		,000
	N	32246	32246	32246	32246	32246	32246	32246	32246
SPEED	Pearson Correlation	-,013*	,312**	,000	,007	,038**	-,082**	-,041**	1
	Sig. (2-tailed)	,015	,000	,937	,196	,000	,000	,000	
	N	32246	32246	32246	32246	32246	32246	32246	32246

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

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

Συντελεστής συσχέτισης ανεξάρτητων μεταβλητών στο πρότυπο των
καθυστερήσεων για τη διαδρομή εντός του κέντρου

Correlations

		DIST_BEG	INTERSECT	MONDAY	TUESDAY	WEDNESDAY	VOLSPEED
DIST_BEG	Pearson Correlation	1	-,092**	,000	,003	,005	,584**
	Sig. (2-tailed)		,000	1,000	,820	,649	,000
	N	7099	7099	7099	7099	7099	7099
INTERSECT	Pearson Correlation	-,092**	1	,000	,002	,000	,045**
	Sig. (2-tailed)	,000		,995	,867	,990	,000
	N	7099	7099	7099	7099	7099	7099
MONDAY	Pearson Correlation	,000	,000	1	-,173**	-,264**	,039**
	Sig. (2-tailed)	1,000	,995		,000	,000	,001
	N	7099	7099	7099	7099	7099	7099
TUESDAY	Pearson Correlation	,003	,002	-,173**	1	-,283**	,042**
	Sig. (2-tailed)	,820	,867	,000		,000	,000
	N	7099	7099	7099	7099	7099	7099
WEDNESDAY	Pearson Correlation	,005	,000	-,264**	-,283**	1	,010
	Sig. (2-tailed)	,649	,990	,000	,000		,421
	N	7099	7099	7099	7099	7099	7099
VOLSPEED	Pearson Correlation	,584**	,045**	,039**	,042**	,010	1
	Sig. (2-tailed)	,000	,000	,001	,000	,421	
	N	7099	7099	7099	7099	7099	7099

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