



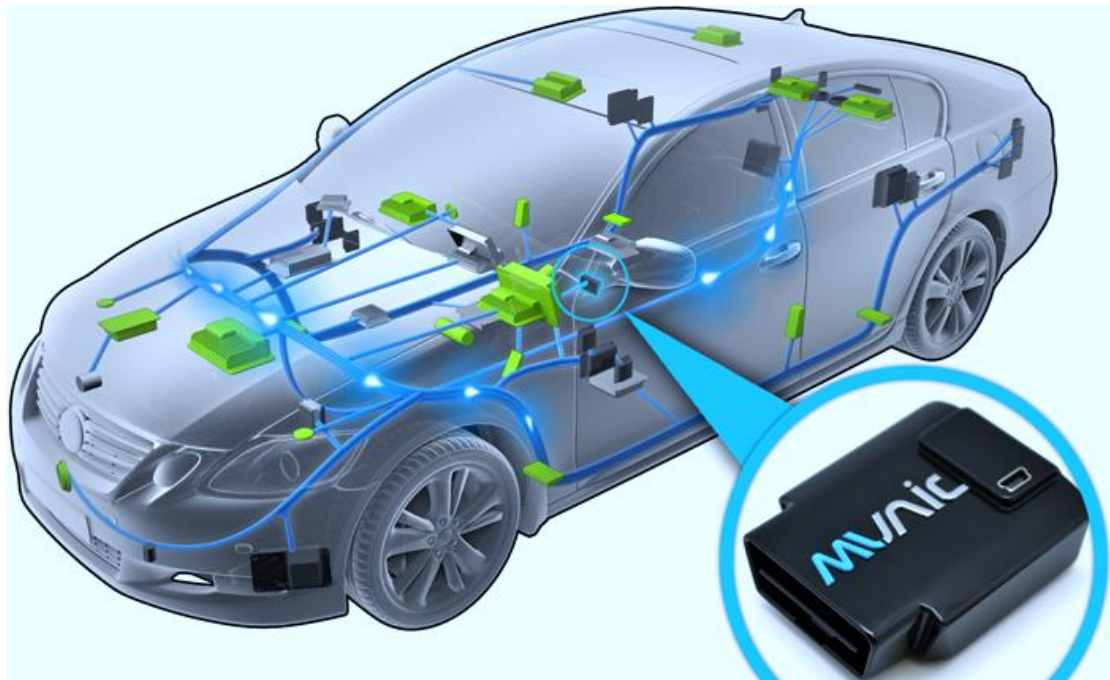
ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ ΚΑΙ ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ ΥΠΟΔΟΜΗΣ

ΔΙΕΡΕΥΝΗΣΗ ΤΗΣ ΚΥΚΛΟΦΟΡΙΑΚΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΤΟΥ ΟΔΗΓΟΥ ΜΕ ΧΡΗΣΗ ΤΩΝ ΔΙΑΓΝΩΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ ΤΟΥ ΟΧΗΜΑΤΟΣ

Διπλωματική Εργασία



ΣΑΠΛΑΟΥΡΑΣ ΕΥΣΤΡΑΤΙΟΣ

Επιβλέπων: Γιώργος Γιαννής, Καθηγητής ΕΜΠ

Αθήνα, Οκτώβριος 2015

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω θερμά τον κ. Γιώργο Γιαννή, Καθηγητή της Σχολής Πολιτικών Μηχανικών ΕΜΠ για την εμπιστοσύνη που μου έδειξε με την ανάθεση της παρούσας Διπλωματικής Εργασίας, για την υποστήριξη του καθώς και για την εκπληκτική συνεργασία που είχαμε σε όλα τα στάδια της. Επιπλέον, οφείλω να τον ευχαριστήσω για την σωστή καθοδήγηση αλλά και για τις γενικότερες γνώσεις που μου μετέδωσε.

Παράλληλα, θα ήθελα να ευχαριστήσω τα υπόλοιπα μέλη της εξεταστικής επιτροπής: κ. Ι. Γκόλια, Πρύτανη του Εθνικού Μετσόβιου Πολυτεχνείου και Καθηγητή της Σχολής Πολιτικών Μηχανικών ΕΜΠ και κ. Ε. Βλαχογιάννη, Επίκουρη Καθηγήτρια της Σχολής Πολιτικών Μηχανικών ΕΜΠ, για τις εποικοδομητικές παρατηρήσεις τους αλλά κυρίως για τις γνώσεις που μου προσέφεραν καθ'όλη τη διάρκεια της φοίτησής μου στη σχολή.

Ένα μεγάλο ευχαριστώ ανήκει στον κ. Δημήτρη Τσελέντη, Υποψήφιο Διδάκτορα ΕΜΠ, για τις συμβουλές και τις υποδείξεις του κατά τη διάρκεια εκπόνησης της παρούσας Διπλωματικής Εργασίας. Χωρίς τη βοήθειά του, δεν θα τα είχα καταφέρει ως εδώ. Εξίσου ευχαριστώ και τον κ. Αναστάσιο Θεοφιλάτο, Υποψήφιο Διδάκτορα ΕΜΠ, για τη βοήθειά του και τις υποδείξεις του σε σημαντικά ζητήματα της εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου, που βρίσκονται πάντα δίπλα μου προσφέροντάς μου απεριόριστη αγάπη και συμπαράσταση.

Αθήνα, Οκτώβριος 2015
Σαπλαούρας Ευστράτιος

ΔΙΕΡΕΥΝΗΣΗ ΤΗΣ ΚΥΚΛΟΦΟΡΙΑΚΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΤΟΥ ΟΔΗΓΟΥ ΜΕ ΧΡΗΣΗ ΤΩΝ ΔΙΑΓΝΩΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ ΤΟΥ ΟΧΗΜΑΤΟΣ

Ευστράτιος Σαπλαούρας
Επιβλέπων: Γιώργος Γιαννής, Καθηγητής ΕΜΠ

ΣΥΝΟΨΗ:

Στόχος της παρούσας Διπλωματικής Εργασίας είναι η διερεύνηση της κυκλοφοριακής συμπεριφοράς του οδηγού με χρήση των διαγνωστικών στοιχείων (OBD) του οχήματος. Αξιοποιήθηκε μεγάλος αριθμός δεδομένων που καταγράφονταν ανά δευτερόλεπτο και αφορούσαν στην ακριβή θέση του οχήματος, στην ταχύτητά του, στα σημεία όπου ο οδηγός πραγματοποιούσε απότομη αλλαγή ταχύτητας ή απότομο ελιγμό (events), στην κατανάλωση καυσίμων, κτλ. Αναπτύχθηκαν μαθηματικά στατιστικά μοντέλα γραμμικής και διωνυμικής λογιστικής παλινδρόμησης. Τα αποτελέσματα έδειξαν ισχυρή συσχέτιση μεταξύ της υπέρβασης του ανώτατου ορίου ταχύτητας και της απότομης συμπεριφοράς του οδηγού στην κυκλοφορία, με τα μεγέθη όμως να είναι δυσανάλογα, καθώς επίσης και αύξηση της ταχύτητας των οδηγών τις επικίνδυνες ώρες (00:00 – 04:00).

Λέξεις κλειδιά: κυκλοφοριακή συμπεριφορά, διαγνωστικά στοιχεία, γραμμική και λογιστική παλινδρόμηση, απότομη αλλαγή ταχύτητας, απότομος ελιγμός, υπέρβαση ανώτατου ορίου ταχύτητας

INVESTIGATION OF DRIVER'S BEHAVIOR USING THE DIAGNOSTIC ELEMENTS OF THE VEHICLE

Efstratios Saplaouras

Supervisor: George Yannis, Professor NTUA

ABSTRACT:

The aim of this diploma thesis is to investigate the driver's behavior through the on-board diagnostics of the vehicle. A large data set was exploited, which were recorded per second and concerned the exact position of the vehicle, its speed, the spots where the driver was performing a sudden change of speed or a sudden maneuver, the fuel consumption, etc. Mathematical statistical models were developed using linear and binary logistic regression. The results demonstrated a strong correlation between over-speeding and driver harsh behavior, although the two dimensions are disproportionate, as well as that higher driving speeds during risky hours (00:00-04:00).

Key words: driver's behavior, on-board diagnostics, linear and logistic regression, sudden change of speed, sudden maneuver, over-speeding

ΠΕΡΙΛΗΨΗ:

Αντικείμενο της παρούσας Διπλωματικής Εργασίας αποτελεί η διερεύνηση της κυκλοφοριακής συμπεριφοράς του οδηγού με χρήση των διαγνωστικών στοιχείων του οχήματος. Δεδομένου ότι η δυνατότητα αξιοποίησης των στοιχείων αυτών είναι σχετικά πρόσφατη, πρόκειται για ένα θέμα το οποίο δεν έχει μελετηθεί ιδιαίτερα ούτε διεθνώς ούτε στην Ελλάδα, ωστόσο, όπως προκύπτει και από τη βιβλιογραφική ανασκόπηση, τα τελευταία χρόνια χάρη στην εξέλιξη της τεχνολογίας, παρατηρείται μία αυξανόμενη ενασχόληση του επιστημονικού κόσμου γύρω από το συγκεκριμένο ζήτημα.

Η **συλλογή των απαιτούμενων στοιχείων** για τη διερεύνηση της κυκλοφοριακής συμπεριφοράς του οδηγού πραγματοποιήθηκε με μεταφορά των σχετικών στοιχείων από το σύστημα διαγνωστικών του οχήματος, μέσω μίας ειδικής συσκευής που κατέγραφε δεδομένα ανά δευτερόλεπτο.

Μετά από κατάλληλη επεξεργασία και μία σειρά δοκιμών αναπτύχθηκαν μαθηματικά μοντέλα με τη μέθοδο της **διωνυμικής λογιστικής παλινδρόμησης** για τις παρακάτω εξαρτημένες μεταβλητές, από ένα μεγάλο αριθμό στοιχείων ανά δευτερόλεπτο που συλλέχθηκαν από τα διαγνωστικά του οχήματος:

- ✓ Απότομη Δεξιά Στροφή
- ✓ Απότομη Αριστερή Στροφή
- ✓ Απότομη Επιβράδυνση
- ✓ Απότομη Επιτάχυνση
- ✓ Παραβίαση Ανώτατου Ορίου Ταχύτητας

Αναπτύχθηκε και ένα μαθηματικό μοντέλο με τη μέθοδο της **γραμμικής παλινδρόμησης** για την ταχύτητα, καθότι πρόκειται για συνεχή μεταβλητή. Όλα τα μαθηματικά μοντέλα συνοψίζονται στον παρακάτω πίνακα.

Πίνακας: Συγκεντρωτικός πίνακας αποτελεσμάτων μαθηματικών μοντέλων

Ανεξάρτητες Μεταβλητές	Απότομη Δεξιά Στροφή	Απότομη Αριστερή Στροφή	Απότομη Επιβράδυνση	Απότομη Επιτάχυνση	Παραβίαση Ανώτατου Ορίου Ταχύτητας	Ταχύτητα Οχήματος
Σταθερό	-0.129	-0.046	-0.469	0.826	-3.575	38.022
Ημέρα Οδήγησης						
Επικίνδυνες Ώρες						2.375
Τύπος Οδού						
Απότομη Επιτάχυνση	-1.651	-2.158			-4.316	-16.032
Απότομη Επιβράδυνση	-1.714	-2.022			-5.216	-10.006
Απότομη Δεξιά Στροφή						
Απότομη Αριστερή Στροφή						
Παραβίαση Ανώτατου Ορίου Ταχύτητας	-5.281	-6.083	-5.549	-3.906		38.385
Ταχύτητα Οχήματος				-0.041	0.122	
Κατανάλωση Καυσίμων						
Στροφές ανά λεπτό						

Από τα διάφορα στάδια εκπόνησης της παρούσας Διπλωματικής Εργασίας προέκυψε μια σειρά συμπερασμάτων, όπως αυτά συνοψίζονται παρακάτω:

- 1) Τα δεδομένα τα οποία βρίσκονται στο σύστημα διαγνωστικών του οχήματος μπορούν να αποθηκευτούν ηλεκτρονικά σε υπολογιστή και περιέχουν ιδιαιτέρως σημαντικές πληροφορίες οι οποίες, μετά από κατάλληλη επεξεργασία και ανάπτυξη μαθηματικών μοντέλων, μπορούν να χρησιμεύσουν στην εξαγωγή χρήσιμων συμπερασμάτων για την **κυκλοφοριακή συμπεριφορά του οδηγού**.
- 2) Βασικός παράγοντας για την εμφάνιση κάποιου απότομου συμβάντος κατά τη διάρκεια της οδήγησης είναι η **παραβίαση του ορίου ταχύτητας**, όπως αποδείχθηκε από την ανάπτυξη των μαθηματικών μοντέλων. Η μαθηματική σχέση που προκύπτει για κάθε εξαρτημένη μεταβλητή αποδεικνύει την ισχυρή συσχέτιση μεταξύ της παραβίασης του ορίου ταχύτητας και των εξεταζομένων μεταβλητών. Στην περίπτωση της απότομης επιβράδυνσης, το over-speeding αποτελεί τη μοναδική μεταβλητή που επηρεάζει την εμφάνιση ενός τέτοιου συμβάντος.
- 3) Ωστόσο, στα σημεία όπου ο οδηγός βρίσκεται άνω των ορίων ταχύτητας, δείχνει να είναι αρκετά **προσεκτικός** όσον αφορά την οδήγησή του και να μην προβαίνει ούτε σε απότομη επιτάχυνση ή επιβράδυνση ούτε σε απότομους ελιγμούς. Αυτό φαίνεται μέσω των αρνητικών συντελεστών που προκύπτουν στις μαθηματικές σχέσεις για τις μεταβλητές που εξετάστηκαν. Το αποτέλεσμα, πιθανώς, καταδεικνύει την εμπειρία του οδηγού να διαχειριστεί κατάλληλα το όχημά του ακόμα και σε υψηλές ταχύτητες, χωρίς να προβαίνει σε επικίνδυνες ενέργειες.
- 4) Η μόνη περίπτωση όπου η παραβίαση του ορίου ταχύτητας επιδρά θετικά σε κάποια μεταβλητή είναι στην γραμμική ανάλυση παλινδρόμησης και στην **ταχύτητα του οχήματος**. Όταν ο οδηγός υπερβαίνει το όριο ταχύτητας, είναι λογικό η ταχύτητα του οχήματος να αυξάνεται. Στην περίπτωση της ταχύτητας του οχήματος αναπτύχθηκε η **γραμμική παλινδρόμηση** καθώς περιγράφει καλύτερα τις συνεχείς μεταβλητές.
- 5) Εξετάζοντας τους **απότομους ελιγμούς** του οδηγού, το συμπέρασμα που εξάγεται είναι ότι βρίσκονται σε άμεση συσχέτιση με τις απότομες εναλλαγές ταχύτητας και, φυσικά, με την υπέρβαση του ορίου ταχύτητας. Παρόλα αυτά, **το αντίστροφο δεν ισχύει**. Ειδικότερα, οι αναλύσεις έδειξαν ότι η εμφάνιση απότομων επιταχύνσεων ή επιβραδύνσεων κατά τη διάρκεια της οδήγησης δεν μπορούν να θεωρηθούν ως συνάρτηση των απότομων δεξιών ή αριστερών στροφών τις οποίες πραγματοποιεί ο οδηγός στη διαδρομή του.
- 6) Η ώρα οδήγησης δείχνει να επηρεάζει μόνο την ταχύτητα του οχήματος και όχι τη συνολική κυκλοφοριακή συμπεριφορά του οδηγού. Κατά τη διάρκεια **των επικίνδυνων ωρών** (00:00 - 4:00), παρατηρείται αύξηση της ταχύτητας του

οχήματος, κάτι το οποίο συμβαίνει –πιθανότατα– λόγω μικρότερου κυκλοφοριακού φόρτου στις οδούς. Παρόλα αυτά, δεν προκύπτει ότι στο συγκεκριμένο χρονικό διάστημα ο οδηγός είτε είναι πιο απρόσεκτος είτε πιο επικίνδυνος καθώς στα υπόλοιπα μαθηματικά μοντέλα, το επίπεδο σημαντικότητας της μεταβλητής των επικίνδυνων ωρών ήταν ιδιαίτερος χαμηλό.

- 7) Οι υπόλοιπες μεταβλητές που υπήρχαν στη βάση δεδομένων και εξετάστηκαν όπως η ημέρα οδήγησης, ο τύπος της οδού, η κατανάλωση καυσίμων και οι στροφές ανά λεπτό **δεν φαίνεται να έχουν ουσιαστική επιρροή** στη συμπεριφορά του οδηγού. Σε όλα τα μαθηματικά μοντέλα που αναπτύχθηκαν, οι συγκεκριμένες μεταβλητές εμφάνιζαν ένα ιδιαίτερα χαμηλό επίπεδο σημαντικότητας, καταδεικνύοντας την μη επίδρασή τους στην επίτευξη του στόχου της παρούσας διπλωματικής εργασίας.
- 8) Η στατιστική επεξεργασία των περισσότερων στοιχείων πραγματοποιήθηκε με τη μέθοδο της **λογιστικής ανάλυσης παλινδρόμησης**, που αποδείχθηκε κατάλληλη για τέτοιου είδους ανάλυση. Η ανάλυση των στοιχείων με αυτή τη μέθοδο οδήγησε στην ανάπτυξη αξιόπιστων μαθηματικών μοντέλων διερεύνησης της κυκλοφοριακής συμπεριφοράς του οδηγού. Η επιλογή της συγκεκριμένης μεθόδου έγινε καθώς οι μεταβλητές προς εξέταση ήταν διακριτές, λαμβάνοντας τις τιμές 0 και 1.
- 9) Τα σαφή και ερμηνεύσιμα αποτελέσματα της παρούσας εργασίας καταδεικνύουν την **καταλληλότητα της μεθόδου ανάλυσης** σχετικά με την διερεύνηση της κυκλοφοριακής συμπεριφοράς του οδηγού μέσω των διαγνωστικών στοιχείων του οχήματος. Η αποτελεσματικότητα της μεθόδου επιτρέπει την εφαρμογή της σε έρευνες με παρόμοιο αντικείμενο έπειτα από κατάλληλη προσαρμογή. Βέβαια, οι μεταβλητές θα πρέπει να επιλέγονται ανάλογα με την εξεταζόμενη περίπτωση.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ.....	3
ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ	4
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ.....	5
1. ΕΙΣΑΓΩΓΗ.....	6
1.1 Γενική Ανασκόπηση.....	6
1.2 Στόχος Διπλωματικής Εργασίας	10
1.3 Μεθοδολογία	10
1.4 Δομή της Διπλωματικής Εργασίας	11
2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ	14
2.1 Γενικά	14
2.2 Συναφείς Έρευνες.....	14
2.2.1 Development of a driving data recorder (1994)	16
2.2.2 TrafficView: traffic data dissemination using car-to-car communication (2004)	17
2.2.3 Feedback by technology: Attitudes and opinions of truck drivers (2005)	18
2.2.4 In-vehicle data recorders for monitoring and feedback on drivers' behavior (2008)	19
2.2.5 Effects of in-vehicle monitoring on the driving behavior of teenagers (2010)	20
2.2.6 Modeling the behavior of novice young drivers during the first year after licensure (2010).....	21
2.2.7 Providing accident detection in vehicular networks through OBD-II devices and Android-based smartphones (2011)	22
2.2.8 Driving behavior and traffic safety: an acceleration-based safety evaluation procedure for smartphones (2014).....	23
2.2.9 Determining the sampling time frame for In-Vehicle Data Recorder measurement in assessing drivers (2014).....	24
2.3 Σύνοψη	25
3. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ	26
3.1 Εισαγωγή.....	26
3.2 Βασικές Έννοιες Στατιστικής	26
3.3 Συσχέτιση Μεταβλητών – Συντελεστής Συσχέτισης	28
3.4 Βασικές Κατανομές	28
3.4.1 Κανονική Κατανομή	29
3.4.2 Κατανομή Poisson	29
3.4.3 Αρνητική Διωνυμική Κατανομή	30
3.5 Μαθηματικά Πρότυπα	30

3.5.1	Γραμμική Παλινδρόμηση	30
3.5.1.1	Μέθοδος Ελαχίστων Τετραγώνων	32
3.5.2	Λογιστική Ανάλυση Παλινδρόμησης	32
3.5.3	Λογαριθμοκανονική Παλινδρόμηση	34
3.6	Διαδικασία Ανάπτυξης και Κριτήρια Αποδοχής Μοντέλου	34
3.7	Λειτουργία του Ειδικού Στατιστικού Λογισμικού	37
4.	ΣΥΛΛΟΓΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΣΤΟΙΧΕΙΩΝ	39
4.1	Εισαγωγή	39
4.2	Συλλογή στοιχείων	39
4.3	Επεξεργασία Στοιχείων	40
4.3.1	Διαμόρφωση Δεδομένων	40
4.3.2	Περιγραφικά Στατιστικά (Descriptive Statistics)	41
4.3.3	Διαμόρφωση βάσης δεδομένων	43
4.3.4	Εισαγωγή δεδομένων στο ειδικό στατιστικό πρόγραμμα	46
5.	ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ – ΑΠΟΤΕΛΕΣΜΑΤΑ	47
5.1	Εισαγωγή	47
5.2	Ανάπτυξη και Εφαρμογή Μοντέλων	47
5.2.1	Δεδομένα εισόδου – Καθορισμός Μεταβλητών	47
5.2.2	Λογιστική Ανάλυση Παλινδρόμησης	48
5.2.3	Περιγραφή Μοντέλων Λογιστικής Παλινδρόμησης	49
5.2.4	Γραμμική Ανάλυση Παλινδρόμησης	49
5.2.5	Τελικά Μοντέλα για την Κυκλοφοριακή Συμπεριφορά του Οδηγού	50
5.3	Αποτελέσματα Μοντέλων	54
5.4	Εξήγηση συνολικών αποτελεσμάτων	56
5.5	Γενικές Παρατηρήσεις	57
6.	ΣΥΜΠΕΡΑΣΜΑΤΑ	59
6.1	Σύνοψη Αποτελεσμάτων	59
6.2	Συνολικά Συμπεράσματα	61
6.3	Προτάσεις για Βελτίωση της Κυκλοφοριακής Συμπεριφοράς	63
6.4	Προτάσεις για Περαιτέρω Έρευνα	63
7.	ΒΙΒΛΙΟΓΡΑΦΙΑ	65

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

ΑΡΙΘΜΗΣΗ	ΤΙΤΛΟΣ	ΣΕΛΙΔΑ
3.1	Κρίσιμες τιμές του συντελεστή t	36
4.1	Όρια ταχύτητας βάσει ΚΟΚ	42
4.2	Descriptive Statistics - driver 1	43
4.3	Descriptive Statistics - driver 2	43
4.4	Ονομασία ανά τύπο οδού	44
4.5	Ονομασία ανά ημέρα	44
5.1	Περίληψη Μοντέλου (Model Summary)	50
5.2	Πίνακας Κατάταξης (Classification Table) ^a	50
5.3	Μεταβλητές στην εξίσωση (Variables in the Equation)	50
5.4	Περίληψη Μοντέλου (Model Summary)	51
5.5	Πίνακας Κατάταξης (Classification Table) ^a	51
5.6	Μεταβλητές στην εξίσωση (Variables in the Equation)	51
5.7	Περίληψη Μοντέλου (Model Summary)	51
5.8	Πίνακας Κατάταξης (Classification Table) ^a	51
5.9	Μεταβλητές στην εξίσωση (Variables in the Equation)	52
5.10	Περίληψη Μοντέλου (Model Summary)	52
5.11	Πίνακας Κατάταξης (Classification Table) ^a	52
5.12	Μεταβλητές στην εξίσωση (Variables in the Equation)	52
5.13	Περίληψη Μοντέλου (Model Summary)	52
5.14	Πίνακας Κατάταξης (Classification Table) ^a	53
5.15	Μεταβλητές στην εξίσωση (Variables in the Equation)	53
5.16	Ανάλυση Διακύμανσης (ANOVA) ^a	53
5.17	Περίληψη Μοντέλου (Model Summary)	53
5.18	Συντελεστές (Coefficients) ^a	54
6.1	Συγκεντρωτικός πίνακας αποτελεσμάτων μαθηματικών μοντέλων	60

ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

ΑΡΙΘΜΗΣΗ	ΤΙΤΛΟΣ	ΣΕΛΙΔΑ
1.1	Θύματα τροχαίων ατυχημάτων, Σύγκριση 2013-2014 [πηγή: Ελληνική Αστυνομία (www.astynomia.gr)]	6
1.2	Νεκροί ανά εκατομμύρια πληθυσμού, 2014 [πηγή: European Transport Safety Council (www.etsc.eu)]	7
1.3	Θύματα τροχαίων ατυχημάτων, Σύγκριση Α'εξαμήνου 2014-2015 [πηγή: Ελληνική Αστυνομία (www.astynomia.gr)]	8
1.4	Διάγραμμα ροής των σταδίων εκπόνησης της Διπλωματικής Εργασίας	11
3.1	Ευθεία ελαχίστων τετραγώνων	32

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

ΑΡΙΘΜΗΣΗ	ΤΙΤΛΟΣ	ΣΕΛΙΔΑ
1.1	Χάρτης οδικού δικτύου με χιλιομετρικές θέσεις συχνότερων τροχαίων σοβαρών ατυχημάτων, 2014 [πηγή: Ελληνική Αστυνομία (www.astynomia.gr)]	9
4.1	Σχηματική απεικόνιση λειτουργίας συστήματος	39
4.2	Απόσπασμα αρχικού αρχείου csv	40
4.3	Απόσπασμα δεδομένων σε αρχείο Excel	41
4.4	Χάρτης διαδρομής οδηγού	42
4.5	Εισαγωγή των στοιχείων στο πεδίο δεδομένων	46
4.6	Καθορισμός των μεταβλητών στο πεδίο μεταβλητών	46

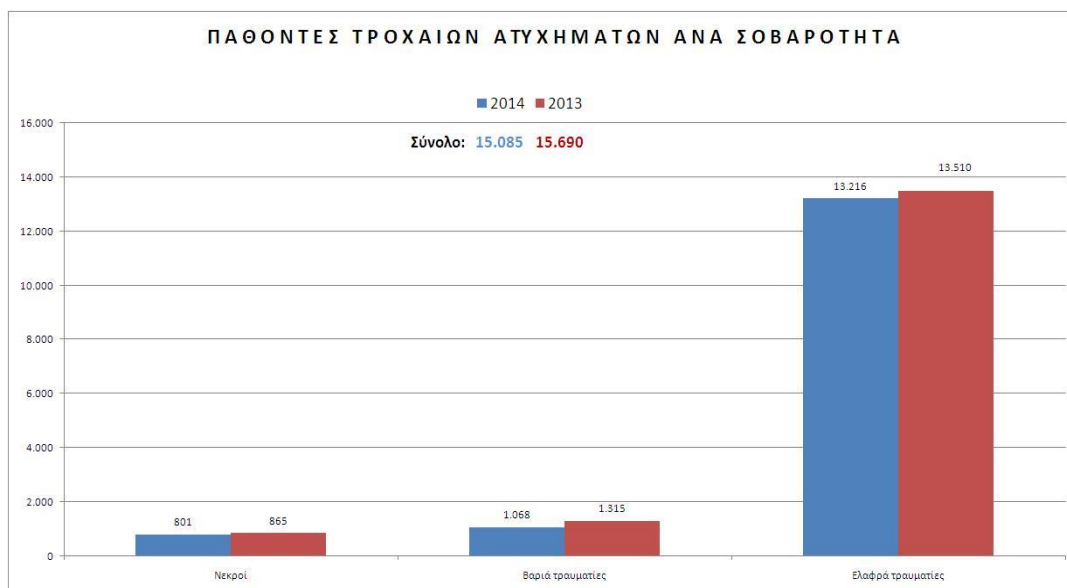
1. ΕΙΣΑΓΩΓΗ

1.1 Γενική Ανασκόπηση

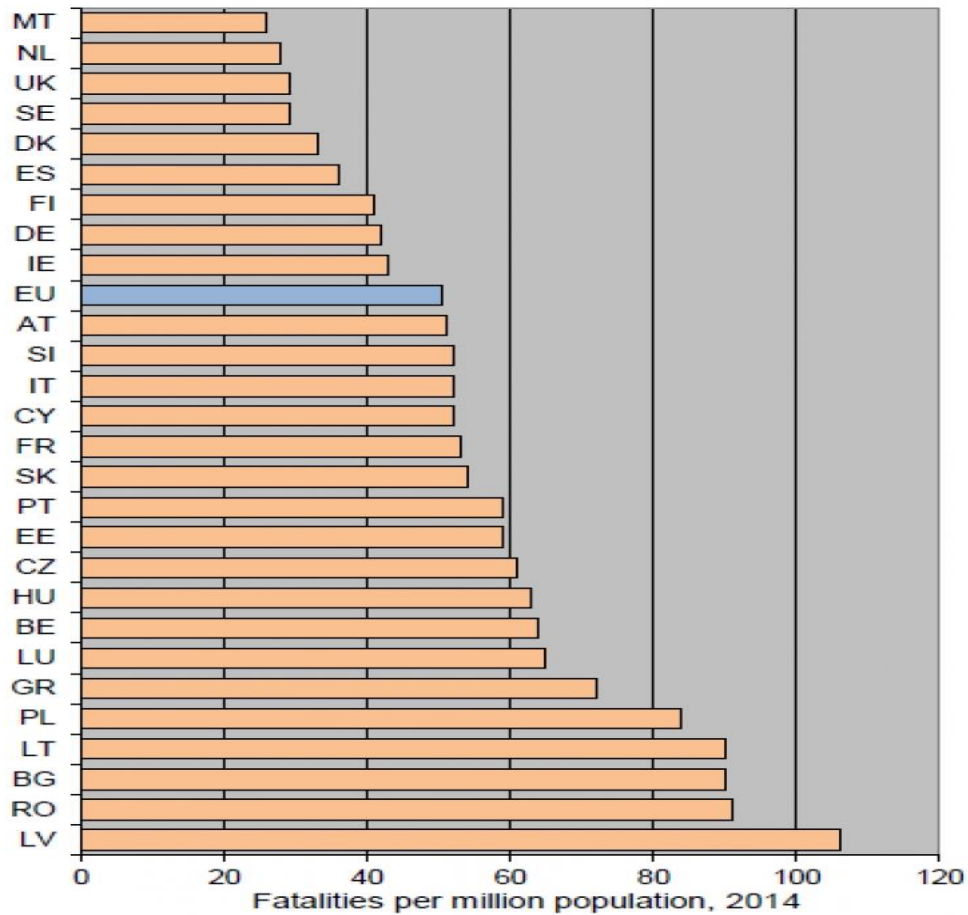
Οι **οδικές μεταφορές** είναι συνυφασμένες με την ύπαρξη της ανθρώπινης ζωής και αποτελούν δείκτη της εξέλιξης των κοινωνιών, καθώς συμβάλλουν στην αναβάθμιση του βιοτικού επιπέδου. Στη σύγχρονη εποχή, η περαιτέρω πρόοδος στον τομέα των μεταφορών, προσώπων και αγαθών, αποτελεί αντικείμενο της επιστήμης του συγκοινωνιολόγου μηχανικού και έχει στόχο την παροχή ασφαλών, γρήγορων, οικονομικών και άνετων μετακινήσεων.

Εδώ και αρκετά χρόνια η **οδική ασφάλεια** αποτελεί αντικείμενο ενδιαφέροντος από τους ερευνητές και τις κυβερνήσεις των χωρών, καθώς η αυξανόμενη ζήτηση για μετακινήσεις οδηγεί καθημερινά σε ατυχήματα, με κοινωνικό και οικονομικό κόστος. Τα τροχαία ατυχήματα εκτιμάται διεθνώς ότι αποτελούν την τρίτη αιτία θανάτου, μετά τις καρδιοπάθειες και τον καρκίνο, ενώ αποτελούν και μία από τις βασικότερες αιτίες πρόκλησης μόνιμης αναπηρίας σε παγκόσμια κλίμακα, κυρίως σε άτομα νεαρής ηλικίας. Αν και έχουν γίνει βήματα προς την κατεύθυνση της μείωσης των τροχαίων ατυχημάτων εκτιμάται ότι το 2010 σε ατυχήματα σκοτώθηκαν 1,3 εκατομμύρια άνθρωποι **παγκοσμίως** και τραυματίστηκαν 50 εκατομμύρια (IRTAD, 2011).

Η χώρα μας, δυστυχώς, εμφανίζει υψηλό δείκτη τροχαίων ατυχημάτων. Σύμφωνα με την τροχαία, το 2014 στην **Ελλάδα** καταγράφηκαν 11.679 ατυχήματα, με 15.085 παθόντες εκ των οποίων οι 801 ήταν νεκροί και οι 1.068 βαριά τραυματίες. Την ίδια χρονιά στην **Ευρώπη** 25.640 άνθρωποι έχασαν τη ζωή τους σε οδικά ατυχήματα (European Commission, 2015).

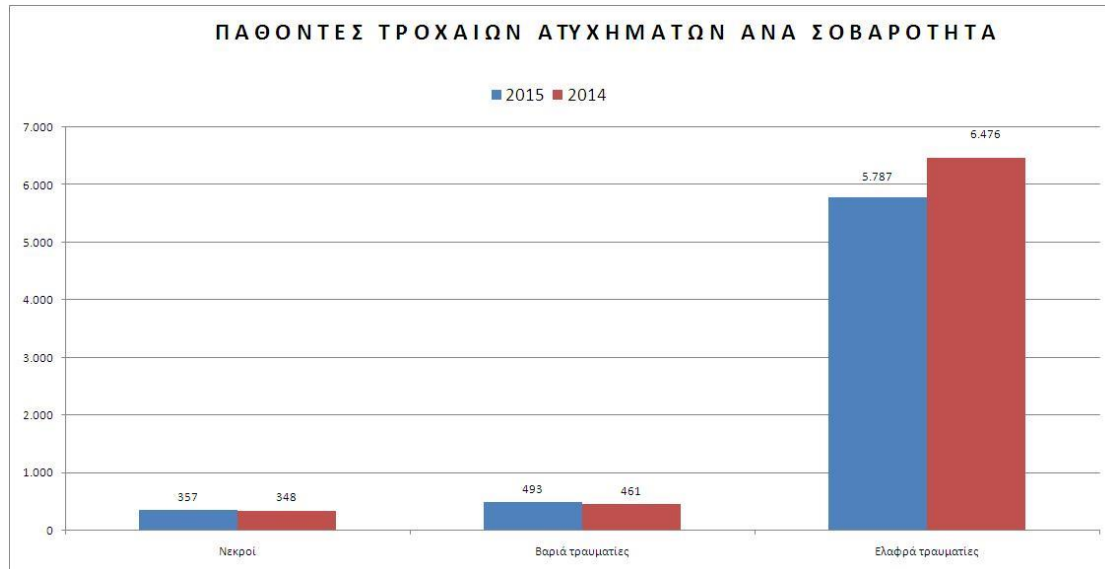


Διάγραμμα 1.1: Θύματα τροχαίων ατυχημάτων, Σύγκριση 2013-2014 [πηγή: Ελληνική Αστυνομία (www.astynomia.gr)]



Διάγραμμα 1.2: Νεκροί ανά εκατομμύρια πληθυσμού, 2014 [πηγή: European Transport Safety Council (www.etsc.eu)]

Δυστυχώς, στο πρώτο εξάμηνο του 2015 φαίνεται να έχουμε μια μικρή αύξηση στον αριθμό των νεκρών και βαριά τραυματιών στη χώρα μας σε σχέση με την αντίστοιχη περίοδο κατά το 2014.



Διάγραμμα 1.3: Θύματα τροχαίων ατυχημάτων, Σύγκριση Α'εξαμήνου 2014-2015 [πηγή: Ελληνική Αστυνομία (www.astynomia.gr)]

✓ Τα αίτια των οδικών ατυχημάτων

Το θέμα της οδικής ασφάλειας είναι περίπλοκο καθώς εξαρτάται από πολλούς **παράγοντες** οι σπουδαιότεροι εκ των οποίων κατά σειρά αυξανόμενης σπουδαιότητας είναι:

- ❖ το όχημα
- ❖ η οδός και το περιβάλλον
- ❖ οι χρήστες της οδού

Στις περισσότερες περιπτώσεις δύο ή και τρεις από τους παραπάνω παράγοντες συμβάλλουν στο ατύχημα. Η πολυπλοκότητα και η έλλειψη λεπτομερούς καταγραφής και ανάλυσης των συνθηκών υπό τις οποίες έγινε το ατύχημα, δεν επιτρέπουν πάντα την αντικειμενική διαπίστωση της συμβολής κάθε παράγοντα. Εν τούτοις, διάφορες μελέτες ατυχημάτων σε βάθος δείχνουν ότι ο χρήστης της οδού μόνος ή σε συνδυασμό με τους άλλους δύο παράγοντες, αποτελεί την κύρια αιτία των οδικών ατυχημάτων.

Το όχημα

Τα χαρακτηριστικά των οχημάτων με τα οποία σχετίζονται τα ατυχήματα είναι:

- η ηλικία του οχήματος
- το μέγεθος του οχήματος
- ο τύπος του οχήματος

Η οδός και το περιβάλλον

Ατυχήματα προκαλούν οι παρακάτω συνθήκες στην οδό και γενικότερα στο περιβάλλον:

- Ανεπαρκή γεωμετρικά χαρακτηριστικά
- Χαμηλά πρότυπα κατασκευής
- Κακή μελέτη, τοποθέτηση και κατασκευή παρόδιων στοιχείων

- Κακή οργάνωση της κυκλοφορίας
- Ανεπαρκής έλεγχος και σήμανση κατά τη διάρκεια των κατασκευών
- Δυσμενείς περιβαλλοντικές συνθήκες



Εικόνα 1.1: Χάρτης οδικού δικτύου με χιλιομετρικές θέσεις συχνότερων τροχαίων σοβαρών ατυχημάτων, 2014 [πηγή: Ελληνική Αστυνομία (www.astynomia.gr)]

Οι χρήστες της οδού

Η συμπεριφορά του χρήστη της οδού, είτε ως οδηγός ή επιβάτης ενός οχήματος είτε ως πεζός, αποτελεί το σημαντικότερο παράγοντα για τα οδικά ατυχήματα. Τα ατυχήματα που αποδίδονται στη χρήση της οδού προκαλούνται κυρίως από την παράβαση των κανόνων της οδικής κυκλοφορίας, όπως:

- υπερβολική ταχύτητα
- λανθασμένη προσπέραση
- στροφή ή στάση
- απρόσεκτη οδήγηση
- μέθη
- μη συμμόρφωση με τη σήμανση και σηματοδότηση
- απρόσεκτη διάσχιση οδών από πεζούς.

Γίνεται, επομένως, απαραίτητη η ανάλυση και η συνεχής παρακολούθηση της κυκλοφοριακής συμπεριφοράς του οδηγού είτε με χρήση προσομοιωτή οδήγησης είτε με χρήση δεδομένων σε πραγματικό χρόνο.

1.2 Στόχος Διπλωματικής Εργασίας

Σύμφωνα με όσα αναφέρθηκαν παραπάνω, στόχος της παρούσας Διπλωματικής Εργασίας είναι η **διερεύνηση της κυκλοφοριακής συμπεριφοράς του οδηγού με χρήση των διαγνωστικών στοιχείων του οχήματος.**

Συγκεκριμένα, θα εξεταστεί το πώς τα διάφορα στοιχεία που συνθέτουν τον τρόπο οδήγησης του οδηγού (ταχύτητα, επιτάχυνση, επιβράδυνση κτλ) αλληλεπιδρούν μεταξύ τους και καθορίζουν την οδηγική του συμπεριφορά. Προκειμένου να επιτευχθεί αυτή η προσπάθεια, απαιτείται η εφαρμογή κατάλληλων μεθόδων ανάλυσης των δεδομένων. Επομένως, επιμέρους στόχος της Διπλωματικής Εργασίας αποτελεί η **επιλογή της κατάλληλης μεθόδου** για την ανάπτυξη ενός μαθηματικού μοντέλου, που θα αποτυπώνει επαρκώς τη σχέση μεταξύ εξαρτημένων και ανεξάρτητων μεταβλητών.

Εκτιμάται ότι τα αποτελέσματα που θα προκύψουν, με το τέλος της Διπλωματικής Εργασίας, θα επιτρέψουν την κατανόηση του βαθμού και του τρόπου με τον οποίο τα διαγνωστικά στοιχεία του οχήματος είναι σε θέση να εκφράσουν την κυκλοφοριακή συμπεριφορά του οδηγού.

1.3 Μεθοδολογία

Στο υποκεφάλαιο αυτό περιγράφεται συνοπτικά η μέθοδος που ακολουθήθηκε για την επίτευξη του στόχου της Διπλωματικής Εργασίας.

Αρχικά καθορίστηκε το αντικείμενο που θα εξέταζε η παρούσα Διπλωματική Εργασία καθώς και ο επιδιωκόμενος στόχος. Για την υλοποίηση του στόχου πραγματοποιήθηκε ευρεία **βιβλιογραφική ανασκόπηση**. Αναζητήθηκαν, δηλαδή, έρευνες με θέμα συναφές με εκείνο της Διπλωματικής Εργασίας τόσο σε ελληνικό, όσο και σε διεθνές επίπεδο.

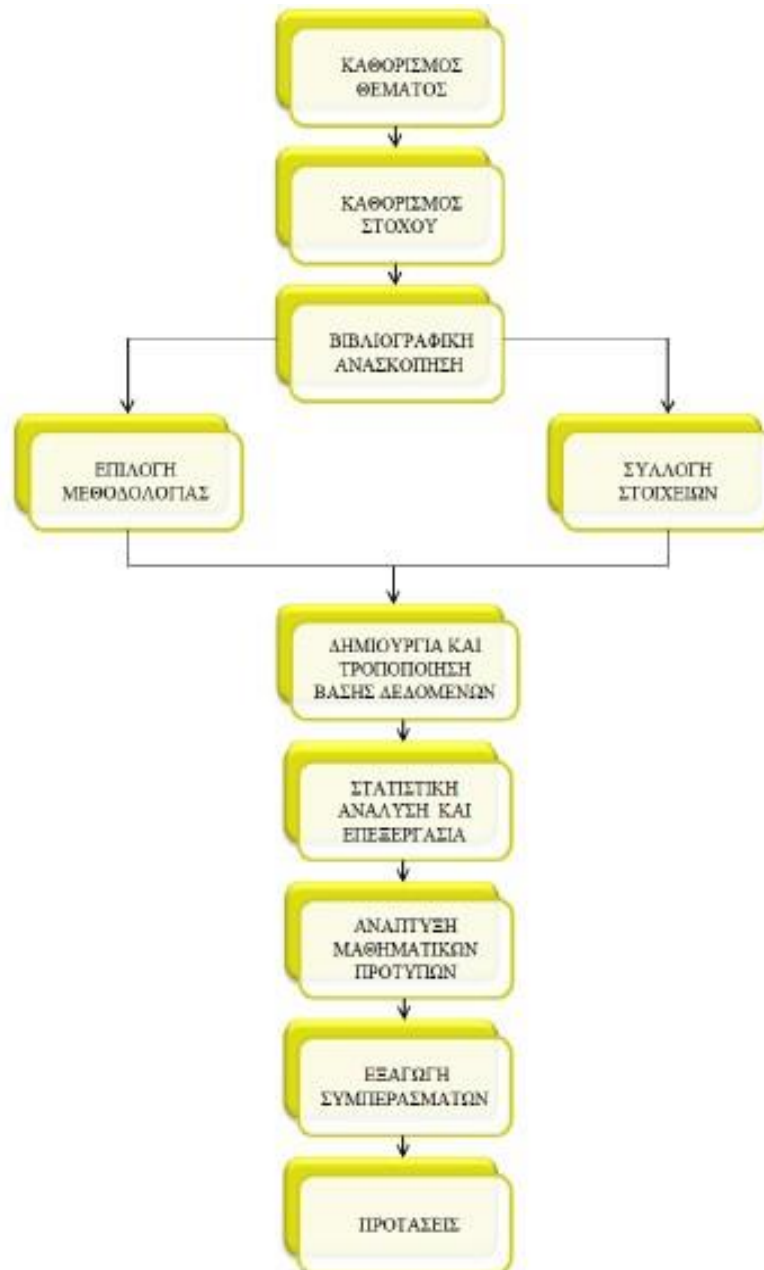
Μετά την ολοκλήρωση της αναζήτησης βιβλιογραφικών αναφορών, σειρά είχε η εύρεση του τρόπου **συλλογής των στοιχείων**, τα οποία θα μας βοηθούσαν να εκτιμήσουμε την κυκλοφοριακή συμπεριφορά του οδηγού.

Στη συνέχεια, τα στοιχεία καταχωρήθηκαν σε ηλεκτρονική βάση δεδομένων (Microsoft Excel), η οποία βελτιώθηκε σταδιακά, έως ότου αποκτήσει την τελική της μορφή. Ακολούθησε η **επιλογή της μεθόδου** στατιστικής επεξεργασίας των στοιχείων και η εισαγωγή της βάσης δεδομένων στο ειδικό λογισμικό στατιστικής ανάλυσης (SPSS 21.0).

Την επιλογή της κατάλληλης μεθόδου διαδέχτηκε η ανάπτυξη των μαθηματικών μοντέλων και η **παρουσίαση των αποτελεσμάτων**, στο πλαίσιο της οποίας

πραγματοποιήθηκε περιγραφή της διερεύνησης της κυκλοφοριακής συμπεριφοράς του οδηγού με χρήση των διαγνωστικών στοιχείων του οχήματος.

Στη συνέχεια, παρουσιάζονται, υπό μορφή διαγράμματος ροής, τα διαδοχικά στάδια που ακολουθήθηκαν κατά την εκπόνηση της Διπλωματικής Εργασίας.



Διάγραμμα 1.4: Διάγραμμα ροής των σταδίων εκπόνησης της Διπλωματικής Εργασίας

1.4 Δομή της Διπλωματικής Εργασίας

Στο υποκεφάλαιο αυτό παρουσιάζεται η δομή της Διπλωματικής Εργασίας, μέσω της συνοπτικής αναφοράς στο περιεχόμενο των κεφαλαίων της.

Στο **Κεφάλαιο 1** γίνεται μια εισαγωγή στο αντικείμενο της Διπλωματικής Εργασίας, προκειμένου να προσδιοριστεί ο στόχος της. Αρχικά, γίνεται μια γενική ανασκόπηση, όπου παρουσιάζεται το πλαίσιο της Διπλωματικής Εργασίας και αναφέρονται τα δεδομένα όπως έχουν προκύψει ως σήμερα, όσον αφορά την οδική ασφάλεια. Παρατίθενται, επίσης, τα βασικά στατιστικά στοιχεία για την Ελλάδα. Τέλος, παρουσιάζεται η μεθοδολογία που ακολουθήθηκε, περιγράφοντας συνοπτικά τα βασικά στάδια της εκτέλεσης της παρούσας εργασίας.

Στο **Κεφάλαιο 2** παρουσιάζονται τα αποτελέσματα της βιβλιογραφικής ανασκόπησης, όπως αυτά προέκυψαν από την αναζήτηση ερευνών με παρεμφερές αντικείμενο. Παρατίθενται εργασίες από την Ελλάδα και το εξωτερικό, οι οποίες έχουν δημοσιευθεί σε συνέδρια, επιστημονικά περιοδικά, άρθρα ή συγγράματα. Πραγματοποιείται σύνοψη της κάθε εργασίας που περιλαμβάνει το πλαίσιο της έρευνας, τη μεθοδολογία και τα βασικά αποτελέσματα, με έμφαση στα στοιχεία που παρουσιάζουν συνάφεια με το αντικείμενο της παρούσας Διπλωματικής Εργασίας. Το κεφάλαιο κλείνει με τη σύνοψη και κριτική αξιολόγηση των ερευνών, προκειμένου να διαπιστωθεί εάν και κατά πόσον κάποιες από αυτές είναι ικανές να συμβάλλουν στην παρούσα Διπλωματική Εργασία.

Στο **Κεφάλαιο 3** παρουσιάζεται το θεωρητικό υπόβαθρο στο οποίο στηρίχθηκε η ανάλυση των στοιχείων. Αρχικά περιγράφονται βασικές μαθηματικές και στατιστικές έννοιες και στη συνέχεια αναλύονται οι προϋποθέσεις εφαρμογής και τα επιμέρους στοιχεία της διωνυμικής λογιστικής παλινδρόμησης (binary logistic regression). Ακολούθως, παρουσιάζεται η διαδικασία ανάπτυξης μοντέλου και οι απαραίτητοι στατιστικοί έλεγχοι στους οποίους υποβάλλονται. Το κεφάλαιο ολοκληρώνεται με μία σύντομη αναφορά στα βήματα που ακολουθούνται για την επεξεργασία των δεδομένων στο ειδικό λογισμικό στατιστικής ανάλυσης (SPSS 21.0).

Στο **Κεφάλαιο 4** γίνεται παρουσίαση των διαδικασιών της συλλογής και επεξεργασίας των στοιχείων στα οποία στηρίχθηκε η Διπλωματική Εργασία. Αρχικά, γίνεται αναφορά στην πηγή προέλευσης των στοιχείων και στην ειδική επεξεργασία τους στο Microsoft Excel ώστε να αποκτήσουν την τελική τους μορφή. Ακολούθως περιγράφεται η κωδικοποίησή τους στο ειδικό λογισμικό στατιστικής ανάλυσης.

Το **Κεφάλαιο 5** είναι από τα σημαντικότερα της Διπλωματικής Εργασίας, καθώς περιλαμβάνει την αναλυτική περιγραφή της μεθοδολογίας που εφαρμόστηκε ως την εξαγωγή των τελικών αποτελεσμάτων. Αρχικά, περιγράφονται τα βήματα που ακολουθήθηκαν για την εφαρμογή της μεθοδολογίας και παρουσιάζεται η διαδικασία ανάπτυξης του μαθηματικού μοντέλου. Παρουσιάζονται τα δεδομένα εισόδου και εξόδου, με ιδιαίτερη έμφαση στους στατιστικούς ελέγχους αξιοπιστίας των αποτελεσμάτων, ενώ τα τελικά αποτελέσματα συνοδεύονται από τις αντίστοιχες μαθηματικές σχέσεις και την περιγραφή τους για την ευκολότερη κατανόησή τους.

Το **Κεφάλαιο 6** αποτελεί το τελευταίο κεφάλαιο της παρούσας Διπλωματικής Εργασίας. Σε αυτό περιλαμβάνονται τα συνολικά συμπεράσματα που προέκυψαν ύστερα από την ερμηνεία των μαθηματικών μοντέλων. Αποτελούν μία σύνθεση αρκετών ποσοτικοποιημένων στοιχείων σε συνδυασμό με τα επιμέρους

αποτελέσματα του προηγούμενου κεφαλαίου. Επιπρόσθετα, καταγράφονται προτάσεις για περαιτέρω έρευνα στο αντικείμενο της Διπλωματικής Εργασίας είτε με άλλες μεθόδους, είτε με εξέταση πρόσθετων παραμέτρων και μεταβλητών.

Στο **Κεφάλαιο 7** παρατίθεται ο κατάλογος των βιβλιογραφικών αναφορών. Ο κατάλογος αυτός περιλαμβάνει αναφορές που αφορούν τόσο σε έρευνες που παρουσιάστηκαν στα κεφάλαια της εισαγωγής και της βιβλιογραφικής ανασκόπησης, όσο και σε στατιστικές έννοιες και μεθόδους που αναλύθηκαν στο θεωρητικό υπόβαθρο.

2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

2.1 Γενικά

Το κεφάλαιο αυτό αφορά στη βιβλιογραφική ανασκόπηση και περιλαμβάνει έρευνες στον τομέα της οδικής ασφάλειας, το αντικείμενο και η μεθοδολογία των οποίων παρουσιάζει συνάφεια με την παρούσα διπλωματική εργασία. Συγκεκριμένα, παρουσιάζονται έρευνες που αναφέρονται στην κατανόηση της συμπεριφοράς των οδηγών αυτοκινήτων μέσω τεχνολογικών μηχανημάτων, προσαρμοσμένα στον εγκέφαλο του αυτοκινήτου. Για κάθε επιστημονική εργασία παρουσιάζεται σύντομη σύνοψη, με έμφαση στη μεθοδολογία η οποία ακολουθήθηκε και στα αποτελέσματα τα οποία προέκυψαν. Μέσω της ανασκόπησης των μεθοδολογιών των ερευνών αυτών, επιχειρήθηκε ο προσδιορισμός μιας κατάλληλης μεθόδου και κατάλληλων παραμέτρων για την αντιμετώπιση του αντικειμένου της διπλωματικής εργασίας.

2.2 Συναφείς Έρευνες

Στο υποκεφάλαιο αυτό παρουσιάζονται τα βασικά σημεία ερευνών οι οποίες σχετίζονται με τη μελέτη συμπεριφοράς οδηγών αυτοκινήτων, στα οχήματα των οποίων έχουν εγκατασταθεί ειδικά συστήματα. Οι έρευνες αυτές είναι δημοσιευμένες σε επιστημονικά περιοδικά και αποτελούν μέρος της διεθνούς βιβλιογραφίας γύρω από το συγκεκριμένο αντικείμενο της οδικής ασφάλειας. Σκοπός είναι η σύγκριση των αποτελεσμάτων της παρούσας εργασίας με άλλες εργασίες με παρεμφερές περιεχόμενο.

Οι έρευνες που παρατίθενται είναι οι εξής:

1. Ohta, Tohru, and Shouji Nakajima. "**Development of a driving data recorder.**" JSAE Review 15.3 (1994): 255-258.
2. Nadeem, Tamer, et al. "**TrafficView: traffic data dissemination using car-to-car communication.**" ACM SIGMOBILE Mobile Computing and Communications Review 8.3 (2004): 6-19.
3. Huang, Yueng-Hsiang, et al. "**Feedback by technology: Attitudes and opinions of truck drivers.**" Transportation Research Part F: Traffic Psychology and Behaviour 8.4 (2005): 277-297.
4. Toledo, Tomer, Oren Musicant, and Tsippy Lotan. "**In-vehicle data recorders for monitoring and feedback on drivers' behavior.**" *Transportation Research Part C: Emerging Technologies* 16.3 (2008): 320-331.
5. Farmer, Charles M., Bevan B. Kirley, and Anne T. McCartt. "**Effects of in-vehicle monitoring on the driving behavior of teenagers.**" Journal of Safety Research 41.1 (2010): 39-45.

6. Prato, Carlo Giacomo, et al. "**Modeling the behavior of novice young drivers during the first year after licensure.**" *Accident Analysis & Prevention* 42.2 (2010): 480-486.
7. Zaldivar, Jorge, et al. "**Providing accident detection in vehicular networks through OBD-II devices and Android-based smartphones.**" *Local Computer Networks (LCN)*, 2011 IEEE 36th Conference on. IEEE, 2011.
8. Vaiana, Rosolino, et al. "**Driving behavior and traffic safety: an acceleration-based safety evaluation procedure for smartphones.**" *Modern Applied Science* 8.1 (2014): p88.
9. Shichrur, Rachel, Adi Sarid, and Navah Z. Ratzon. "**Determining the sampling time frame for In-Vehicle Data Recorder measurement in assessing drivers.**" *Transportation research part C: emerging technologies* 42 (2014): 99-106.

2.2.1 Development of a driving data recorder (1994)

Η έρευνα αυτή αποτελεί μία από τις πρώτες προσπάθειες ανάπτυξης ενός οργάνου καταγραφής δεδομένων από τον εγκέφαλο του αυτοκινήτου. Ο σκοπός της είναι να παρουσιάσει μία γενική επισκόπηση του οργάνου αυτού (DDR) και να παραθέσει τρόπους χρησιμοποίησής του.

Βασικά στοιχεία:

- Το DDR είναι αρκετά μικρό και ελαφρύ ώστε να μην δημιουργεί πρόβλημα στο όχημα.
- Είναι, επίσης, σχεδιασμένο ώστε να μην καταναλώνει μεγάλη ποσότητα μπαταρίας από το αυτοκίνητο.
- Μπορεί να χρησιμοποιηθεί για 9.000 ώρες οδήγησης ή 100.000 χιλιόμετρα.
- Τα δεδομένα, τα οποία αποθηκεύονται αρχικά σε μία κάρτα μνήμης, μπορούν να αναλυθούν μέσω προσωπικού Η/Υ.

Υπάρχουν δύο τρόποι με τους οποίους μπορούν να εγγραφούν τα δεδομένα οδήγησης:

- a) Σύστημα που βασίζεται στο χρόνο, όπου τα δεδομένα τοποθετούνται σε διαδοχική σειρά
- b) Σύστημα που βασίζεται στη συχνότητα, όπου η τοποθέτηση των δεδομένων γίνεται με βάση είτε τη συχνότητα χρήσης είτε τη συχνότητα εμφάνισης.

Ανάλογα με τις εντολές εγγραφής, το DDR μπορεί να χρησιμοποιηθεί για να συλλέγει στοιχεία όπως ο συνολικός χρόνος οδήγησης, η συνολική απόσταση που διένυσε ο οδηγός, πόσες φορές άναψε ή έσβησε η μηχανή του αυτοκινήτου κ.ο.κ.

Εν κατακλείδι, το DDR σχεδιάστηκε για να καταγράψει τις συνθήκες οδήγησης ενός αυτοκινήτου σε κανονικές συνθήκες οδοστρώματος και να αναλύσει τη συμπεριφορά του οδηγού στο δρόμο. Επομένως, ήταν δυνατό να συσσωρευτούν ποσοτικά στοιχεία οδήγησης τα οποία θα καταδεικνύουν το προφίλ του οδηγού.

2.2.2 TrafficView: traffic data dissemination using car-to-car communication (2004)

Σε αυτήν την εργασία παρουσιάζεται το TrafficView, το οποίο αποτελεί μέρος του σχεδίου e-Road. Το σχέδιο αυτό είναι μια προσπάθεια να επιτευχθούν στόχοι, όπως η ασφαλής οδήγηση, ο έλεγχος της κυκλοφορίας ή η διάδοση επειγόντων μηνυμάτων παρέχοντας επεκτάσιμη υποδομή για επικοινωνία μεταξύ οχημάτων. Το TrafficView καθορίζει ένα πλαίσιο αποθήκευσης και διάδοσης πληροφοριών για τα οχήματα που βρίσκονται εν κινήσει. Ο σκοπός του είναι να ανταλλάσσονται πληροφορίες μεταξύ των αυτοκινήτων αναφορικά με την ταχύτητα και τις συντεταγμένες τους. Καθώς ένα όχημα κινείται, εισέρχεται στην εμβέλεια εκπομπής κάποιων οχημάτων και εξέρχεται από την εμβέλεια άλλων. Χρησιμοποιώντας αυτό το σύστημα, ο οδηγός θα μπορεί να γνωρίζει τον κυκλοφοριακό φόρτο τη δεδομένη χρονική στιγμή, κάτι το οποίο χρησιμεύει σε καταστάσεις κακοκαιρίας ή στην επιλογή εναλλακτικής διαδρομής για το ταξίδι του.

Η μεθοδολογία που ακολουθήθηκε αναφέρεται ως μηχανισμός διάδοσης. Κάθε όχημα εκπέμπει πληροφορίες για τον εαυτό του και για τα άλλα οχήματα τα οποία γνωρίζει, ενώ όταν λαμβάνει άλλες πληροφορίες, αναβαθμίζει τη μνήμη του και τις προωθεί στην επόμενη περίοδο εκπομπής.

Οι καταγραφές για τα οχήματα περιλαμβάνουν:

- ✓ Μοναδικό αριθμό αναγνώρισης
- ✓ Ακριβής θέση
- ✓ Ταχύτητα
- ✓ Ώρα εκπομπής

Η εργασία είχε ως κύριο θέμα, στη συνέχεια, την εύρεση του κατάλληλου αλγορίθμου για την επεξεργασία και ανάλυση όλων των δεδομένων που προέκυψαν από το πείραμα. Τα αποτελέσματα έδειξαν ότι ο αλγόριθμος Ratio-based είναι η καλύτερη επιλογή όσον αφορά τον έλεγχο και την ακρίβεια των παραμέτρων που χρησιμοποιούνται. Οι άλλες μέθοδοι προσφέρουν μεγαλύτερη ευκολία στην προσαρμογή των παραμέτρων αλλά το κόστος λειτουργίας τους τις καθιστά ασύμφωρες σε σχέση με την προαναφερθείσα.

2.2.3 Feedback by technology: Attitudes and opinions of truck drivers (2005)

Σκοπός της εργασίας αυτής ήταν να ερευνήσει την άποψη οδηγών φορτηγών αυτοκινήτων για την χρήση συστημάτων καταγραφής οδηγικής συμπεριφοράς που βρίσκονται εντός του οχήματος και παρέχουν πληροφορίες για την βελτίωση της οδικής ασφάλειας. Επιπλέον, είχε ως στόχο να κατανοηθεί ποιος είναι ο καλύτερος τρόπος να λαμβάνουν οι οδηγοί τις πληροφορίες αυτές.

Η έρευνα επικεντρώθηκε στους οδηγούς μεγάλων αποστάσεων, καθώς βρίσκονται στο δρόμο περισσότερη ώρα από τους υπόλοιπους οδηγούς φορτηγών αυτοκινήτων και θα επωφεληθούν περισσότερο από την εξέλιξη της τεχνολογίας. Συνολικά, ερωτήθηκαν 198 οδηγοί και 66 άνθρωποι συμμετείχαν στην συλλογή των ποιοτικών δεδομένων σχετικά με τη στάση τους ως προς την τεχνολογία και την ανατροφοδότηση των πληροφοριών. Τα ερωτηματολόγια που μοιράστηκαν στους οδηγούς επικεντρώθηκαν σε 6 βασικές κατηγορίες ερωτημάτων:

- i. Θα ήθελαν οι οδηγοί να λαμβάνουν περισσότερες πληροφορίες σχετικά με την συμπεριφορά τους στο δρόμο? Επιπλέον, θα προτιμούσαν οι πληροφορίες να έχουν θετική ή αρνητική χροιά?
- ii. Ποια είναι τα πλεονεκτήματα της ανατροφοδότησης μέσω της τεχνολογίας?
- iii. Ποια είναι τα μειονεκτήματα?
- iv. Ποια είναι η προτιμότερη μορφή λήψης πληροφοριών σχετικά με την οδηγική συμπεριφορά? Επίσης, πόσο συχνά θα ήθελαν να ενημερώνονται?
- v. Από ποιον θα ήθελαν οι οδηγοί να λαμβάνουν ανατροφοδότηση?
- vi. Ποιες είναι οι πιο σημαντικές συμπεριφορές ασφαλούς οδήγησης?

Τα αποτελέσματα κατέδειξαν ότι οι οδηγοί μεγάλων αποστάσεων επιθυμούν να λαμβάνουν περισσότερες θετικές πληροφορίες σχετικά με την συμπεριφορά τους στο δρόμο. Παρόλα αυτά, θα προτιμούσαν αυτές οι πληροφορίες να προέρχονται από κάποιους επιβλέποντες ή managers παρά από τεχνολογικά μέσα. Όσον αφορά την μορφή ή τη συχνότητα της ανατροφοδότησης, δεν επέδειξαν κάποια ισχυρή προτίμηση σε μια συγκεκριμένη απάντηση ενώ τόνισαν τη σημασία της προσαρμοστικότητας ενός τεχνολογικού προγράμματος στις απαιτήσεις του εκάστοτε οδηγού.

2.2.4 In-vehicle data recorders for monitoring and feedback on drivers' behavior (2008)

Το αντικείμενο της εργασίας αυτής είναι η διερεύνηση της συμπεριφοράς των οδηγών μέσω ενός συστήματος καταγραφής δεδομένων, ονόματι DriveDiagnostics, το οποίο είναι προσαρμοσμένο στο εσωτερικό του αυτοκινήτου, έχει διαστάσεις 11x6x3 εκατοστά και φορτίζεται από την μπαταρία του οχήματος.

Η μέθοδος της έρευνας ήταν η εξής: Το σύστημα συλλέγει στοιχεία όπως η επιτάχυνση του οχήματος (κατά x, y), η ταχύτητά του, η ακριβής θέση του μέσω συντεταγμένων με τη βοήθεια του GPS, η κατανάλωση καυσίμων, ο συνολικός χρόνος ταξιδιού κτλ. Σε αντίθεση με άλλα παρόμοια συστήματα, η μετάδοση των πληροφοριών μέσω του DriveDiagnostics γίνεται συνεχώς και σε πραγματικό χρόνο, παρόλο που το όχημα βρίσκεται εν κινήσει. Σκοπός της συγκεκριμένης εργασίας είναι να εκτιμήσει τη συμπεριφορά των οδηγών, καταμετρώντας τον τρόπο οδήγησής τους μέσω συγκεκριμένων παραμέτρων (συχνότητα αλλαγής λωρίδων, απότομα φρεναρίσματα και επιταχύνσεις, υπερβολική ταχύτητα). Όλες αυτές οι πληροφορίες μεταφέρονται μέσω ασύρματων δικτύων σε έναν διακομιστή, ο οποίος τις αποθηκεύει ώστε για κάθε οδηγό να υπάρχει ένας συγκεκριμένος φάκελος με τα χαρακτηριστικά του (αριθμός οχήματος, μέση κατανάλωση καυσίμων, συμμετοχή σε οδικά ατυχήματα κτλ.). Τέλος, οι οδηγοί κατατάσσονται σε 3 κατηγορίες (προσεκτικοί, κανονικοί και επιθετικοί), ανάλογα με το προφίλ που προκύπτει από τα δεδομένα.

Τα αποτελέσματα έδειξαν ότι η αρχική έκθεση των συμμετεχόντων στο πείραμα είχε σημαντική επίδραση στη βελτίωση της οδηγικής τους συμπεριφοράς και της οδικής ασφάλειας. Επιπλέον, η πρόσβαση που είχαν στα δεδομένα του συστήματος τους βοήθησε να καταλάβουν τα σφάλματα που διαπράττουν στο δρόμο. Παρόλα αυτά, η έρευνα έδειξε ότι αν δεν υπάρξει περαιτέρω ενδιαφέρον για την καταγραφή της οδηγικής συμπεριφοράς των ανθρώπων, μετά από ένα μικρό χρονικό διάστημα δεν παρατηρείται καμία πρόοδος. Στην εν λόγω εργασία, η επίδραση του DriveDiagnostics και στους 33 οδηγούς που χρησιμοποιήθηκε είχε εξαφανιστεί με το πέρας 5 μηνών.

2.2.5 Effects of in-vehicle monitoring on the driving behavior of teenagers (2010)

Ο κύριος σκοπός αυτής της εργασίας είναι να προσδιορίσει αν η οδηγική συμπεριφορά των εφήβων βελτιώνεται με την τοποθέτηση μιας συσκευής παρακολούθησης και ανατροφοδότησης πληροφοριών μέσα στο όχημά τους.

Για τις ανάγκες της έρευνας, χρησιμοποιήθηκαν 85 οδηγοί ηλικίας 16 και 17 χρόνων, οι οποίοι χωρίστηκαν τυχαία σε 4 ομάδες, με τη διαφορά να έγκειται στο αν θα ειδοποιούνται μέσω ηχητικών μηνυμάτων για τυχόν υπέρβαση συγκεκριμένων ορίων και αν θα ενημερώνονται ηλεκτρονικά οι γονείς τους για αυτές τις παραβάσεις. Η συσκευή μπορούσε να ανιχνεύσει όλα τα περιστατικά απότομης επιτάχυνσης ή επιβράδυνσης, υπερβολικής ταχύτητας και μη χρήσης της ζώνης ασφαλείας. Κάθε φορά που ο οδηγός πραγματοποιούσε μία απότομη αλλαγή ταχύτητας ή δεν φορούσε τη ζώνη του, ένα σήμα στελνόταν σε κεντρικό ηλεκτρονικό υπολογιστή καταγράφοντας την εκάστοτε παράβαση. Όσον αφορά τα όρια ταχύτητας, μέσω του GPS ήταν δυνατό να εντοπιστεί η ακριβής θέση του οχήματος και να συγκριθεί η ταχύτητά του με τα όρια ταχύτητας του συγκεκριμένου τμήματος της οδού όπου βρισκόταν. Τα δεδομένα αυτά, επίσης, αποθηκεύονταν στον ίδιο κεντρικό υπολογιστή. Όσοι οδηγοί είχαν ενταχθεί στις ομάδες 1 και 2, είχαν ηλεκτρονική ειδοποίηση μέσω χαρακτηριστικών ήχων κάθε φορά που είχαν απότομη συμπεριφορά ή δεν φορούσαν ζώνη ασφαλείας.

Τα αποτελέσματα ήταν ιδιαίτερα ενθαρρυντικά για τη χρήση της ζώνης ασφαλείας, καθώς το 94% των συμμετεχόντων την φορούσε από τη στιγμή που επιβιβαζόταν στο αυτοκίνητο, ενώ το υπόλοιπο 6% συμμορφωνόταν με τον κώδικα οδικής κυκλοφορίας όταν άκουγε τον χαρακτηριστικό ήχο. Οι οδηγοί που ενημερωνόντουσαν μέσω ηχητικών μηνυμάτων για υπέρβαση του ορίου ταχύτητας προσάρμοζαν άμεσα τη συμπεριφορά τους καθώς δήλωσαν ότι τους ενοχλούσε ο ήχος. Σε όσους, όμως, δεν ερχόταν ειδοποίηση στο όχημά τους και απλά μάθαιναν οι γονείς τους για την παράβαση του ορίου ταχύτητας, δεν παρατηρήθηκε κάποια συμμόρφωση ως προς τους κανόνες ασφαλείας. Η συμπεριφορά των οδηγών με απότομη επιτάχυνση ή επιβράδυνση δεν βελτιωνόταν σημαντικά με την εκπομπή του ηχητικού μηνύματος παρά μόνο στις περιπτώσεις όπου οι γονείς τους μάθαιναν για αυτές τις παραβάσεις.

2.2.6 Modeling the behavior of novice young drivers during the first year after licensure (2010)

Οι αρχάριοι οδηγοί συχνά εμφανίζονται ευάλωτοι μπροστά στον κίνδυνο ατυχήματος, κάτι το οποίο μεταφράζεται σε μεγάλη συμμετοχή τους στον κατάλογο των οδικών ατυχημάτων. Σκοπός της συγκεκριμένης έρευνας είναι η ανάλυση της συμπεριφοράς των νέων οδηγών με τη χρήση ενός συστήματος καταγραφής δεδομένων, το οποίο βρίσκεται εντός του οχήματος.

Για τις ανάγκες της εργασίας εξετάστηκαν 62 οδηγοί (36 άνδρες – 26 γυναίκες) ηλικίας 17 χρονών κατά την περίοδο των πρώτων 12 μηνών μετά την απόκτηση της άδειας οδήγησης. Για τους 3 πρώτους μήνες, ήταν απαραίτητο να υπάρχει στο αυτοκίνητο ένας γονιός κατά τη διάρκεια της οδήγησης, ενώ μετά το πέρας αυτού του χρονικού διαστήματος, μπορούσαν να οδηγούν μόνοι. Το σύστημα IVDR (In-Vehicle Data Recorder) κατέγραφε δεδομένα όπως η ταχύτητα και η επιτάχυνση του οχήματος ώστε να μπορεί να διακρίνει πότε ο οδηγός πραγματοποιεί επικίνδυνους ελιγμούς, οι οποίοι κατηγοριοποιούνταν ανάλογα με το είδος τους (αλλαγή λωρίδας, απότομο φρενάρισμα ή επιτάχυνση κτλ.) και τη σοβαρότητά τους (χαμηλή, μέση, υψηλή). Όλες αυτές οι πληροφορίες μεταδίδονταν μέσω ασύρματων δικτύων σε έναν διακομιστή, ο οποίος διατηρούσε αρχείο για κάθε ένα από τα εξεταζόμενα οχήματα, δημιουργώντας έτσι το προφίλ του κάθε αρχάριου οδηγού. Οι γονείς και οι νεαροί οδηγοί ήταν σε θέση να ελέγξουν τα δεδομένα αυτά μέσω μιας εφαρμογής που τους επέτρεπε να έχουν πρόσβαση στο προφίλ τους.

Όπως ήταν αναμενόμενο, τα αποτελέσματα επιβεβαίωσαν ότι οι άνδρες οδηγοί είναι πιο επιρρεπείς στον κίνδυνο και ότι όταν οδηγούν χωρίς επίβλεψη, οι αρχάριοι οδηγοί γίνονται πιο «επιθετικοί». Όσον αφορά την οδήγηση με παρόντες τους γονείς, ο κίνδυνος ατυχήματος αποτελεί συνάρτηση και της οδηγικής συμπεριφοράς του ενήλικα. Τέλος, σημαντικό ρόλο στην μείωση της επικινδυνότητας των νέων οδηγών παίζει και ο έλεγχος των δεδομένων που προκύπτουν από το IVDR. Οι νεαροί οι οποίοι ελέγχονται ηλεκτρονικά από τους γονείς τους καθ' όλη τη διάρκεια της έρευνας γίνονται πιο προσεκτικοί όταν οδηγούν μόνοι τους, ενώ αντίθετα όσοι δεν έχουν την συνεχή επίβλεψη των γονέων μέσω της ηλεκτρονικής εφαρμογής, δεν δείχνουν διάθεση να βελτιώσουν την τάση που έχουν για επικίνδυνη οδήγηση.

2.2.7 Providing accident detection in vehicular networks through OBD-II devices and Android-based smartphones (2011)

Σκοπός της έρευνας αυτής είναι η ανάπτυξη μιας εφαρμογής για κινητά τηλεφώνά με λογισμικό Android η οποία παρακολουθεί το όχημα μέσω του συστήματος OBD-II (On Board Diagnostics) και ανιχνεύει οδικά ατυχήματα.

Το σύστημα OBD αναπτύχθηκε στις Ηνωμένες Πολιτείες Αμερικής και είχε ως στόχο την ανίχνευση μηχανικών προβλημάτων στο όχημα, τα οποία προκαλούσαν υψηλά επίπεδα εκπομπής ρύπων, πάνω από τα αποδεκτά όρια. Για να το επιτύχει αυτό, το σύστημα είναι συνδεδεμένο με την μηχανή του αυτοκινήτου και όταν ανιχνεύσει ένα πρόβλημα, το αποθηκεύει στη μνήμη του ώστε να επιδιορθωθεί αργότερα από τους ειδικούς. Η μέθοδος της εργασίας αυτής στηρίχθηκε κατά ένα σκέλος στο συγκεκριμένο σύστημα. Το κινητό τηλέφωνο συνδέεται μέσω Bluetooth με το μηχανήμα OBD-II και λαμβάνει πληροφορίες σχετικά με την κατάσταση του οχήματος. Επιπλέον, μέσω GPS, γνωρίζει την ταχύτητα και την ακριβή τοποθεσία του. Για την ανίχνευση του οδικού ατυχήματος, 2 παράμετροι έχουν ληφθεί υπόψιν. Η μία αφορά την ενεργοποίηση του συστήματος των αερόσακων και η άλλη την επιτάχυνση του αυτοκινήτου, όπου πολύ υψηλές τιμές του G λαμβάνουν χώρα μόνον όταν σοβαρά ατυχήματα συμβαίνουν, καθώς ούτε τα μονοθέσια της Formula 1 μπορούν να πιάσουν τέτοιες τιμές. Επομένως, μόλις ανιχνευθεί ότι ένα ατύχημα έχει γίνει, αμέσως ειδοποιείται είτε με γραπτό μήνυμα είτε με email ένα άτομο της επιλογής του οδηγού, το οποίο έχει αποφασιστεί από την αρχή της έρευνας, και γίνεται κλήση στην άμεση βοήθεια και την αστυνομία. Οι συντεταγμένες του οχήματος που λαμβάνονται από το GPS είναι απαραίτητες για την όσο το δυνατόν πιο γρήγορη άφιξη του ασθενοφόρου.

Τα αποτελέσματα της έρευνας ήταν άκρως ενθαρρυντικά όσον αφορά την οδική ασφάλεια και τον χρόνο διακομιδής των εμπλεκόμενων σε ατύχημα στο κοντινότερο νοσοκομείο. Ο χρόνος ειδοποίησης από την ώρα που συνέβη το ατύχημα υπολογίστηκε στα 3 δευτερόλεπτα, με τον μέγιστο χρόνο να φτάνει τα 6 δευτερόλεπτα. Η διαφορά έγκειται στην απόδοση σύνδεσης του κινητού τηλεφώνου στο διαδίκτυο.

2.2.8 Driving behavior and traffic safety: an acceleration-based safety evaluation procedure for smartphones (2014)

Αυτή η εργασία πραγματοεύεται την ανάπτυξη μιας πρωτότυπης εφαρμογής για ηλεκτρονικές, κινητές συσκευές η οποία θα εκτιμάει τον βαθμό ασφάλειας των οδηγών που βρίσκονται εν κινήσει μετρώντας την διαμήκη και πλευρική τους επιτάχυνση, μέσω χρήσης των συστημάτων GPS . Όπου απαιτείτο, η εφαρμογή θα εμφανίζει μια προειδοποίηση στους οδηγούς ώστε να συμμορφώσουν την συμπεριφορά τους στον δρόμο.

Η μεθοδολογία που ακολουθήθηκε ήταν η εξής: Τοποθετώντας σε έναν άξονα x, y τις επιταχύνσεις του οχήματος (πλευρική και διαμήκης αντίστοιχα), εκτιμήθηκε η συμπεριφορά του οδηγού στο τιμόνι (επιθετικός ή μη). Χρησιμοποιήθηκε, επίσης, ο κύκλος τριβής του οχήματος, ο οποίος είναι συνάρτηση των χαρακτηριστικών των ελαστικών του αυτοκινήτου και των χαρακτηριστικών του οδοστρώματος. Λαμβάνοντας, επιπλέον, υπόψη την εμπειρία του οδηγού και τον τύπο του αυτοκινήτου δημιουργήθηκε το “Διάγραμμα Οδηγικής Συμπεριφοράς” (Driving Style Diagram – DSD), το οποίο αποτελεί μία συσχέτιση όλων των παραμέτρων που αναφέρθηκαν παραπάνω. Το πείραμα πραγματοποιήθηκε με το ίδιο όχημα και το ίδιο smartphone και πήραν μέρος 5 οδηγοί με διαφορετικά οδηγικά χαρακτηριστικά. Η συμπεριφορά κάθε οδηγού εκτιμήθηκε με βάση το ποσοστό υπέρβασης των ορίων επιτάχυνσης του DSD.

Η έρευνα είχε ως αποτέλεσμα ότι η καλύτερη τιμή για να διαχωρίσεις τους επιθετικούς με τους ασφαλείς οδηγούς μπορεί να οριστεί στο 9% ως ποσοστό υπέρβασης των ορίων επιτάχυνσης του DSD.

2.2.9 Determining the sampling time frame for In-Vehicle Data Recorder measurement in assessing drivers (2014)

Η παρούσα εργασία περιγράφει μία μέθοδο προσδιορισμού ενός κατάλληλου χρονικού πλαισίου ώστε το In-Vehicle Data Recorder να αποτελεί ένα αξιόπιστο εργαλείο αξιολόγησης για την ανάλυση προτύπων οδήγησης και αλλαγών συμπεριφοράς των οδηγών σε βάθος χρόνου.

Η μέθοδος βασίστηκε σε 64 οδηγούς ταξί, στα οχήματα των οποίων εγκαταστάθηκε μία συσκευή καταγραφής δεδομένων η οποία αποθήκευε στοιχεία όπως η ταχύτητα, η θέση, οριζόντια και κατακόρυφη επιτάχυνση και ελιγμούς. Επιπλέον, στατιστικές αναλύσεις διενεργήθηκαν για τον προσδιορισμό του χρόνου κατά τον οποίο το ποσοστό των ανεπιθύμητων συμβάντων οδήγησης σταθεροποιείται χρησιμοποιώντας, αρχικά, έναν δείκτη σύγκλισης, μέγεθος που μειώνεται καθώς τα δεδομένα συγκλίνουν, και το αναμενόμενο σφάλμα για διαφορετικούς χρόνους δειγματοληψίας, με σκοπό την σύγκριση του σφάλματος αυτού σε διαφορετικές χρονικές περιόδους.

Τα ευρήματα έδειξαν ότι οι 300 οδηγικές ώρες αποτελούν ένα ικανό και αξιόπιστο δείγμα ώστε να σχηματιστεί ένα σταθερό μοτίβο οδήγησης. Στην περίπτωση των οδηγών ταξί, οι 300 ώρες μεταφράζονται σε ενάμιση μήνα οδήγησης, θεωρώντας 7 ώρες καθημερινής ξεκούρασης. Δειγματοληψία μικρότερη των 100 ωρών δεν καταλήγει σε αξιόπιστο μέτρο αξιολόγησης οδηγικής συμπεριφοράς. Τέλος, το πλεονέκτημα του IVDR έγκειται στην μέτρηση ακριβών και αξιόπιστων δεδομένων, καταγράφοντας την συχνότητα των ανεπιθύμητων συμβάντων σε πραγματικό χρόνο.

2.3 Σύνοψη

Στο παρόν κεφάλαιο παρουσιάστηκαν και αναλύθηκαν ορισμένες από τις σημαντικότερες έρευνες που σχετίζονται με την εκτίμηση της συμπεριφοράς του οδηγού μέσω συστημάτων καταγραφής δεδομένων που βρίσκονται εντός του οχήματος, καθώς και τη συμβολή της χρήσης αυτών των νέων τεχνολογιών στην βελτίωση της οδικής ασφάλειας. Από την ανάλυση των παραπάνω ερευνών και την συνολική εξέτασή τους προκύπτουν τα παρακάτω κύρια σημεία:

- ✓ Για τη σωστή και αποτελεσματική λειτουργία των συστημάτων καταγραφής δεδομένων, απαιτείται **σύνδεση με Η/Υ** ώστε να αποθηκεύονται και να επεξεργάζονται άμεσα οι πληροφορίες.
- ✓ Οι πληροφορίες αυτές σχετίζονται με τον **τρόπο οδήγησης** του οδηγού (ταχύτητα, επιτάχυνση, επιβράδυνση, απόσταση με τα υπόλοιπα αυτοκίνητα κτλ) καθώς και με τα **μηχανικά χαρακτηριστικά** του οχήματος (κατανάλωση καυσίμων, πίεση ελαστικών κοκ).
- ✓ Είτε πρόκειται για νέους είτε για έμπειρους οδηγούς, όταν υπάρχει **ηχητική ειδοποίηση εντός του οχήματος** σχετικά με κάποια πιθανή παράβαση στην οποία έχουν υποπέσει, εμφανίζεται η τάση να βελτιώνουν άμεσα την οδηγική τους συμπεριφορά.
- ✓ Αντιθέτως, όταν αυτή η παράβαση απλώς καταγράφεται στα αρχεία του Η/Υ και δεν ενημερώνεται ο οδηγός, δεν παρατηρήθηκε κάποια βελτίωση στην οδική ασφάλεια.
- ✓ Σε βάθος χρόνου, **σχεδόν κανένας** οδηγός δεν ανατρέχει στο αρχείο οδήγησής του ώστε να ελέγξει τις οδηγικές παραβάσεις τις οποίες έχει διαπράξει, καθιστώντας αυτήν τη δυνατότητα του συστήματος σχεδόν ανούσια.
- ✓ Όσον αφορά οδηγούς οι οποίοι δεν είναι απόλυτα εξοικειωμένοι με την τεχνολογία, εμφανίζεται η προτίμησή τους στη λήψη πληροφοριών σχετικά με την οδηγική τους συμπεριφορά μέσω άλλων ανθρώπων παρά μέσω των συστημάτων καταγραφής δεδομένων.
- ✓ Τα συστήματα αυτά μπορούν να συμβάλλουν όχι μόνο στην βελτίωση της οδικής ασφάλειας και της συμπεριφοράς των οδηγών στο δρόμο, αλλά και στην **έγκαιρη πρόληψη και αντιμετώπιση οδικών ατυχημάτων**.
- ✓ Η προσπάθεια εκτίμησης της συμπεριφοράς του οδηγού μέσω τεχνολογικά προηγμένων συστημάτων καταγραφής έχει ξεκινήσει πριν από περίπου 30 χρόνια και συνεχώς **εξελίσσεται και βελτιώνεται**.
- ✓ Τέλος, είναι προφανές ότι υπάρχει **κενό γνώσης** γύρω από το εν λόγω θέμα, ωστόσο η εξέλιξη της τεχνολογίας και κατ' επέκταση των συστημάτων καταγραφής δεδομένων θα επιφέρει ακόμα πιο σημαντικά αποτελέσματα στον τομέα της οδικής ασφάλειας και βελτίωσης της συμπεριφοράς των οδηγών στο δρόμο.

3. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

3.1 Εισαγωγή

Στο κεφάλαιο αυτό, παρουσιάζεται η θεωρία στην οποία βασίζεται η στατιστική ανάλυση της Διπλωματικής Εργασίας. Η μέθοδος που επιλέχθηκε για την ανάλυση των στατιστικών στοιχείων της εργασίας ήταν η διωνυμική λογιστική παλινδρόμηση (binary logistic regression). Ο κύριος λόγος για τον οποίο επιλέχθηκε η συγκεκριμένη μέθοδος είναι ότι οι μεταβλητές που εξετάζονται είναι διακριτές και λαμβάνουν μόνο δύο τιμές (0 και 1). Ωστόσο, στην πορεία αναζήτησης μιας καλύτερης συσχέτισης μεταξύ των μεταβλητών επιλέχθηκε για την ανάλυση των στοιχείων και η γραμμική παλινδρόμηση (linear regression). Ο κύριος λόγος στον οποίο στηρίχθηκε η επιλογή της μεθόδου αυτής συνίσταται στο ότι μία από τις εξαρτημένες μεταβλητές του προβλήματος (ταχύτητα οχήματος) λαμβάνει συνεχείς τιμές.

Στη συνέχεια του κεφαλαίου, αναλύονται τα θεωρητικά στοιχεία που αφορούν τη διωνυμική λογιστική παλινδρόμηση και τη γραμμική παλινδρόμηση, εντοπίζονται τα κριτήρια αποδοχής ενός μαθηματικού προτύπου ενώ παρουσιάζεται και ο τρόπος που επιτεύχθηκαν τα μαθηματικά μοντέλα πρόβλεψης μέσω λειτουργιών του ειδικού στατιστικού λογισμικού. Μεγάλο τμήμα των θεωρητικών στοιχείων αυτών έχει αναπτυχθεί και σε προηγούμενες διπλωματικές.

3.2 Βασικές Έννοιες Στατιστικής

Ο όρος **πληθυσμός** (population) αναφέρεται στο σύνολο των παρατηρήσεων του χαρακτηριστικού που ενδιαφέρει τη στατιστική έρευνα. Πρόκειται για ένα σύνολο στοιχείων που είναι τελειώς καθορισμένα. Ένας πληθυσμός μπορεί να είναι πραγματικός ή θεωρητικός.

Ο όρος **δείγμα** (sample) αναφέρεται σε ένα υποσύνολο του πληθυσμού. Οι περισσότερες στατιστικές έρευνες στηρίζονται σε δείγματα, αφού οι ιδιότητες του πληθυσμού είναι συνήθως αδύνατο να καταγραφούν. Όλα τα στοιχεία που ανήκουν στο δείγμα ανήκουν και στον πληθυσμό, χωρίς απαραίτητα να ισχύει το αντίστροφο. Τα συμπεράσματα που θα προκύψουν από την μελέτη του δείγματος θα ισχύουν με ικανοποιητική ακρίβεια για ολόκληρο τον πληθυσμό μόνο εάν το δείγμα είναι αντιπροσωπευτικό του πληθυσμού.

Με τον όρο **μεταβλητές** (variables) ορίζονται τα χαρακτηριστικά που ενδιαφέρουν να μετρηθούν και να καταγραφούν σε ένα σύνολο ατόμων. Οι μεταβλητές διακρίνονται στις παρακάτω κατηγορίες:

- Ποιοτικές μεταβλητές** (qualitative variables): Είναι οι μεταβλητές των οποίων οι δυνατές τιμές είναι κατηγορίες διαφορετικές μεταξύ τους. Η χρήση αριθμών για την παράσταση των τιμών μιας τέτοιας μεταβλητής είναι καθαρά συμβολική και δεν έχει την έννοια της μέτρησης.

b) **Ποσοτικές μεταβλητές** (quantitative variables). Είναι οι μεταβλητές με τιμές αριθμούς, που όμως έχουν τη σημασία της μέτρησης. Η ηλικία και ο αριθμός παιδιών μιας οικογένειας συνιστούν τέτοιες μεταβλητές. Οι ποσοτικές μεταβλητές διακρίνονται με τη σειρά τους σε δύο μεγάλες κατηγορίες τις διακριτές (ή ασυνεχείς) και τις συνεχείς.

- Σε μία **διακριτή** μεταβλητή η μικρότερη μη μηδενική διαφορά που μπορούν να έχουν δύο τιμές της είναι σταθερή ποσότητα. Ένα τέτοιο παράδειγμα είναι ο αριθμός των μελών της οικογένειας.
- Αντίθετα, σε μία **συνεχή** μεταβλητή δύο τιμές μπορούν να διαφέρουν κατά οποιαδήποτε μικρή ποσότητα. Ως παράδειγμα αναφέρουμε την ηλικία, για την οποία η διαφορά ανάμεσα σε δύο τιμές θα μπορούσε να είναι χρόνια, μήνες, ημέρες, ώρες, λεπτά, δευτερόλεπτα. Στην πράξη, συνεχής θεωρείται μια μεταβλητή όταν μπορεί να πάρει όλες τις τιμές σε ένα διάστημα, διαφορετικά θεωρείται διακριτή.

Μέτρα κεντρικής τάσης (measures of central tendency): Σε περίπτωση ανάλυσης ενός δείγματος x_1, x_2, \dots, x_n η **μέση τιμή** υπολογίζεται σύμφωνα με τη σχέση:

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

Μέτρα διασποράς και μεταβλητότητας (measures of variability): Στην περίπτωση όπου τα δεδομένα αποτελούν ένα δείγμα, η **διακύμανση** συμβολίζεται με s^2 και ορίζεται ως:

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

όπου \bar{x} είναι ο δειγματικός μέσος, δηλαδή η μέση τιμή των παρατηρήσεων του δείγματος.

Η μαθηματική σχέση που δίνει την **τυπική απόκλιση** του δείγματος είναι:

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

Για την περίπτωση συμμετρικά κατανομημένου δείγματος δεδομένων σύμφωνα με έναν εμπειρικό κανόνα, προκύπτει ότι:

- το διάστημα $(-s, +s)$ περιέχει περίπου το 68% των δεδομένων
- το διάστημα $(-2s, +2s)$ περιέχει περίπου το 95% των δεδομένων
- το διάστημα $(-3s, +3s)$ περιέχει περίπου το 99% των δεδομένων

Η **συνδιακύμανση** (covariance of the two variables) αποτελεί ένα μέτρο της σχέσης μεταξύ δύο περιοχών δεδομένων και δίνεται από τη σχέση:

$$\text{cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Μέτρα αξιοπιστίας είναι το επίπεδο εμπιστοσύνης, που ορίζεται ως η αναλογία των περιπτώσεων που μια εκτίμηση θα είναι σωστή και το επίπεδο σημαντικότητας, το οποίο ορίζεται ως η αναλογία των περιπτώσεων που ένα συμπέρασμα είναι εσφαλμένο.

3.3 Συσχέτιση Μεταβλητών – Συντελεστής Συσχέτισης

Έστω X, Y δύο τυχαίες και συνεχείς μεταβλητές. Ο βαθμός της γραμμικής συσχέτισης των δύο μεταβλητών X και Y , οι οποίες έχουν διασπορά σ_X^2 και σ_Y^2 αντίστοιχα και συνδιασπορά $\sigma_{XY} = \text{cov}[X, Y]$ καθορίζεται με τον **συντελεστή συσχέτισης** (correlation coefficient), ο οποίος συμβολίζεται με ρ και ορίζεται ως:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Ο συντελεστής συσχέτισης ρ εκφράζει το βαθμό και τον τρόπο που οι δύο μεταβλητές συσχετίζονται. Δεν εξαρτάται από την μονάδα μέτρησης των X και Y και λαμβάνει τιμές στο διάστημα $[-1, +1]$. Τιμές κοντά στο $+1$ δηλώνουν ισχυρή θετική συσχέτιση, τιμές κοντά στο -1 δηλώνουν ισχυρή αρνητική συσχέτιση και τιμές κοντά στο 0 δηλώνουν γραμμική ανεξαρτησία των X και Y .

Η **εκτίμηση** του συντελεστή συσχέτισης ρ γίνεται με την αντικατάσταση στην παραπάνω εξίσωση της συνδιασποράς σ_{XY} και των διασπορών σ_X και σ_Y , από όπου προκύπτει τελικά η έκφραση της εκτιμήτριας r :

$$r(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{[\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}]}$$

3.4 Βασικές Κατανομές

Όπως είναι γνωστό από τη θεωρία της στατιστικής, για να μελετηθούν τα διάφορα στατιστικά μεγέθη πρέπει να είναι γνωστή η μορφή της κατανομής που ακολουθούν οι τιμές τους. Παρακάτω παρατίθενται οι σημαντικότερες στατιστικές κατανομές που χρησιμοποιούνται στην ανάλυση των οδικών ατυχημάτων.

3.4.1 Κανονική Κατανομή

Μία από τις πιο σημαντικές κατανομές πιθανότητας για συνεχείς μεταβλητές είναι η κανονική κατανομή ή κατανομή του Gauss. Η συνάρτηση πυκνότητας της κατανομής αυτής είναι:

$$f(x) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2s^2}}$$

όπου m και s είναι σταθερές ίσες με την μέση τιμή και την τυπική απόκλιση αντίστοιχα.

3.4.2 Κατανομή Poisson

Είναι γνωστό ότι η πιο κατάλληλη κατανομή για την περιγραφή τελείως τυχαίων διακριτών γεγονότων είναι η κατανομή Poisson. Μια τυχαία μεταβλητή X (όπως π.χ. το πλήθος των ατυχημάτων ή των νεκρών από οδικά ατυχήματα) θεωρείται ότι ακολουθεί κατανομή Poisson με παράμετρο λ ($\lambda > 0$), και γράφεται $X \sim P(\lambda)$, όταν έχει συνάρτηση μάζας πιθανότητας την:

$$f(x) = \frac{\mu^x e^{-\mu}}{x!}$$

όπου $x = 0, 1, 2, 3, \dots$ και $x! = x \cdot (x-1) \cdot \dots \cdot 3 \cdot 2 \cdot 1$

Η μέση τιμή και η διασπορά κατά Poisson είναι $E\{x\} = \mu$ και $\sigma^2\{x\} = \mu$ και είναι ίσες μεταξύ τους.

Η κατανομή Poisson αφορά στον αριθμό των “συμβάντων” σε ορισμένο χρονικό ή χωρικό διάστημα. Γενικά, ο αριθμός X των συμβάντων σε χρονικό (ή χωρικό) διάστημα t ακολουθεί την κατανομή Poisson αν **(α)** ο ρυθμός λ , έστω των συμβάντων είναι χρονικά σταθερός και **(β)** οι αριθμοί των συμβάντων σε ξένα διαστήματα αποτελούν ανεξάρτητα ενδεχόμενα (Κοκολάκης και Σπηλιώτης, 1999).

Η κατανομή Poisson είναι κατάλληλη για την ανάπτυξη μοντέλων που αφορούν φαινόμενα που εμφανίζονται σπάνια και των οποίων οι εμφανίσεις είναι ανεξάρτητες μεταξύ τους, δηλαδή η εμφάνιση του φαινομένου μια φορά δεν επηρεάζει την επόμενη.

Ο αριθμός των παθόντων είναι μία μεταβλητή, η οποία παρουσιάζει όμοιες ιδιότητες με την μεταβλητή του αριθμού των ατυχημάτων και γενικά υποστηρίζεται ότι τα οδικά ατυχήματα ακολουθούν συνήθως κατανομή Poisson (Charman 1971, Zahavi 1962) ή κανονική κατανομή (Hojati 2011).

3.4.3 Αρνητική Διωνυμική Κατανομή

Μία άλλη πολύ σημαντική κατανομή που χρησιμοποιείται στην οδική ασφάλεια είναι η αρνητική διωνυμική κατανομή. Η χρήση της κατανομής αυτής ενδείκνυται για περιπτώσεις όπου η διακύμανση των στοιχείων του δείγματος είναι μεγαλύτερη από τον μέσο όρο. Αυτό μπορεί να παρατηρηθεί σε φαινόμενα που εμφανίζουν περιοδικές μεταβολές (όπως για παράδειγμα αριθμός αφίξεων οχημάτων που αφορούν σε μικρά χρονικά διαστήματα (π.χ. 10 sec) σε κάποιο σημείο μετά από φωτεινό σηματοδότη. Μια τυχαία μεταβλητή X θεωρείται ότι ακολουθεί την αρνητική διωνυμική κατανομή με παραμέτρους k, p (k : θετικός ακέραιος, $0 < p < 1$), και γράφεται $X \sim NB(k, p)$, όταν έχει συνάρτηση μάζας πιθανότητας την:

$$P(x) = \binom{X + K - 1}{X} p^K (1 - p)^X$$

όπου $X = 0, 1, 2, 3, \dots$

Μία συνήθης πρακτική στον έλεγχο στατιστικών υποθέσεων, είναι ο υπολογισμός της τιμής της πιθανότητας p (probability-value ή p -value). Η πιθανότητα p είναι το μικρότερο επίπεδο σημαντικότητας α που οδηγεί στην απόρριψη της μηδενικής υπόθεσης H_0 έναντι της εναλλακτικής H_1 . Είναι μία σημαντική τιμή, διότι ποσοτικοποιεί τη στατιστική απόδειξη που υποστηρίζει την εναλλακτική υπόθεση. Γενικά, όσο πιο μικρή είναι η τιμή της πιθανότητας p , τόσο περισσότερες είναι οι αποδείξεις για την απόρριψη της μηδενικής υπόθεσης H_0 έναντι της εναλλακτικής H_1 . Εάν η τιμή p είναι μικρότερη ή ίση του επιπέδου σημαντικότητας α , τότε η μηδενική υπόθεση H_0 απορρίπτεται.

3.5 Μαθηματικά Πρότυπα

3.5.1 Γραμμική Παλινδρόμηση

Ο κλάδος της στατιστικής, ο οποίος εξετάζει τη σχέση μεταξύ δύο ή περισσότερων μεταβλητών ώστε να είναι δυνατή η πρόβλεψη της μίας από τις υπόλοιπες, ονομάζεται ανάλυση παλινδρόμησης (regression analysis).

Με τον όρο εξαρτημένη μεταβλητή εννοείται η μεταβλητή της οποίας η τιμή πρόκειται να προβλεφθεί, ενώ ο όρος ανεξάρτητη μεταβλητή αναφέρεται σε εκείνη την μεταβλητή, η οποία χρησιμοποιείται για την πρόβλεψη της εξαρτημένης μεταβλητής. Η ανεξάρτητη μεταβλητή δεν θεωρείται τυχαία αλλά παίρνει καθορισμένες τιμές. Η εξαρτημένη μεταβλητή αντίθετα, θεωρείται τυχαία και "καθοδηγείται" από την ανεξάρτητη μεταβλητή. Προκειμένου να προσδιοριστεί αν μια ανεξάρτητη μεταβλητή ή ένας συνδυασμός ανεξάρτητων μεταβλητών προκαλεί τη μεταβολή της εξαρτημένης μεταβλητής, κρίνεται απαραίτητη η ανάπτυξη μαθηματικών μοντέλων. Η ανάπτυξη ενός μαθηματικού μοντέλου αποτελεί μια στατιστική διαδικασία που συμβάλλει στην ανάπτυξη εξισώσεων που περιγράφουν

τη σχέση μεταξύ των ανεξάρτητων μεταβλητών και της εξαρτημένης. Επισημαίνεται ότι η επιλογή της μεθόδου ανάπτυξης ενός μοντέλου βασίζεται στο αν η εξαρτημένη μεταβλητή λαμβάνει συνεχείς ή διακριτές τιμές.

Στην περίπτωση που η **εξαρτημένη μεταβλητή είναι συνεχές μέγεθος** και ακολουθεί κανονική κατανομή χρησιμοποιείται η μέθοδος της γραμμικής παλινδρόμησης, της οποίας η πιο απλή περίπτωση είναι η **απλή γραμμική παλινδρόμηση** (simple linear regression). Στην απλή γραμμική παλινδρόμηση υπάρχει μόνο μία ανεξάρτητη μεταβλητή X και μία εξαρτημένη μεταβλητή Y , η οποία προσεγγίζεται ως μία γραμμική συνάρτηση του X . Η τιμή y_i της μεταβλητής Y , για κάθε τιμή x_i της μεταβλητής X , δίνεται από την σχέση:

$$y_i = a + \beta x_i + \varepsilon_i$$

Το πρόβλημα της παλινδρόμησης είναι η εύρεση των παραμέτρων a και β που εκφράζουν όσο το δυνατόν καλύτερα τη γραμμική εξάρτηση της εξαρτημένης μεταβλητής Y από την ανεξάρτητη μεταβλητή X . Κάθε ζεύγος τιμών (a, β) καθορίζει και μία διαφορετική γραμμική σχέση που εκφράζεται γεωμετρικά από ευθεία γραμμή και οι δύο παράμετροι ορίζονται ως εξής:

- Ο σταθερός όρος a είναι η τιμή του y για $x=0$
- Ο συντελεστής β του x είναι η κλίση (slope) της ευθείας ή αλλιώς ο **συντελεστής παλινδρόμησης** (regression coefficient). Εκφράζει την μεταβολή της μεταβλητής Y όταν η μεταβλητή X μεταβληθεί κατά μία μονάδα.

Η τυχαία μεταβλητή ε_i λέγεται **σφάλμα παλινδρόμησης** (regression error) και ορίζεται ως η διαφορά της y_i από τη δεσμευμένη μέση τιμή $E(Y | X = x_i)$ όπου $E(Y | X = x_i) = a + \beta x_i$. Για την ανάλυση της γραμμικής παλινδρόμησης γίνονται οι παρακάτω υποθέσεις:

- Η μεταβλητή X είναι ελεγχόμενη για το πρόβλημα υπό μελέτη, δηλαδή είναι γνωστές οι τιμές της χωρίς καμιά αμφιβολία.
- Η εξάρτηση της Y από τη X είναι γραμμική.
- Το σφάλμα παλινδρόμησης έχει μέση τιμή μηδέν για κάθε τιμή της X και η διασπορά του είναι σταθερή και δεν εξαρτάται από τη X , δηλαδή $E(\varepsilon_i) = 0$ και $\text{Var}(\varepsilon_i) = \sigma_\varepsilon^2$.

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

Σε περίπτωση που η εξαρτημένη μεταβλητή Ψ **εξαρτάται γραμμικά από περισσότερες από μια ανεξάρτητες μεταβλητές** X ($X_1, X_2, X_3, \dots, X_n$) τότε γίνεται αναφορά στην **πολλαπλή γραμμική παλινδρόμηση** (multiple linear regression). Η εξίσωση η οποία αποτυπώνει τη σχέση ανάμεσα στην εξαρτημένη και τις ανεξάρτητες μεταβλητές έχει τη γενικότερη μορφή:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni} + \varepsilon_i$$

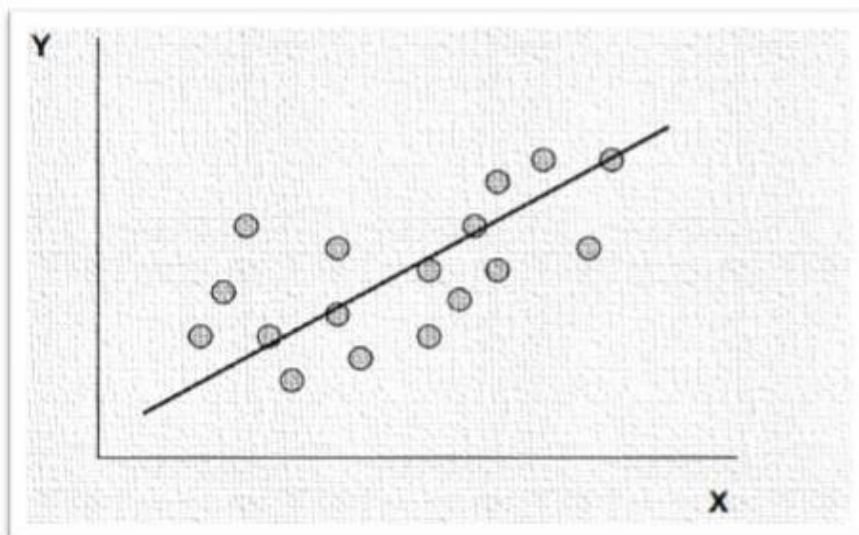
Οι υποθέσεις της πολλαπλής γραμμικής παλινδρόμησης είναι ίδιες με εκείνες της απλής γραμμικής παλινδρόμησης, δηλαδή υποθέτει κανείς ότι τα σφάλματα ε_i της παλινδρόμησης (όπως και η τυχαία μεταβλητή Y για κάθε τιμή της X) ακολουθούν κανονική κατανομή με σταθερή διασπορά. Γενικά το πρόβλημα και η εκτίμηση της πολλαπλής γραμμικής παλινδρόμησης δεν διαφέρει ουσιαστικά από εκείνο της απλής γραμμικής παλινδρόμησης. Ένα καινούριο στοιχείο στην πολλαπλή γραμμική παλινδρόμηση είναι ότι πριν προχωρήσει κανείς στην εκτίμηση των παραμέτρων πρέπει να ελέγξει αν πράγματι πρέπει να συμπεριληφθούν όλες οι ανεξάρτητες μεταβλητές στο μοντέλο. Εκείνο που απαιτείται να εξασφαλιστεί είναι η μηδενική συσχέτιση των ανεξάρτητων μεταβλητών, δηλαδή θα πρέπει να ισχύει:

$$\rho(X_i, X_j) \quad i \neq j \rightarrow 0$$

3.5.1.1 Μέθοδος Ελαχίστων Τετραγώνων

Η εκτίμηση των παραμέτρων του μοντέλου της πολλαπλής γραμμικής παλινδρόμησης γίνεται με τη μέθοδο ελαχίστων τετραγώνων (least squares method). Σύμφωνα με τη μέθοδο ελαχίστων τετραγώνων, ο προσδιορισμός των συντελεστών β_i δίνει μια προσεγγιστική ευθεία που συνδέει τις τιμές της μεταβλητής Y , δοθέντων των τιμών της μεταβλητής X .

Η ευθεία που προκύπτει ονομάζεται ευθεία παλινδρόμησης της Y πάνω στην X . Σκοπός είναι το άθροισμα των τετραγώνων των κατακόρυφων αποστάσεων των σημείων (X, Y) από την ευθεία να είναι ελάχιστο. Παρακάτω δίνεται ένα ενδεικτικό διάγραμμα της ευθείας ελαχίστων τετραγώνων.



Διάγραμμα 3.1: Ευθεία ελαχίστων τετραγώνων

3.5.2 Λογιστική Ανάλυση Παλινδρόμησης

Στα μοντέλα γραμμικής παλινδρόμησης που αναφέρθηκαν παραπάνω, ισχύει η προϋπόθεση ότι η εξαρτημένη μεταβλητή είναι συνεχής. Στην περίπτωση που η

εξαρτημένη μεταβλητή είναι διακριτή, εφαρμόζεται η **λογιστική ανάλυση παλινδρόμησης**. Η λογιστική ανάλυση παλινδρόμησης χρησιμοποιείται για τη δημιουργία μοντέλων πρόβλεψης και ταξινόμησης. Είναι δυνατή η έκβαση μιας κατηγορικής μεταβλητής με δύο κατηγορίες με τη χρήση ενός συνόλου συνεχών και διακριτών μεταβλητών. Σε αντίθεση με τη γραμμική παλινδρόμηση, η εξαρτημένη μεταβλητή είναι η πιθανότητα η έκβαση του αποτελέσματος να ισούται με 1. Χρησιμοποιείται ο νεπέριος λογάριθμος για την πιθανότητα ή τον λόγο πιθανοφάνειας (likelihood ratio), η εξαρτημένη μεταβλητή να είναι 1 σύμφωνα με τον παρακάτω τύπο:

$$Y = \text{logit}(P) = \ln\left(\frac{P_i}{1 - P_i}\right) = B_0 + B_i x_i$$

όπου B_0 : η σταθερά του μοντέλου

B_i : οι παραμετρικές εκτιμήτριες για τις ανεξάρτητες μεταβλητές ($x_i = 1, 2, \dots, n$) και

n : το σύνολο των εξαρτημένων μεταβλητών.

Η πιθανότητα κυμαίνεται από 0 έως 1, ενώ ο νεπέριος λογάριθμος $\ln\left(\frac{P_i}{1 - P_i}\right)$ λαμβάνει τιμές από $-\infty$ έως $+\infty$. Όταν οι πιθανές κατηγορίες της εξαρτημένης μεταβλητής είναι δύο, η ανάλυση ονομάζεται **διωνυμική** λογιστική παλινδρόμηση (binary logistic regression), ενώ σε περίπτωση πλήθους κατηγοριών περισσότερων των δύο χρησιμοποιείται η **πολυωνυμική** λογιστική παλινδρόμηση (multinomial logistic regression).

Η λογική της λογιστικής παλινδρόμησης είναι παρόμοια με εκείνη της γραμμικής (πολλαπλής) παλινδρόμησης, με τη διαφορά ότι επειδή η εξαρτημένη μεταβλητή είναι κατηγορική, δεν προβλέπονται τιμές της εξαρτημένης μεταβλητής, αλλά ταξινόμηση σε μία εκ των (δύο) κατηγοριών της (group membership).

Τα μοντέλα λογιστικής παλινδρόμησης υπολογίζουν την **καμπυλόγραμμη σχέση** ανάμεσα στην κατηγορική επιλογή Y και στις μεταβλητές X_i οι οποίες μπορεί να είναι συνεχείς ή διακριτές. Η καμπύλη της λογιστικής παλινδρόμησης είναι προσεγγιστικά γραμμική στις μεσαίες τιμές και λογαριθμική στις ακραίες τιμές. Με απλό μετασχηματισμό της παραπάνω σχέσης οδηγούμαστε στην εξής νέα σχέση:

$$\frac{P_i}{1 - P_i} = e^{B_0 + B_i x_i}$$

Η θεμελιώδης εξίσωση για την λογιστική παλινδρόμηση δείχνει ότι όταν μια τιμή μιας ανεξάρτητης μεταβλητής αυξηθεί κατά μία μονάδα και όλες οι υπόλοιπες παραμείνουν σταθερές, ο **νέος λόγος πιθανοφάνειας** $\frac{P_i}{1 - P_i}$ δίδεται από τη σχέση:

$$\left(\frac{P_i}{1 - P_i}\right) = e^{B_0 + B_i x_i + 1} = e^{B_0} e^{B_i} e^{B_i}$$

Έτσι παρατηρείται ότι όταν η εξαρτημένη μεταβλητή X_i αυξηθεί κατά μία μονάδα, με όλες τις υπόλοιπες μεταβλητές να παραμένουν σταθερές, η πιθανότητα αυξάνεται κατά ένα συντελεστή e^{B_i} .

3.5.3 Λογαριθμοκανονική Παλινδρόμηση

Μέσω της λογαριθμοκανονικής παλινδρόμησης (lognormal regression) δίνεται η δυνατότητα ανάπτυξης ενός μοντέλου που συσχετίζει δύο ή περισσότερες μεταβλητές. Η σχέση που συνδέει την εξαρτημένη με τις ανεξάρτητες μεταβλητές είναι γραμμική. Στη λογαριθμοκανονική παλινδρόμηση οι συντελεστές των μεταβλητών του μοντέλου είναι οι συντελεστές της γραμμικής παλινδρόμησης. Υπολογίζονται από την ανάλυση της παλινδρόμησης με βάση την αρχή των ελαχίστων τετραγώνων.

Η λογαριθμοκανονική παλινδρόμηση βασίζεται στην υπόθεση ότι τα στοιχεία που περιέχονται στη βάση δεδομένων είναι μη αρνητικά, ο φυσικός λογάριθμος της ανεξάρτητης μεταβλητής ακολουθεί την κανονική κατανομή και ο αριθμητικός μέσος είναι σχετικά μεγάλος. Η μαθηματική σχέση που περιγράφει τη μέθοδο αυτή είναι η εξής:

$$\log(y_i) = b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_kx_{ki} + \varepsilon_i$$

όπου y : η εξαρτημένη μεταβλητή

b_0, b_1, \dots, b_k : οι συντελεστές μερικής παλινδρόμησης

$x_{1i}, x_{2i}, \dots, x_{ki}$: οι εξαρτημένες μεταβλητές και

ε_i : το σφάλμα παλινδρόμησης.

3.6 Διαδικασία Ανάπτυξης και Κριτήρια Αποδοχής Μοντέλου

Όπως αναφέρθηκε και σε προηγούμενο εδάφιο, κάθε μοντέλο που αναπτύσσεται, για να θεωρηθεί αποδεκτό πρέπει να πληροί κάποιες βασικές προϋποθέσεις. Αρχικά πρέπει να ισχύει η **κανονικότητα**. Βάσει της προϋπόθεσης αυτής, απαιτείται οι τιμές της μεταβλητής να ακολουθούν κανονική κατανομή.

➤ Μη συσχέτιση των ανεξάρτητων μεταβλητών

Βασική προϋπόθεση είναι η μη συσχέτιση των ανεξάρτητων μεταβλητών. Οι ανεξάρτητες μεταβλητές πρέπει να είναι γραμμικώς ανεξάρτητες μεταξύ τους, δηλαδή να ισχύει $\rho(x_i, x_j) \forall i \neq j \rightarrow 0$ γιατί διαφορετικά δεν είναι δυνατή η εξακρίβωση της επιρροής της κάθε μεταβλητής στο αποτέλεσμα. Αν δηλαδή σε ένα μοντέλο εισάγονται δύο μεταβλητές που σχετίζονται αρκετά μεταξύ τους, εμφανίζονται προβλήματα μεροληψίας και επάρκειας.

➤ Λογική ερμηνεία των πρόσημων

Σημαντικό κριτήριο για την αποδοχή ενός μοντέλου μετά τη διαμόρφωση του είναι οι τιμές και τα πρόσημα των συντελεστών παλινδρόμησης β . Πρέπει αρχικά να υπάρχει λογική ερμηνεία των πρόσημων τους. Το θετικό πρόσημο του συντελεστή δηλώνει αύξηση της εξαρτημένης μεταβλητής με την αύξηση της ανεξάρτητης. Αντίθετα, αρνητικό πρόσημο επιφέρει μείωση της εξαρτημένης μεταβλητής με την αύξηση της ανεξάρτητης.

➤ **Κριτήριο λόγου πιθανοφάνειας (Likelihood Ratio Test-LRT)**

Σημαντικό ρόλο στην επιλογή των μεταβλητών των μοντέλων της λογιστικής παλινδρόμησης παίζει και η πιθανοφάνεια. Για την εκτίμηση της επιρροής των παραμέτρων β χρησιμοποιείται η μέθοδος της μέγιστης πιθανοφάνειας. Για να επιτευχθεί υψηλή πιθανοφάνεια επιχειρείται ο λογάριθμος των συναρτήσεων πιθανοφάνειας $L = -\log(\text{likelihood})$ να είναι όσο το δυνατόν μικρότερος, καθώς προτιμώνται μοντέλα με μικρότερο λογάριθμο της συνάρτησης πιθανοφάνειας. Μοντέλα που περιέχουν πολλές μεταβλητές είναι περισσότερο σύνθετα και χρειάζεται ένας κανόνας να αποφασίζει εάν η μείωση του λογαρίθμου της πιθανοφάνειας αξίζει την αυξημένη πολυπλοκότητα και για το σκοπό αυτό χρησιμοποιείται το Likelihood Ratio Test (LRT) ή αλλιώς κριτήριο λόγου πιθανοφάνειας.

Σύμφωνα με το συγκεκριμένο κριτήριο εάν ισχύει:

$$LRT = -2(L(b) - L(0)) > \chi_{b,0.05}^2$$

Όπου:

- $L(b)$ είναι ο λογάριθμος πιθανοφάνειας του μοντέλου με τις b μεταβλητές
- $L(0)$ είναι ο λογάριθμος πιθανοφάνειας του μοντέλου χωρίς τις b μεταβλητές
- $\chi_{b,0.05}^2$ η τιμή του κριτηρίου χ^2 για b βαθμούς ελευθερίας σε επίπεδο σημαντικότητας 5%

Το μοντέλο είναι στατιστικά προτιμότερο από το μοντέλο χωρίς τις μεταβλητές και γίνονται δεκτές οι μεταβλητές ως σημαντικές. Επισημαίνεται ότι οι διακριτές μεταβλητές με κατηγορίες k έχουν $k-1$ βαθμούς ελευθερίας, ενώ οι συνεχείς έχουν πάντοτε ένα βαθμό ελευθερίας.

➤ **Στατιστική αξιολόγηση των παραμέτρων**

Η στατιστική αξιολόγηση των παραμέτρων πραγματοποιείται μέσω του ελέγχου t -test (κριτήριο της κατανομής Student). Με τον τρόπο αυτό είναι δυνατό να διαπιστωθεί εάν οι παράμετροι που υπολογίστηκαν διαφέρουν σημαντικά από το 0, προσδιορίζεται η στατιστική σημαντικότητα των ανεξάρτητων μεταβλητών και καθορίζονται ποιες μεταβλητές τελικά θα συμπεριληφθούν στο τελικό μοντέλο. Ο συντελεστής t εκφράζεται από τη σχέση:

$$t_{\text{stat}} = \frac{\beta_i}{s.e.}$$

όπου $s.e.$: το τυπικό σφάλμα των σταθερών παραμέτρων (standard error)

Βάσει της παραπάνω σχέσης, όσο μειώνεται το τυπικό σφάλμα τόσο αυξάνεται η τιμή του t_{stat} και συνεπώς αυξάνεται η επάρκεια (efficiency). Όσο μεγαλύτερη είναι η τιμή του t_{stat} τόσο μεγαλύτερη είναι η επιρροή της συγκεκριμένης μεταβλητής στο τελικό αποτέλεσμα. Στον παρακάτω πίνακα παρουσιάζονται οι τιμές του t_{stat} για κάθε επίπεδο εμπιστοσύνης.

Πίνακας 3.1: Κρίσιμες τιμές του συντελεστή t

Βαθμός Ελευθερίας	Επίπεδο Εμπιστοσύνης				
	0,90	0,95	0,975	0,99	0,995
80	1,296	1,671	2,000	2,390	2,660
120	1,289	1,658	1,980	2,358	2,617
∞	1,282	1,645	1,960	2,326	2,576

Όπως φαίνεται από τον παραπάνω πίνακα για ένα δείγμα περί τα 80 και επίπεδο εμπιστοσύνης 95% είναι $t^* = 1,671$ και για επίπεδο εμπιστοσύνης 90% είναι $t^* = 1,3$. Έτσι αν για παράδειγμα μια μεταβλητή έχει τιμή t^* ίση με $-3,8$, η απόλυτη τιμή της τιμής t είναι $3,8$ δηλαδή μεγαλύτερη από $1,671$ και επομένως η μεταβλητή είναι αποδεκτή και στατιστικά σημαντική για το 95% των περιπτώσεων.

Όσον αφορά στα μοντέλα λογιστικής ανάλυσης παλινδρόμησης, ισχύει ό,τι και στην απλή γραμμική παλινδρόμηση, με τη διαφορά ότι στη λογιστική ανάλυση παλινδρόμησης το αντίστοιχο t -test έχει την ονομασία Wald. Η τιμή του Wald για κάθε μεταβλητή πρέπει να είναι μεγαλύτερη του $1,7$ όπως ακριβώς ισχύει και για τον συντελεστή t .

➤ Συντελεστής προσαρμογής R^2

Μετά τον έλεγχο στατιστικής εμπιστοσύνης εξετάζεται η ποιότητα του μοντέλου. Η ποιότητα του μοντέλου καθορίζεται με βάση τον συντελεστή προσαρμογής. Ο συντελεστής R^2 χρησιμοποιείται ως κριτήριο καλής προσαρμογής των δεδομένων στο γραμμικό μοντέλο και ορίζεται από τη σχέση:

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

$$\begin{aligned} \text{Όπου: } SSR &= \sum_{i=1}^n (y_i - \hat{y})^2 = \beta^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\ SST &= \sum_{i=1}^n (y_i - \bar{y})^2 \end{aligned}$$

Τα αρχικά SSR και SST έχουν προέλθει από τις φράσεις υπόλοιπο άθροισμα τετραγώνων (Residual Sum of Squares) και συνολικό άθροισμα τετραγώνων (Total Sum of Squares) αντίστοιχα. Με \hat{y} συμβολίζεται η προβλεπόμενη τιμή της εξαρτημένες μεταβλητές από τις ανεξάρτητες.

Ο συντελεστής αυτός εκφράζει το ποσοστό της μεταβλητότητας της μεταβλητής Y που εξηγείται από την μεταβλητή X . Λαμβάνει τιμές από 0 έως 1. Όσο πιο κοντά βρίσκεται η τιμή του R^2 στην μονάδα, τόσο πιο ισχυρή γίνεται η γραμμική σχέση εξάρτησης των μεταβλητών Y και X . Ο συντελεστής R^2 έχει συγκριτική αξία, κάτι το οποίο σημαίνει ότι δεν υπάρχει συγκεκριμένη τιμή του R^2 που είναι αποδεκτή ή απορριπτέα, αλλά μεταξύ δύο ή περισσότερων μοντέλων επιλέγεται ως καταλληλότερο εκείνο με τη μεγαλύτερη τιμή του συντελεστή R^2 . Ο συντελεστής R^2 μπορεί να χρησιμοποιηθεί ως μέτρο ισχυρότητας της γραμμικής σχέσης ανεξάρτητα από το αν το X παίρνει καθορισμένες τιμές ή αν είναι τυχαία μεταβλητή.

➤ Ελαστικότητα

Η ελαστικότητα αντικατοπτρίζει την **ευαισθησία** μιας εξαρτημένης μεταβλητής στην μεταβολή μιας ή περισσότερων ανεξάρτητων μεταβλητών. Πιο συγκεκριμένα η τιμή της ελαστικότητας ερμηνεύεται ως το ποσοστό επί της εκατό της μεταβολής της εξαρτημένης μεταβλητής που προκαλείται από μια μεταβολή της ανεξάρτητης μεταβλητής κατά 1%. Η ελαστικότητα, για γραμμικά πρότυπα, δίδεται από τη σχέση:

$$e_i = \frac{\Delta Y_i}{Y_i} \frac{X_i}{\Delta X_i} = \beta_i \frac{X_i}{Y_i}$$

Επισημαίνεται ότι η παραπάνω σχέση εφαρμόζεται αποκλειστικά σε **συνεχείς** μεταβλητές. Για **διακριτές** μεταβλητές χρησιμοποιείται η **έννοια της ψευδοελαστικότητας**, η οποία περιγράφει τη μεταβολή στην τιμή της πιθανότητας επιλογής κατά τη μετάβαση από τη μία τιμή της διακριτής μεταβλητής στην άλλη. Η σχέση που υπολογίζει την τιμή της ψευδοελαστικότητας για διακριτές μεταβλητές είναι η παρακάτω:

$$E_{x_{ink}}^{P_i} = e^{\beta_{ik}} \frac{\sum_{i=1}^I e^{\beta_i x_n}}{\sum_{i=1}^I e^{\Delta(\beta_i x_n)}} - 1$$

Όπου:

- I, το πλήθος των πιθανών επιλογών
- x_{ink} , η τιμή της μεταβλητής k για την εναλλακτική i του ατόμου n
- $\Delta(\beta_i x_n)$, η τιμή της συνάρτησης που καθορίζει την κάθε επιλογή αφού η τιμή της x_{nk} έχει μεταβληθεί από 0 σε 1
- $\beta_i x_n$, η αντίστοιχη τιμή όταν η x_{nk} έχει τιμή 0
- β_{ik} , η τιμή της παραμέτρου της μεταβλητής x_{nk}

Όσον αφορά στο **σφάλμα της εξίσωσης του μοντέλου**, πρέπει να πληρούνται κάποιες βασικές προϋποθέσεις:

- να ακολουθεί κανονική κατανομή
- να έχει σταθερή διασπορά, $\text{Var}(\varepsilon_i) = \sigma_\varepsilon^2 = c$
- να έχει μηδενική συσχέτιση, $\rho(x_i, x_j) = 0 \forall i \neq j$

Η **διασπορά του σφάλματος** εξαρτάται από τον συντελεστή R^2 . Όσο μεγαλύτερο είναι το R^2 τόσο μικρότερη είναι η διασπορά του σφάλματος, δηλαδή τόσο καλύτερη είναι η πρόβλεψη που βασίζεται στην ευθεία παλινδρόμησης.

3.7 Λειτουργία του Ειδικού Στατιστικού Λογισμικού

Η στατιστική ανάλυση των δεδομένων έγινε με τη χρήση ειδικού στατιστικού λογισμικού. Χρησιμοποιείται η εντολή **Analyze** για την ανάλυση αυτήν. Η εντολή περιλαμβάνει τις παρακάτω επιλογές:

- **Descriptive Statistics:** Διαδικασίες για την παραγωγή περιγραφικών αποτελεσμάτων. Εδώ βρίσκεται η επιλογή Options. Πρόκειται για χρήσιμες στατιστικές περιγραφικές συναρτήσεις (μέσος, τυπική απόκλιση, μέγιστο, ελάχιστο).

- **Correlate:** Η διαδικασία που μετράει τη συσχέτιση ανάμεσα σε ζευγάρια μεταβλητών. Από εδώ επιλέγεται η εντολή Bivariate correlations. Οι μεταβλητές που ενδιαφέρουν εισάγονται στο πλαίσιο Variables και χρησιμοποιείται ο συντελεστής συσχέτισης **Pearson** αν πρόκειται για συνεχείς μεταβλητές και ο συντελεστής συσχέτισης **Spearman** αν πρόκειται για διακριτές μεταβλητές.
- **Regression:** Η διαδικασία εκτελεί διάφορα είδη αναλύσεων παλινδρόμησης, μία εκ των οποίων είναι η λογιστική που επιλέξαμε για την ανάλυση των δεδομένων μας. Λόγω της φύσης των εξαρτημένων μεταβλητών, επιλέχθηκε η διωνυμική λογιστική παλινδρόμηση (binary logistic). Η μεταβλητή που ενδιαφέρει (εξαρτημένη μεταβλητή) εισάγεται στο πλαίσιο Dependent. Οι επεξηγηματικές μεταβλητές με τις οποίες θα εξηγηθεί η μεταβλητότητα της εξαρτημένης μεταβλητής, εισάγονται στο πλαίσιο Factor(s) εάν είναι διακριτές ή στο πλαίσιο Covariate(s) εάν είναι συνεχείς. Στο πλαίσιο Method μπορεί να επιλεγεί μια μέθοδος για τη βέλτιστη επιλογή επεξηγηματικών μεταβλητών. Αυτή συνήθως αφήνεται Enter που σημαίνει ότι στο μοντέλο εισέρχονται όσες μεταβλητές βρίσκονται στο πλαίσιο Covariate(s) με τη σειρά που αναγράφονται εκεί, αν και δοκιμάστηκαν και άλλες μέθοδοι.

Τέλος, τα αποτελέσματα εμφανίζονται στα δεδομένα εξόδου. Για τον έλεγχο καταλληλότητας του μοντέλου εφαρμόζονται τα κριτήρια που προαναφέρθηκαν.

Επιδιώκεται:

- ✓ Οι τιμές και τα πρόσημα των συντελεστών παλινδρόμησης β_i να μπορούν να εξηγηθούν λογικά.
- ✓ Ο **σταθερός όρος** της εξίσωσης, που εκφράζει το σύνολο των παραμέτρων που δε λήφθησαν υπόψη, να είναι κατά το δυνατό μικρότερος.
- ✓ Η τιμή του στατιστικού ελέγχου **Wald** να είναι μεγαλύτερη από την τιμή 1,7 για επίπεδο εμπιστοσύνης 95% και το επίπεδο σημαντικότητας να είναι μικρότερο από 5%.
- ✓ Τα τελικά μοντέλα να έχουν όσο το δυνατόν χαμηλότερο **LRT**, ειδικά σε σχέση με τα αρχικά (μοντέλα χωρίς μεταβλητές).
- ✓ Ο συντελεστής συσχέτισης R^2 να είναι κατά το δυνατό μεγαλύτερος, αν και δευτερεύων στην ανάλυση λογιστικής παλινδρόμησης.

4. ΣΥΛΛΟΓΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΣΤΟΙΧΕΙΩΝ

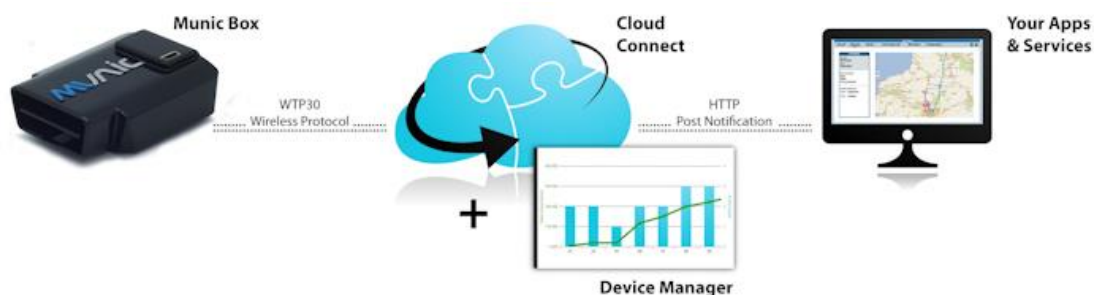
4.1 Εισαγωγή

Η παρούσα Διπλωματική Εργασία έχει ως σκοπό τη διερεύνηση της κυκλοφοριακής συμπεριφοράς του οδηγού μέσω των διαγνωστικών στοιχείων του οχήματος. Μετά τη βιβλιογραφική ανασκόπηση ερευνών συναφών με το αντικείμενο της Διπλωματικής Εργασίας, ήταν απαραίτητη η συλλογή των στοιχείων και η κατάλληλη επεξεργασία τους.

Στο κεφάλαιο αυτό παρουσιάζεται η διαδικασία συλλογής και επεξεργασίας των στοιχείων, έτσι ώστε να δοθεί μια πλήρης εικόνα για την ποιότητα και αξιοπιστία των δεδομένων που χρησιμοποιήθηκαν. Γίνεται αναφορά στα προβλήματα που προέκυψαν κατά την επεξεργασία τους, παρουσιάζεται ο τρόπος κωδικοποίησής τους καθώς και η αρχική επεξεργασία που υπέστησαν στο πρόγραμμα Microsoft Excel.

4.2 Συλλογή στοιχείων

Όλα τα δεδομένα που είναι απαραίτητα για την διερεύνηση της κυκλοφοριακής συμπεριφοράς του οδηγού προέρχονται από μία συσκευή που βρίσκεται εντός του οχήματος και είναι συνδεδεμένη με την κεντρική ηλεκτρονική μονάδα ελέγχου. Πρόκειται για ένα μηχάνημα διαστάσεων 27 x 48 x 49.5 cm το οποίο στέλνει άμεσα πληροφορίες σχετικά με την οδήγηση του χρήστη σε συσκευές όπως το κινητό του ή ο ηλεκτρονικός του υπολογιστής.



Εικόνα 4.1: Σχηματική απεικόνιση λειτουργίας συστήματος [Πηγή: munic.box]

Η σύνδεση μεταξύ της ενσωματωμένης συσκευής στο όχημα και του Cloud Connect (υπηρεσία αποθήκευσης δεδομένων στο διαδίκτυο, όπου αποκτάς πρόσβαση μέσω δημιουργίας λογαριασμού) γίνεται μέσω ενός προηγμένου πρωτοκόλλου το οποίο επιτρέπει συνεχή και σε πραγματικό χρόνο μεταφορά των δεδομένων, συμπίεση όλων των αρχείων και βελτιωμένη διαχείριση των δεδομένων παρακολούθησης του οδηγού.

Στις μέρες μας, κάθε όχημα περιλαμβάνει ένα περίπλοκο σύστημα δικτύων (OBD – On Board Diagnostics) το οποίο είναι σε θέση να μας παρέχει όποια πληροφορία χρειάζεται σχετικά με το όχημα, όπως η χρήση των ζωνών ασφαλείας, η εκπομπή

ρύπων, η πίεση των ελαστικών κτλ. Ωστόσο, το εν λόγω μηχανήμα μας δίνει τη δυνατότητα να πάμε ένα βήμα παραπέρα από το πρότυπο OBD και αποθηκεύοντας όλα τα δεδομένα που προκύπτουν από την χρήση του οχήματος από τον οδηγό, μπορούμε να τα επεξεργαστούμε καταλλήλως και να εκτιμήσουμε την κυκλοφοριακή του συμπεριφορά, θέμα το οποίο αποτελεί και το αντικείμενο της παρούσας Διπλωματικής Εργασίας.

4.3 Επεξεργασία Στοιχείων

4.3.1 Διαμόρφωση Δεδομένων

Τα δεδομένα καταγράφονται από το μηχανήμα ανά δευτερόλεπτο (second-based data) για κάθε ημέρα που ο οδηγός χρησιμοποίησε το όχημα και στέλνονται στη συνδεδεμένη συσκευή σε μορφή αρχείου csv.

CSV είναι αρχεία στα οποία αποθηκεύονται μέσα τους δεδομένα "τύπου πίνακα". Πιο απλά, είναι στην ουσία αρχεία που περιέχουν μέσα τους δεδομένα υπό μορφή Excel, δηλαδή περιέχουν πολλαπλά πεδία με μία ή περισσότερες τιμές για κάθε τέτοιο πεδίο, όπως συμβαίνει με έναν πίνακα excel πχ με πολλαπλές στήλες και γραμμές. Το csv είναι τα αρχικά των λέξεων Comma Separated Values που υποδηλώνει στην ουσία μία σειρά μεταβλητών με τιμές μέσα τους, χωρισμένες απλά με ένα κόμμα.

Για τις ανάγκες της παρούσας Διπλωματικής Εργασίας, τα δεδομένα αποθηκεύονταν σε ηλεκτρονικό υπολογιστή.

Date	Recorded_at	Longitude	Latitude	GPRMC_VALID	GPS_SPEED	GPS_DIR	DIO_IGNITION	ODO_FULL	GPS_PDOP	GPS_FIXED_SAT_NUM	BEHAVE_ID	BEHAVE_GPS_SPEED_BEGIN	BEHAVE_GPS_SPEED_END
2015/09/09	17:48:12	23.80495	38.02055	A,0,13.01	True	16887,2,10	-	-	-	-	-	-	-
2015/09/09	17:48:11	23.80495	38.02055	A,0,13.01	True	16887,2,10	-	-	-	-	-	-	-
2015/09/09	17:48:10	23.80495	38.02055	A,0,13.01	True	16887,2,10	-	-	-	-	-	-	-
2015/09/09	17:48:09	23.80495	38.02055	A,0,13.01	True	16887,2,10	-	-	-	-	-	-	-
2015/09/09	17:48:08	23.80495	38.02055	A,0,13.01	True	16887,2,10	-	-	-	-	-	-	-
2015/09/09	17:48:07	23.80495	38.02055	A,0,13.01	True	16887,2,10	-	-	-	-	-	-	-
2015/09/09	17:48:06	23.80495	38.02055	A,0,13.01	True	16887,2,10	-	-	-	-	-	-	-
2015/09/09	17:48:05	23.80495	38.02055	A,0,13.01	True	16887,2,10	-	-	-	-	-	-	-
2015/09/09	17:48:04	23.80495	38.02055	A,0,13.01	True	16887,2,10	-	-	-	-	-	-	-
2015/09/09	17:48:03	23.80495	38.02055	A,0,13.01	True	16887,2,10	-	-	-	-	-	-	-
2015/09/09	17:48:02	23.80495	38.02055	A,0,13.01	True	16887,2,10	-	-	-	-	-	-	-
2015/09/09	17:48:01	23.80495	38.02055	A,0,13.01	True	16887,2,10	-	-	-	-	-	-	-
2015/09/09	17:48:00	23.80495	38.02055	A,0,13.01	True	16887,2,9	-	-	-	-	-	-	-
2015/09/09	17:47:59	23.80495	38.02055	A,0,13.01	True	16887,2,9	-	-	-	-	-	-	-
2015/09/09	17:47:58	23.80495	38.02055	A,0,13.01	True	16887,2,10	-	-	-	-	-	-	-
2015/09/09	17:47:57	23.80495	38.02055	A,0,13.01	True	16887,2,9	-	-	-	-	-	-	-
2015/09/09	17:47:56	23.80495	38.02055	A,0,13.01	True	16887,2,9	-	-	-	-	-	-	-
2015/09/09	17:47:55	23.80495	38.02055	A,0,13.01	True	16887,2,9	-	-	-	-	-	-	-
2015/09/09	17:47:54	23.80495	38.02055	A,0,13.01	True	16887,2,9	-	-	-	-	-	-	-

Εικόνα 4.2: Απόσπασμα αρχικού αρχείου csv

Για την ευχερέστερη ανάγνωση και επεξεργασία των στοιχείων, έγινε η απαραίτητη μετατροπή σε αρχείο Excel.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Date	Recorded_at	Longitude	Latitude	GPRMC_VALID	GPS_SPEED	GPS_DIR	DIO_IGNITION	ODO_FULL	GPS_POOP	S_FIXED_SAT_NU	BEHAVE_ID
2						km/h	1/100 degrees		km			
3	14/7/2015	17:28:55	23.80502	38.0205	A	0	31.81	TRUE	10950	2	8	-
4	14/7/2015	17:28:54	23.80502	38.0205	A	0	31.81	TRUE	10950	2	8	-
5	14/7/2015	17:28:53	23.80502	38.0205	A	0	31.81	TRUE	10950	2	8	-
6	14/7/2015	17:28:51	23.80502	38.0205	A	0	31.81	TRUE	10950	2	8	-
7	14/7/2015	17:28:50	23.80502	38.0205	A	0	31.81	TRUE	10943	2	8	-
8	14/7/2015	17:28:49	23.80502	38.0205	A	0	31.81	TRUE	10943	2	7	-
9	14/7/2015	17:28:48	23.80502	38.0205	A	0	31.81	TRUE	10943	2	7	-
10	14/7/2015	17:28:47	23.80502	38.0205	A	0	31.81	TRUE	10943	2	7	-
11	14/7/2015	17:28:46	23.80502	38.0205	A	0	31.81	TRUE	10943	2	7	-
12	14/7/2015	17:28:45	23.80502	38.0205	A	0	31.81	TRUE	10943	2	7	-
13	14/7/2015	17:28:44	23.80502	38.0205	A	0	31.81	TRUE	10943	2	7	-
14	14/7/2015	17:28:43	23.80502	38.0205	A	0	31.81	TRUE	10943	2	7	-
15	14/7/2015	17:28:42	23.80502	38.0205	A	0	31.81	TRUE	10943	2	7	-
16	14/7/2015	17:28:41	23.80502	38.0205	A	0	31.81	TRUE	10943	2	8	-
17	14/7/2015	17:28:40	23.80502	38.0205	A	0	31.81	TRUE	10943	2	8	-
18	14/7/2015	17:28:39	23.80502	38.0205	A	0	31.81	TRUE	10943	2	8	-
19	14/7/2015	17:28:38	23.80502	38.0205	A	0	31.81	TRUE	10943	2	8	-
20	14/7/2015	17:28:37	23.80502	38.0205	A	0	31.81	TRUE	10943	2	7	-
21	14/7/2015	17:28:36	23.80502	38.0205	A	0	31.81	TRUE	10943	2	8	-

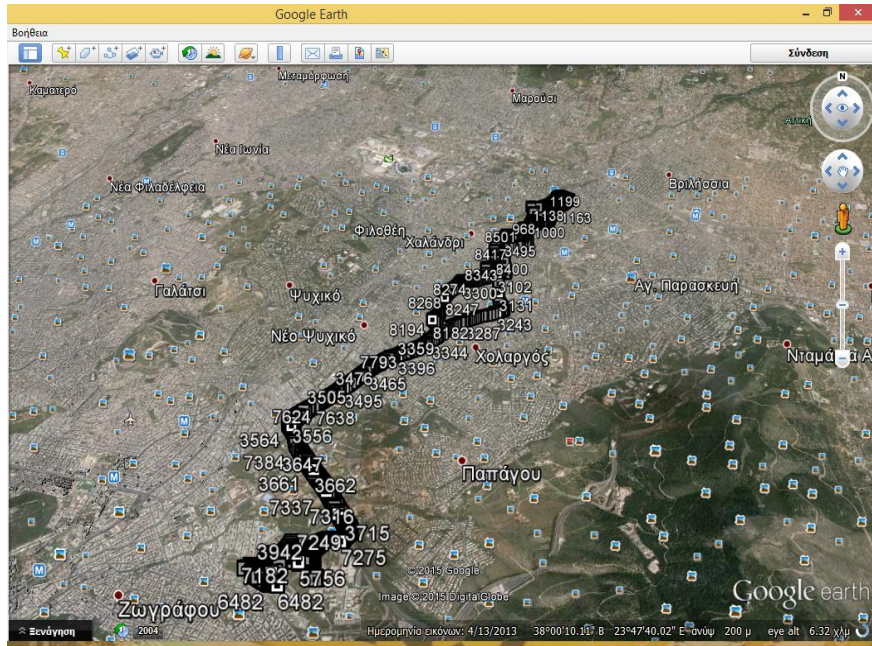
Εικόνα 4.3: Απόσπασμα δεδομένων σε αρχείο Excel

Τα δεδομένα που συλλέχθηκαν αφορούσαν 2 οδηγούς (driver 1, driver 2) οι οποίοι οδήγησαν 49 και 54 ημέρες αντίστοιχα, τους μήνες Ιούνιο και Ιούλιο. Κάθε ημέρα αποθηκευόταν σε ξεχωριστό αρχείο Excel, το οποίο περιελάμβανε όλες τις απαραίτητες πληροφορίες για τον οδηγό (ταχύτητα, συντεταγμένες, ώρα και ημέρα οδήγησης, κατανάλωση καυσίμων, κατάσταση μπαταρίας οχήματος, διανυόμενη απόσταση κτλ.).

Η πιο σημαντική, ίσως, πληροφορία που μας παρείχε το μηχανήμα ήταν η καταγραφή κάποιου γεγονότος (event). Ως event ορίζεται η στιγμή εκείνη όπου ο οδηγός προβαίνει σε κάποια απότομη αλλαγή ταχύτητας (επιτάχυνση ή φρενάρισμα) ή σε κάποιον απότομο ελιγμό (πλευρική επιτάχυνση). Κάθε φορά, λοιπόν, όπου είχαμε κάποιο event κατά τη διάρκεια οδήγησης, το μηχανήμα την κατέγραφε και μας έδινε παράλληλα την τιμή της επιτάχυνσης (ή επιβράδυνσης) στο σημείο αυτό. Κάθε τιμή κάτω των $1,50 \text{ m/s}^2$ κρίθηκε αμελητέα και δεν θεωρήθηκε ως event. Το όριο αυτό ορίστηκε μετά από έλεγχο σε έρευνες με παρόμοιο περιεχόμενο.

4.3.2 Περιγραφικά Στατιστικά (Descriptive Statistics)

- ✓ Με βάση τις συντεταγμένες και με τη βοήθεια του διαδικτύου (www.gpsvisualizer.com, google earth) βρέθηκε η διαδρομή που ακολουθούσε ανά ημέρα ο κάθε οδηγός, ώστε να μπορούσαμε να την χωρίσουμε ανά τύπο οδού (επαρχιακή οδός, κατοικημένη περιοχή, οδός ήπιας κυκλοφορίας, κεντρική αρτηρία με ύπαρξη σηματοδότησης, αυτοκινητόδρομος, εθνική οδός).



Εικόνα 4.4: Χάρτης διαδρομής οδηγού

- ✓ Με βάση την ώρα οδήγησης ανά ημέρα, βρέθηκε το πότε και για ποιο χρονικό διάστημα ο κάθε οδηγός χρησιμοποιεί το όχημά του τις επικίνδυνες ώρες (risky hours), διάστημα το οποίο ορίζεται μεταξύ 00.00 – 04.00.
- ✓ Με βάση την ταχύτητα του οχήματος σε κάθε θέση και σε συνδυασμό με τον τύπο οδού στον οποίον βρισκόταν ο οδηγός, εξετάστηκε το ενδεχόμενο της παραβίασης ταχύτητας. Τα όρια ταχύτητας λήφθηκαν υπόψη βάσει του ΚΟΚ (αναπροσαρμογή Ν.3542/02.03.2007/ΦΕΚ.50α).

Πίνακας 4.1: Όρια ταχύτητας βάσει ΚΟΚ

Τύπος Οδού	Όριο ταχύτητας (km/h)
Αυτοκινητόδρομος	130
Εθνική Οδός	110
Κεντρική αρτηρία με ύπαρξη σηματοδότησης	70
Κατοικημένη περιοχή	50
Επαρχιακή Οδός	50
Δρόμος ήπιας κυκλοφορίας	30

- ✓ Τέλος, καταγράφησαν αναλυτικά οι θέσεις και ο αριθμός των γεγονότων (events) που πραγματοποίησε ο κάθε οδηγός ανά ημέρα αλλά και συνολικά κατά τη διάρκεια του διμήνου που εξετάστηκε.

Όλα τα παραπάνω αποσκοπούσαν στην εξαγωγή των περιγραφικών στατιστικών (descriptive statistics) για τον κάθε οδηγό.

Πίνακας 4.2: Descriptive Statistics - driver 1

<i>Driver 1</i>	Χιλιόμετρα	Συμβάντα ταχύτητας	Απότομη Επιτάχυνση	Απότομη Επιβράδυνση	Συμβάντα στροφών	Απότομη Δεξιά Στροφή	Απότομη Αριστερή Στροφή
Δρόμος ήπιας κυκλοφορίας	151.85	398	163	235	370	171	199
Κεντρική αρτηρία με ύπαρξη σηματοδότησης	431.80	997	473	524	491	261	230
Κατοικημένη περιοχή	476.40	714	332	382	414	233	181
Εθνική Οδός	586.05	86	33	53	84	41	43
Επαρχιακή Οδός	219.50	58	12	46	37	16	21
Αυτοκινητόδρομος	1653.40	400	173	227	301	185	116
Σύνολο	3519.00	2653	1186	1467	1697	907	790

Πίνακας 4.3: Descriptive Statistics - driver 2

<i>Driver 2</i>	Χιλιόμετρα	Συμβάντα ταχύτητας	Απότομη Επιτάχυνση	Απότομη Επιβράδυνση	Συμβάντα στροφών	Απότομη Δεξιά Στροφή	Απότομη Αριστερή Στροφή
Δρόμος ήπιας κυκλοφορίας	11.96	84	39	45	34	16	18
Κεντρική αρτηρία με ύπαρξη σηματοδότησης	999.80	1537	822	715	741	421	320
Κατοικημένη περιοχή	318.97	524	269	255	247	133	114
Εθνική Οδός	0.30	1	1	0	0	0	0
Επαρχιακή Οδός	0.00	0	0	0	0	0	0
Αυτοκινητόδρομος	124.88	125	70	55	98	45	53
Κεντρική αρτηρία χωρίς την ύπαρξη σηματοδότησης (Λ. Συγγρού)	6.09	0	0	0	0	0	0
Σύνολο	1462.00	2271	1201	1070	1120	615	505

4.3.3 Διαμόρφωση βάσης δεδομένων

Πριν την εισαγωγή των δεδομένων στο ειδικό στατιστικό πρόγραμμα έγινε κατάλληλη επεξεργασία κάθε φύλλου Excel με σκοπό να δημιουργηθεί ένα ενιαίο αρχείο και για τους 2 οδηγούς, κωδικοποιώντας τις μεταβλητές οι οποίες θα μας ενδιέφεραν για την ανάλυσή μας. Ωστόσο, λόγω του τεράστιου όγκου δεδομένων

ήταν πρακτικά αδύνατο να ενοποιήσουμε όλα τα στοιχεία που είχαμε στη διάθεσή μας σε ένα αρχείο. Επομένως, προτιμήθηκε η λύση δύο ξεχωριστών αρχείων Excel, καθένα από τα οποία περιείχε όλα τα κυκλοφοριακά δεδομένα για κάθε οδηγό ξεχωριστά, για ολόκληρο το χρονικό διάστημα για το οποίο εξετάστηκε.

Για να μπορέσουμε να πραγματοποιήσουμε όσο το δυνατόν πιο σωστά τις αναλύσεις μας, ήταν απαραίτητη η κωδικοποίηση συγκεκριμένων μεταβλητών και η πρόσθηση νέων στήλων όπως:

- ❖ **Τύπος Οδού (Type of road):** Όπως προαναφέρθηκε, μέσω των συντεταγμένων είχαμε τη δυνατότητα να βρίσκουμε οποιαδήποτε στιγμή την τοποθεσία του οχήματος. Σε μια νέα στήλη, λοιπόν, σημειώθηκε ο τύπος της οδού με την ονομασία να δίδεται στον παρακάτω πίνακα:

Πίνακας 4.4: Ονομασία ανά τύπο οδού

Τύπος Οδού	Ονομασία
Αυτοκινητόδρομος	6
Εθνική Οδός	4
Κεντρική αρτηρία με ύπαρξη σηματοδότησης	2
Κατοικημένη περιοχή	3
Επαρχιακή Οδός	5
Δρόμος ήπιας κυκλοφορίας	1

- ❖ **Επικίνδυνες ώρες (Risky Hours):** Η νέα στήλη που δημιουργήθηκε περιείχε το 1, όταν η οδήγηση γινόταν μεταξύ 00.00 – 04.00, και το 0 καθ' όλη την υπόλοιπη διάρκεια της οδήγησης.
- ❖ **Ταυτότητα Οδηγού (Driver ID):** Εύλογα προκύπτει ότι, όταν οδηγούσε ο 1^{ος} οδηγός η στήλη περιείχε το νούμερο 1, ενώ αντίστοιχα, όταν οδηγούσε ο 2^{ος} οδηγός περιείχε το νούμερο 2.
- ❖ **Ημέρα οδήγησης (Day):** Με βάση την ημερομηνία οδήγησης, καταγράφηκε και η συγκεκριμένη ημέρα που χρησιμοποιήθηκε το όχημα. Η ονομασία δίδεται στον παρακάτω πίνακα:

Πίνακας 4.5: Ονομασία ανά ημέρα

Ημέρα	Ονομασία
Δευτέρα	1
Τρίτη	2
Τετάρτη	3
Πέμπτη	4
Παρασκευή	5
Σάββατο	6
Κυριακή	7

- ❖ **Υπέρβαση ορίου ταχύτητας (OverSpeed):** Η εν λόγω στήλη περιείχε το 1 στα σημεία όπου ο οδηγός κινούταν άνω του εκάστοτε ορίου ταχύτητας και το 0 στα σημεία όπου οδηγούσε σύμφωνα με τον ΚΟΚ.
- ❖ **Απότομη Επιτάχυνση (Sudden Acceleration):** Στα σημεία όπου υπήρχε η καταγραφή κάποιου γεγονότος ταχύτητας με τιμή άνω των $1,50 \text{ m/s}^2$ σημειώθηκε ο αριθμός 1. Σε αντίθετη περίπτωση, ο αριθμός 0.
- ❖ **Απότομη Επιβράδυνση (Sudden Deceleration):** Στα σημεία όπου υπήρχε η καταγραφή κάποιου γεγονότος ταχύτητας με τιμή κάτω των $-1,50 \text{ m/s}^2$ σημειώθηκε ο αριθμός 1. Σε αντίθετη περίπτωση, ο αριθμός 0.
- ❖ **Απότομη Δεξιά στροφή (Sudden Right Turn):** Στα σημεία όπου υπήρχε η καταγραφή κάποιου γεγονότος στροφής με τιμή άνω των $1,50 \text{ m/s}^2$ σημειώθηκε ο αριθμός 1. Σε αντίθετη περίπτωση, ο αριθμός 0.
- ❖ **Απότομη Αριστερή στροφή (Sudden Left Turn):** Στα σημεία όπου υπήρχε η καταγραφή κάποιου γεγονότος στροφής με τιμή κάτω των $-1,50 \text{ m/s}^2$ σημειώθηκε ο αριθμός 1. Σε αντίθετη περίπτωση, ο αριθμός 0.

Το τελικό αρχείο Excel για κάθε οδηγό περιείχε μόνο τις μεταβλητές που μας ενδιέφεραν για την επεξεργασία στο ειδικό στατιστικό πρόγραμμα και κρατήθηκαν οι χρονικές στιγμές όπου είχαμε κάποιο σημαντικό γεγονός (απότομη αλλαγή ταχύτητας ή απότομο ελιγμό) ή παράβαση ορίου ταχύτητας.

4.3.4 Εισαγωγή δεδομένων στο ειδικό στατιστικό πρόγραμμα

Μετά τη διαμόρφωση του τελικού πίνακα Excel, πραγματοποιήθηκε εισαγωγή των στοιχείων στο πεδίο δεδομένων (data view) του ειδικού λογισμικού στατιστικής ανάλυσης.

	Date	Day	DriverID	Recorded_at	Trips	FrequencyofTrips	RiskyHour	Longitude	Latitude	Distance	TypeofRoad
1	04-Jun-2015	4,0	2,0	17:52:09.00	6,0	2,0	,0	23,79712	38,01259	48,714019298346340	.
2	04-Jun-2015	4,0	2,0	18:59:33.00	6,0	3,0	,0	23,79440	38,01357	122,822585676238130	.
3	05-Jun-2015	5,0	2,0	11:04:57.00	6,0	4,0	,0	23,79470	38,01845	159,515409717120660	.
4	05-Jun-2015	5,0	2,0	11:05:04.00	6,0	4,0	,0	23,79470	38,01845	159,515409717120660	.
5	05-Jun-2015	5,0	2,0	11:05:12.00	6,0	4,0	,0	23,79470	38,01845	159,515409717120660	.
6	05-Jun-2015	5,0	2,0	11:05:23.00	6,0	4,0	,0	23,79300	38,01713	176,836025309216650	.
7	05-Jun-2015	5,0	2,0	11:05:36.00	6,0	4,0	,0	23,79345	38,01663	183,537908166382380	.
8	05-Jun-2015	5,0	2,0	11:05:40.00	6,0	4,0	,0	23,79360	38,01647	185,128461125334600	.
9	05-Jun-2015	5,0	2,0	11:05:50.00	6,0	4,0	,0	23,79293	38,01586	191,490903791161680	.
10	05-Jun-2015	2,0	2,0	11:10:50.00	6,0	4,0	,0	23,79338	38,01663	.	.
11	05-Jun-2015	2,0	2,0	11:10:51.00	6,0	4,0	,0	23,79353	38,01669	.	.
12	05-Jun-2015	2,0	2,0	11:10:52.00	6,0	4,0	,0	23,79368	38,01677	.	.
13	05-Jun-2015	2,0	2,0	11:12:46.00	6,0	4,0	,0	23,79454	38,01864	.	.
14	05-Jun-2015	2,0	2,0	11:12:47.00	6,0	4,0	,0	23,79454	38,01864	.	.
15	05-Jun-2015	2,0	2,0	11:12:48.00	6,0	4,0	,0	23,79454	38,01864	.	.
16	05-Jun-2015	2,0	2,0	11:12:49.00	6,0	4,0	,0	23,79454	38,01864	.	.
17	05-Jun-2015	2,0	2,0	11:12:50.00	6,0	4,0	,0	23,79454	38,01864	.	.
18	05-Jun-2015	2,0	2,0	11:12:51.00	6,0	4,0	,0	23,79454	38,01864	.	.
19	05-Jun-2015	2,0	2,0	11:18:36.00	6,0	4,0	,0	23,79463	38,01887	.	.
20	05-Jun-2015	2,0	2,0	11:18:37.00	6,0	4,0	,0	23,79462	38,01890	.	.
21	05-Jun-2015	2,0	2,0	11:18:38.00	6,0	4,0	,0	23,79462	38,01890	.	.
22	05-Jun-2015	2,0	2,0	11:18:39.00	6,0	4,0	,0	23,79457	38,01900	.	.

Εικόνα 4.5: Εισαγωγή των στοιχείων στο πεδίο δεδομένων

Στη συνέχεια καθορίστηκε στο πεδίο μεταβλητών (variable view) το όνομα, ο τύπος, ο αριθμός των ψηφίων κάθε μεταβλητής και το αν η μεταβλητή είναι συνεχής (scale) ή διακριτή (nominal ή ordinal). Επισημαίνεται ότι nominal ονομάζεται μια μεταβλητή όταν οι απαντήσεις της παρουσιάζονται βάσει κατάταξης.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Date	Date	11	0		None	None	11	Right	Scale	Input
2	Day	Numeric	12	1		None	None	12	Right	Nominal	Input
3	DriverID	Numeric	12	1		None	None	12	Right	Nominal	Input
4	Recorded_at	Date	11	2		None	None	11	Right	Scale	Input
5	Trips	Numeric	12	1		None	None	12	Right	Nominal	Input
6	Frequencyof...	Numeric	12	1		None	None	12	Right	Scale	Input
7	RiskyHour	Numeric	12	1		None	None	12	Right	Nominal	Input
8	Longitude	Numeric	12	5		None	None	12	Right	Scale	Input
9	Latitude	Numeric	12	5		None	None	12	Right	Scale	Input
10	Distance	Numeric	16	15		None	None	16	Right	Scale	Input
11	TypeofRoad	Numeric	12	1		None	None	12	Right	Nominal	Input
12	ODO_FULLL	Numeric	12	1		None	None	12	Right	Scale	Input
13	BEHAVE_ID	Numeric	12	1		None	None	12	Right	Nominal	Input
14	SuddenAcc	Numeric	12	1		None	None	12	Right	Nominal	Input
15	SuddenDec	Numeric	12	1		None	None	12	Right	Nominal	Input
16	SuddenRigh...	Numeric	12	1		None	None	12	Right	Nominal	Input
17	SuddenLeft...	Numeric	12	1		None	None	12	Right	Nominal	Input
18	MDI_OBD_...	Numeric	12	1		None	None	12	Right	Scale	Input
19	OverSpeed	Numeric	12	1		None	None	12	Right	Nominal	Input
20	MDI_OBD_...	Numeric	12	1		None	None	12	Right	Scale	Input
21	MDI_OBD_...	Numeric	12	1		None	None	12	Right	Scale	Input
22	MDI_OBD_...	Numeric	12	1		None	None	12	Right	Scale	Input
23	COUNTER	Numeric	12	1		None	None	12	Right	Nominal	Input

Εικόνα 4.6: Καθορισμός των μεταβλητών στο πεδίο μεταβλητών

5. ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ – ΑΠΟΤΕΛΕΣΜΑΤΑ

5.1 Εισαγωγή

Το κεφάλαιο αυτό περιλαμβάνει την αναλυτική περιγραφή της εφαρμογής της μεθοδολογίας, καθώς και την παρουσίαση των αποτελεσμάτων της Διπλωματικής Εργασίας. Όπως προαναφέρθηκε, ύστερα από τη βιβλιογραφική ανασκόπηση συναφών ερευνών, έγινε η επιλογή της κατάλληλης μεθοδολογίας για την παρούσα εργασία. Η μέθοδος που επιλέχθηκε, η οποία παρουσιάστηκε αναλυτικά στο κεφάλαιο 3 με τίτλο "Θεωρητικό Υπόβαθρο", είναι η Ανάλυση Λογιστικής Παλινδρόμησης.

Πιο συγκεκριμένα, στο κεφάλαιο αυτό περιγράφονται αναλυτικά τα βήματα που ακολουθήθηκαν κατά την εφαρμογή της μεθοδολογίας και παρουσιάζεται η διαδικασία ανάπτυξης κατάλληλων μοντέλων. Ιδιαίτερη έμφαση δίδεται στην παρουσίαση ζητημάτων αξιοπιστίας των δεδομένων και στις διαδικασίες αντιμετώπισης τους. Αναπόσπαστο μέρος των αποτελεσμάτων αποτελούν οι στατιστικοί έλεγχοι που απαιτούνται για την αποδοχή ή την απόρριψη των μαθηματικών μοντέλων.

5.2 Ανάπτυξη και Εφαρμογή Μοντέλων

Στην παράγραφο αυτή, περιγράφεται η διαδικασία ανάλυσης για την ανάπτυξη των μαθηματικών μοντέλων που αφορούν στην διερεύνηση της κυκλοφοριακής συμπεριφοράς των οδηγών. Σκοπός της ανάλυσης λογιστικής παλινδρόμησης είναι να υπολογισθεί ένα στατιστικά σημαντικό μοντέλο για κάθε οδηγό και να εντοπιστούν οι μεταβλητές που επηρεάζουν περισσότερο την συμπεριφορά τους στο δρόμο.

5.2.1 Δεδομένα εισόδου – Καθορισμός Μεταβλητών

Στα στατιστικά μοντέλα προσδιορισμού της κυκλοφοριακής συμπεριφοράς του οδηγού με χρήση των διαγνωστικών στοιχείων του οχήματος εξετάστηκαν όλες οι μεταβλητές που περιγράφηκαν στο υποκεφάλαιο 4.3.3. Η βάση δεδομένων, η οποία ήταν ήδη εισηγμένη στο ειδικό λογισμικό στατιστικής ανάλυσης, επεξεργάστηκε σύμφωνα με τη διαδικασία που παρουσιάστηκε αναλυτικά στο προηγούμενο κεφάλαιο, όπου καθορίστηκε το όνομα και ο τύπος των μεταβλητών.

Σημειώνεται ότι τα τελικά μοντέλα που προέκυψαν ήταν αποτέλεσμα μιας σειράς δοκιμών, κατά τις οποίες αναπτύχθηκαν μαθηματικά μοντέλα που περιελάμβαναν συνδυασμούς όλων των μεταβλητών που καταγράφηκαν. Τα μοντέλα αυτά αξιολογήθηκαν με βάση τα αποτελέσματα των στατιστικών ελέγχων (Wald, πιθανοφάνεια, κλπ.), όπως αυτοί έχουν αναφερθεί σε προηγούμενο κεφάλαιο, αλλά και με βάση τη λογική εξήγηση των αποτελεσμάτων. Στις δοκιμές αυτές απορρίφθηκαν οι μεταβλητές που αποδείχθηκαν ότι δεν έχουν στατιστικά σημαντική επιρροή. Με αυτή τη διαδικασία διαδοχικών δοκιμών και απόρριψης μοντέλων προέκυψαν τα μαθηματικά μοντέλα με τις καλύτερες επιδόσεις στατιστικής σημαντικότητας, όπως αυτά παρουσιάζονται στα επόμενα υποκεφάλαια.

5.2.2 Λογιστική Ανάλυση Παλινδρόμησης

Με στόχο τη διερεύνηση της κυκλοφοριακής συμπεριφοράς των οδηγών χρησιμοποιήθηκαν ως εξαρτημένες μεταβλητές τα παρακάτω δεδομένα:

- i. Απότομη Δεξιά Στροφή (Sudden Right Turn)
- ii. Απότομη Αριστερή Στροφή (Sudden Left Turn)
- iii. Απότομη επιβράδυνση (Sudden Deceleration)
- iv. Απότομη Επιτάχυνση (Sudden Acceleration)
- v. Παραβίαση ανώτατου ορίου ταχύτητας (OverSpeed)

Οι μεταβλητές αυτές είναι διακριτές και λαμβάνουν μόνο δύο τιμές (0 = Μη εμφάνιση αυτού του συμβάντος, 1 = Εμφάνιση). Για το λόγο αυτό η ανάλυση έγινε με (διωνυμική) λογιστική ανάλυση παλινδρόμησης (**binary logistic regression**). Στο SPSS η λογιστική παλινδρόμηση εφαρμόζεται μέσω της ακολουθίας των εντολών analyze → regression → binary logistic. Τη μετάβαση στην επιλογή binary logistic regression διαδέχεται ο καθορισμός των εξαρτημένων και ανεξάρτητων μεταβλητών. Η μεταβλητή που ενδιαφέρει (εξαρτημένη μεταβλητή) εισάγεται στο πλαίσιο Dependent ενώ οι μεταβλητές που θα χρησιμοποιηθούν για να εξηγήσουν την μεταβλητότητα της εξαρτημένης μεταβλητής (ανεξάρτητες μεταβλητές) εισάγονται στο πλαίσιο Covariates.

Σύμφωνα με το Θεωρητικό Υπόβαθρο του κεφαλαίου 3, σε κάθε μοντέλο θα πρέπει να ελεγχθούν οι παρακάτω παράγοντες:

- ✓ Οι τιμές και τα πρόσημα των συντελεστών παλινδρόμησης β_i να μπορούν να εξηγηθούν λογικά.
- ✓ Ο **σταθερός όρος** της εξίσωσης, που εκφράζει το σύνολο των παραμέτρων που δε λήφθησαν υπόψη, να είναι κατά το δυνατό μικρότερος.
- ✓ Η τιμή του στατιστικού ελέγχου **Wald** να είναι μεγαλύτερη από την τιμή 1,7 για επίπεδο εμπιστοσύνης 95% και το επίπεδο σημαντικότητας να είναι μικρότερο από 5%.

- ✓ Τα τελικά μοντέλα να έχουν όσο το δυνατόν χαμηλότερο **LRT**, ειδικά σε σχέση με τα αρχικά (μοντέλα χωρίς μεταβλητές).
- ✓ Ο συντελεστής συσχέτισης **R²** να είναι κατά το δυνατό μεγαλύτερος, αν και δευτερεύων στην ανάλυση λογιστικής παλινδρόμησης.

Ειδικότερα για τον συντελεστή συσχέτισης, στη λογιστική ανάλυση χρησιμοποιούνται αρκετοί ψευδο-συντελεστές (δηλαδή εκτός του R² της γραμμικής παλινδρόμησης) διότι οι αναλύσεις δεν είναι γραμμικού τύπου. Ένας από τους κυριότερους συντελεστές συσχέτισης R² στη λογιστική παλινδρόμηση είναι ο Cox & Snell R² (R Square). Ο συγκεκριμένος συντελεστής είναι ελλιπής καθώς η μέγιστη τιμή του είναι 0,75 αντί για 1,00 και η διακύμανσή του είναι αρκετά μεγάλη (έως και 0,25). Για την αντιμετώπιση αυτών των προβλημάτων, χρησιμοποιείται ο διορθωμένος συντελεστής συσχέτισης Nagelkerke R² (R Square).

Εκτός από τους μαθηματικούς ελέγχους όμως, ο απώτερος σκοπός για κάθε μοντέλο είναι η ικανότητά του να **προβλέψει** με σχετική ακρίβεια το φαινόμενο το οποίο περιγράφει. Στη λογιστική παλινδρόμηση, αυτή η πρόβλεψη πραγματοποιείται για κάθε κατηγορία και περιγράφεται με ποσοστό επιτυχίας.

5.2.3 Περιγραφή Μοντέλων Λογιστικής Παλινδρόμησης

Τα τελικά μοντέλα της λογιστικής παλινδρόμησης παρουσιάζονται παρακάτω. Οι μαθηματικές σχέσεις που αναπτύχθηκαν για να υπολογίσουν την πιθανότητα εμφάνισης κάποιου γεγονότος είτε απότομης αλλαγής ταχύτητας είτε απότομου ελιγμού είτε υπέρβασης ορίου ταχύτητας υπολογίζουν τη συνάρτηση χρησιμότητας U ή αλλιώς Utility Function, οπότε η πιθανότητα εμφάνισης είναι:

$$P = \frac{e^U}{e^U + 1}$$

5.2.4 Γραμμική Ανάλυση Παλινδρόμησης

Τέλος, χρησιμοποιήθηκε ως εξαρτημένη μεταβλητή και η ταχύτητα του οχήματος (OBD_SPEED) ώστε να μπορέσουμε να διακρίνουμε από ποια στοιχεία της κυκλοφοριακής συμπεριφοράς του οδηγού επηρεάζεται σημαντικά. Η μεταβλητή αυτή, όμως, είναι συνεχής και για το λόγο αυτόν η ανάλυση γίνεται με τη μέθοδο της γραμμικής παλινδρόμησης (linear regression). Από τη στιγμή που η εξαρτημένη μεταβλητή μας εξαρτάται γραμμικά από περισσότερες από μία ανεξάρτητες μεταβλητές, η μέθοδος που χρησιμοποιείται ονομάζεται πολλαπλή γραμμική παλινδρόμηση (**multiple linear regression**).

Η εξίσωση η οποία αποτυπώνει τη σχέση ανάμεσα στην εξαρτημένη και τις ανεξάρτητες μεταβλητές έχει τη γενικότερη μορφή:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni} + \varepsilon_i$$

5.2.5 Τελικά Μοντέλα για την Κυκλοφοριακή Συμπεριφορά του Οδηγού

Παρακάτω θα παρουσιαστούν μόνο τα μαθηματικά μοντέλα που αφορούν τον οδηγό 2 (driver 2). Η ανάλυση έγινε και για τον οδηγό 1 (driver 1) με παρόμοια αποτελέσματα.

Για κάθε μοντέλο παρουσιάζονται 3 πίνακες:

- Περίληψη μοντέλου: ο πίνακας που περιλαμβάνει τους συντελεστές συσχέτισης R^2 κατά Cox & Snell και κατά Nagelkerke.
- Πίνακας κατάταξης: παρουσιάζει την πιθανότητα σωστής πρόβλεψης για το σύνολο των προβλέψεων του μοντέλου.
- Μεταβλητές στην εξίσωση: περιλαμβάνει τις ανεξάρτητες μεταβλητές που ορίζονται για κάθε εξαρτημένη μεταβλητή, δείχνοντας την τιμή του συντελεστή β_i , το τυπικό σφάλμα και την σημαντικότητα της κάθε μίας μεταβλητής.

❖ Απότομη Δεξιά Στροφή (Sudden Right Turn):

Πίνακας 5.1: Περίληψη Μοντέλου (Model Summary)

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	3511,459 ^a	,092	,402

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than ,001.

Πίνακας 5.2: Πίνακας Κατάταξης (Classification Table)^a

Observed		Predicted			
		SuddenRightTurn		Percentage Correct	
		,0	1,0		
Step 1	SuddenRightTurn	,0	18855	2129	89,9
		1,0	84	531	86,3
Overall Percentage					89,8

a. The cut value is ,100

Πίνακας 5.3: Μεταβλητές στην εξίσωση (Variables in the Equation)

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	SuddenAcc	-1,651	,124	177,302	1	,000	,192
	SuddenDec	-1,714	,129	176,487	1	,000	,180
	OverSpeed	-5,281	,143	1372,181	1	,000	,005
	Constant	-,129	,092	1,985	1	,159	,879

a. Variable(s) entered on step 1: SuddenAcc, SuddenDec, OverSpeed.

❖ Απότομη Αριστερή Στροφή (Sudden Left Turn):**Πίνακας 5.4:** Περίληψη Μοντέλου (Model Summary)

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	2717,092 ^a	,092	,460

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than ,001.

Πίνακας 5.5: Πίνακας Κατάταξης (Classification Table)^a

Observed		Predicted		
		SuddenLeftTurn		Percentage Correct
		,0	1,0	
Step 1	SuddenLeftTurn ,0	19944	1150	94,5
	1,0	152	353	69,9
Overall Percentage				94,0

a. The cut value is ,100

Πίνακας 5.6: Μεταβλητές στην εξίσωση (Variables in the Equation)

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a						
SuddenDec	-2,022	,135	225,259	1	,000	,132
OverSpeed	-6,083	,181	1128,711	1	,000	,002
SuddenAcc	-2,158	,134	258,704	1	,000	,116
Constant	-,046	,091	,250	1	,617	,955

a. Variable(s) entered on step 1: SuddenDec, OverSpeed, SuddenAcc.

❖ Απότομη Επιβράδυνση (Sudden Deceleration):**Πίνακας 5.7:** Περίληψη Μοντέλου (Model Summary)

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	4191,227 ^a	,181	,557

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than ,001.

Πίνακας 5.8: Πίνακας Κατάταξης (Classification Table)^a

Observed		Predicted		
		SuddenDec		Percentage Correct
		,0	1,0	
Step 1	SuddenDec ,0	20529	0	100,0
	1,0	1070	0	,0
Overall Percentage				95,0

a. The cut value is ,500

Πίνακας 5.9: Μεταβλητές στην εξίσωση (Variables in the Equation)

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a OverSpeed	-5,549	,153	1317,179	1	,000	,004
Constant	-,469	,040	138,258	1	,000	,626

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

❖ Απότομη Επιτάχυνση (Sudden Acceleration):

Πίνακας 5.10: Περίληψη Μοντέλου (Model Summary)

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	4003,867 ^a	,217	,620

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than ,001.

Πίνακας 5.11: Πίνακας Κατάταξης (Classification Table)^a

Observed		Predicted		
		SuddenAcc		Percentage Correct
		,0	1,0	
Step 1 SuddenAcc ,0	18924	1474	92,8	
1,0	44	1157	96,3	
Overall Percentage			93,0	

a. The cut value is ,100

Πίνακας 5.12: Μεταβλητές στην εξίσωση (Variables in the Equation)

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a OverSpeed	-3,906	,196	397,889	1	,000	,020
MDI_OBD_SPEED	-,041	,003	220,115	1	,000	,960
Constant	,826	,081	104,050	1	,000	2,284

a. Variable(s) entered on step 1: OverSpeed, MDI_OBD_SPEED.

❖ Παραβίαση ανώτατου ορίου ταχύτητας (OverSpeed):

Πίνακας 5.13: Περίληψη Μοντέλου (Model Summary)

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	2782,920 ^a	,461	,876

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than ,001.

Πίνακας 5.14: Πίνακας Κατάταξης (Classification Table)^a

Observed		Predicted		
		OverSpeed		Percentage Correct
		,0	1,0	
Step 1	OverSpeed	,0	1,0	
		2059	601	77,4
		31	18908	99,8
	Overall Percentage			97,1

a. The cut value is ,100

Πίνακας 5.15: Μεταβλητές στην εξίσωση (Variables in the Equation)

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a						
MDI_OBD_SPEED	,122	,003	1433,298	1	,000	1,129
SuddenAcc	-4,316	,201	460,604	1	,000	,013
SuddenDec	-5,216	,204	654,726	1	,000	,005
Constant	-3,575	,162	486,721	1	,000	,028

a. Variable(s) entered on step 1: MDI_OBD_SPEED, SuddenAcc, SuddenDec.

❖ Ταχύτητα Οχήματος (OBD_SPEED):

Πίνακας 5.16: Ανάλυση Διακύμανσης (ANOVA)^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5783509,436	4	1445877,359	7008,088	,000 ^b
	Residual	4455177,406	21594	206,316		
	Total	10238686,84	21598			

a. Dependent Variable: MDI_OBD_SPEED

b. Predictors: (Constant), SuddenDec, RiskyHour, SuddenAcc, OverSpeed

Πίνακας 5.17: Περίληψη Μοντέλου (Model Summary)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,752 ^a	,565	,565	14,3637

a. Predictors: (Constant), SuddenDec, RiskyHour, SuddenAcc, OverSpeed

Πίνακας 5.18: Συντελεστές (Coefficients)^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	38,022	,605		62,826	,000
	RiskyHour	2,375	,266	,040	8,919	,000
	OverSpeed	38,385	,611	,579	62,862	,000
	SuddenAcc	-16,032	,715	-,169	-22,431	,000
	SuddenDec	-10,006	,726	-,100	-13,778	,000

a. Dependent Variable: MDI_OBD_SPEED

5.3 Αποτελέσματα Μοντέλων

➤ Απότομη Δεξιά Στροφή (Sudden Right Turn)

Μαθηματική Σχέση:

$$U = -0,129 - 1,651 * (\text{SuddenAcc}) - 1,714 * (\text{SuddenDec}) - 5,281 * (\text{OverSpeed})$$

Περιγραφή Αποτελεσμάτων:

Στην παραπάνω σχέση παρατηρούμε ότι καθοριστικός παράγοντας στην εμφάνιση κάποιου γεγονότος απότομης δεξιάς στροφής από τον οδηγό είναι η **παραβίαση του ανώτατου ορίου ταχύτητας (OverSpeed)** με συντελεστή -5,281. Αυτό υποδηλώνει ότι στα σημεία όπου ο οδηγός έχει υπερβεί το ανώτατο όριο ταχύτητας της οδού, έχει λιγότερες πιθανότητες να προβεί σε απότομο ελιγμό προς τα δεξιά σε σχέση με τα σημεία όπου οδηγεί εντός ορίων ταχύτητας. Συγκεκριμένα, εμφανίζεται λόγος πιθανοτήτων 200 φορές (1/0,005) χαμηλότερος σε σχέση με το λόγο πιθανοτήτων στα σημεία όπου δεν υπάρχει παραβίαση του ανώτατου ορίου ταχύτητας. Επιπλέον, **οι απότομες αλλαγές ταχύτητας (SuddenAcc, SuddenDec)** φαίνεται να επηρεάζουν και αυτές σημαντικά την εμφάνιση γεγονότος απότομης δεξιάς στροφής με συντελεστή -1,651 και -1,714 αντίστοιχα. Στα σημεία όπου ο οδηγός έχει προβεί σε απότομη επιτάχυνση εμφανίζει λόγο πιθανότητας 5,2 φορές (1/0,192) χαμηλότερο να πραγματοποιήσει απότομο ελιγμό προς τα δεξιά σε σχέση με τα σημεία όπου οδηγεί κανονικά. Ο αντίστοιχος λόγος πιθανότητας για τα σημεία στα οποία υπάρχει γεγονός απότομης επιβράδυνσης είναι 5,5 φορές (1/0,180) χαμηλότερος από τα σημεία κανονικής οδήγησης.

➤ Απότομη Αριστερή Στροφή (Sudden Left Turn)

Μαθηματική Σχέση:

$$U = -0,046 - 6,083 * (\text{OverSpeed}) - 2,022 * (\text{SuddenDec}) - 2,158 * (\text{SuddenAcc})$$

Περιγραφή Αποτελεσμάτων:

Όπως αναμενόταν, τα αποτελέσματα της ανάλυσης για τα γεγονότα απότομης αριστερής στροφής είναι παρόμοια με τα αντίστοιχα για τα γεγονότα απότομης δεξιάς στροφής. Κύριο ρόλο παίζει και εδώ η **παραβίαση του ανώτατου ορίου ταχύτητας (OverSpeed)** με συντελεστή -6,083 και λόγο πιθανοτήτων 500 φορές (1/0,002) χαμηλότερο σε σχέση με τα σημεία όπου ο οδηγός υπακούει στον ΚΟΚ. Όσον αφορά τις **απότομες αλλαγές ταχύτητας (SuddenAcc, SuddenDec)**, επηρεάζουν και αυτές την εμφάνιση απότομου αριστερού ελιγμού με συντελεστή -2,158 και -2,022 αντίστοιχα. Ο λόγος πιθανοτήτων είναι 8,6 φορές (1/0,116) και 7,6 φορές (1/0,132) αντίστοιχα χαμηλότερος από τα σημεία της οδού όπου δεν εμφανίζεται απότομη επιτάχυνση ή επιβράδυνση.

➤ Απότομη Επιβράδυνση (Sudden Deceleration)

Μαθηματική Σχέση:

$$U = -0,469 - 5,549 * (\text{OverSpeed})$$

Περιγραφή Αποτελεσμάτων:

Παρατηρείται από την παραπάνω σχέση ότι για την εμφάνιση συμβάντος απότομης επιβράδυνσης κυρίαρχος και μοναδικός παράγοντας είναι η **παραβίαση του ανώτατου ορίου ταχύτητας (OverSpeed)** με συντελεστή -5,549 και λόγο πιθανοτήτων 250 φορές (1/0,004) χαμηλότερο σε σχέση με τα σημεία όπου ο οδηγός υπακούει στον ΚΟΚ.

➤ Απότομη Επιτάχυνση (Sudden Acceleration)

Μαθηματική Σχέση:

$$U = 0,826 - 3,906 * (\text{OverSpeed}) - 0,041 * (\text{OBD_SPEED})$$

Περιγραφή Αποτελεσμάτων:

Κυρίαρχη επίπτωση στα συμβάντα απότομης επιτάχυνσης έχει και εδώ η **παραβίαση του ανώτατου ορίου ταχύτητας (OverSpeed)** από τον οδηγό, με συντελεστή -3,906 και λόγο πιθανοτήτων 50 φορές (1/0,02) χαμηλότερο από τα σημεία όπου ο εξεταζόμενος οδηγεί εντός ορίων ταχύτητας. Όσον αφορά την **ταχύτητα του οχήματος (OBD SPEED)**, έχει συντελεστή -0,041 και δείχνει ότι με αύξηση της ταχύτητας έχουμε μείωση της πιθανότητας εμφάνισης κάποιου γεγονότος απότομης επιτάχυνσης.

➤ Παραβίαση του ανώτατου ορίου ταχύτητας (OverSpeed)

Μαθηματική Σχέση:

$$U = -3,575 + 0,122 * (OBD_SPEED) - 4,316 * (SuddenAcc) - 5,216 * (SuddenDec)$$

Περιγραφή Αποτελεσμάτων:

Από την παραπάνω σχέση προκύπτει ότι με αύξηση της **ταχύτητας του οχήματος (OBD SPEED)** έχουμε μια μικρή αύξηση στην πιθανότητα εμφάνισης παραβίασης του ανώτατου ορίου ταχύτητας ενώ οι άλλες δύο μεταβλητές (**SuddenAcc**, **SuddenDec**) είναι δυσανάλογες με την υπέρβαση του ορίου ταχύτητας. Με συντελεστή -4,316 και -5,216 αντίστοιχα επιφέρουν μείωση της πιθανότητας εμφάνισης παραβίασης του ορίου ταχύτητας με αύξηση της τιμής τους.

➤ Ταχύτητα Οχήματος (OBD SPEED)

Μαθηματική Σχέση:

$$OBD_{SPEED} = 38,022 + 2,375 * (RiskyHour) + 38,385 * (OverSpeed) - 16,032 * (SuddenAcc) - 10,006 * (SuddenDec)$$

Περιγραφή Αποτελεσμάτων:

Όλες οι μεταβλητές που είναι στην παραπάνω σχέση επηρεάζουν σημαντικά την μεταβολή της ταχύτητας, αλλά και εδώ καθοριστικός παράγοντας είναι η **υπερβολική ταχύτητα (OverSpeed)**, η οποία είναι μέγεθος ανάλογο με την ταχύτητα του οχήματος. Παρατηρούμε επίσης ότι η ταχύτητα αυξάνεται, όταν ο οδηγός χρησιμοποιεί το όχημά του κατά τη διάρκεια των **επικίνδυνων ωρών (RiskyHour)**. Αναφορικά με τις **απότομες αλλαγές ταχύτητας (SuddenAcc, SuddenDec)**, με συντελεστή -16,032 και -10,006 αντίστοιχα επιδρούν αρνητικά στην αύξηση της ταχύτητας, πράγμα που σημαίνει ότι στα σημεία όπου ο οδηγός μεταβάλλει απότομα την ταχύτητά του, είναι λιγότερο πιθανό να αναπτύξει και μεγαλύτερη ταχύτητα σε σχέση με τα υπόλοιπα σημεία της οδού.

5.4 Εξήγηση συνολικών αποτελεσμάτων

Παρατηρούμε ότι, παρόλο που το φαινόμενο της παραβίασης του ανώτατου ορίου ταχύτητας είναι συχνό και σημαντικό, ο οδηγός παραμένει προσεκτικός όσον αφορά την κυκλοφοριακή του συμπεριφορά. Δεν προβαίνει σε απότομους ελιγμούς ή σε απότομες αλλαγές ταχύτητας (επιτάχυνση ή επιβράδυνση) όταν βρίσκεται άνω των προβλεπόμενων ορίων για κάθε τύπο οδού δείχνοντας ότι έχει επίγνωση των κινδύνων που μπορεί να προκαλέσει στους άλλους οδηγούς αλλά και στον εαυτό του.

Πιθανότατα, ο οδηγός υπερβαίνει τα όρια ταχύτητας που ορίζει ο ΚΟΚ όχι από αμέλεια ή από απειρία αλλά είτε λόγω βιασύνης είτε λόγω μικρών κυκλοφοριακών φόρτων που του επιτρέπουν να αναπτύξει μεγάλες ταχύτητες. Η διαχείριση της κατάστασης αυτής υποδηλώνει ότι ο οδηγός γνωρίζει αρκετά καλά τους κινδύνους

της οδικής ασφάλειας και δεν κάνει ιδιαίτερα απότομες κινήσεις όταν παραβαίνει τα όρια ταχύτητας.

Η προσοχή, τέλος, του οδηγού φαίνεται και από την ανάπτυξη του μοντέλου για την ταχύτητα του οχήματος καθώς προκύπτει ότι όταν παρατηρείται άυσηση της ταχύτητας, έχουμε μείωση της πιθανότητας εμφάνισης απότομης επιτάχυνσης ή επιβράδυνσης.

5.5 Γενικές Παρατηρήσεις

- ✓ Τα τελικά μοντέλα που παρουσιάζονται παραπάνω για κάθε μεταβλητή είναι αυτά τα οποία **προβλέπουν καλύτερα** την εκάστοτε μεταβλητή. Ειδικότερα για την υπέρβαση του ανώτατου ορίου ταχύτητας (OverSpeed) η πιθανότητα σωστής πρόβλεψης για το σύνολο των προβλέψεων αγγίζει το 97,1%.
- ✓ Μεταξύ των μοντέλων που είχαν ακριβώς το ίδιο ποσοστό πρόβλεψης, επιλέχθηκαν τα μοντέλα τα οποία είχαν **μεγαλύτερο συντελεστή συσχέτισης R^2** (είτε ως Cox & Snell, είτε ως Nagelkerke). Οι τιμές που προκύπτουν στο μοντέλο της υπέρβασης του ορίου ταχύτητας (0,461 και 0,876) κρίνονται κάτι παραπάνω από ικανοποιητικές.
- ✓ Όλοι οι στατιστικοί έλεγχοι ικανοποιούνται για τις μεταβλητές γενικά (Wald πάνω από 1,7 και Sig μικρότερο ή κοντά στο 0,05).
- ✓ Σε κάθε μεταβλητή παρατηρούμε ότι καθοριστικός παράγοντας για την εμφάνισή της είναι η **υπέρβαση του ανώτατου ορίου ταχύτητας** από τον οδηγό και μάλιστα είναι πάντοτε δυσανάλογη με την μεταβλητή αυτή. Στα σημεία, δηλαδή, όπου ο οδηγός τρέχει παραπάνω από όσο επιβάλλει ο ΚΟΚ, είναι λιγότερο πιθανό να εμφανίσει κάποια απότομη αλλαγή ταχύτητας ή να πραγματοποιήσει απότομους ελιγμούς.
- ✓ Στα σημεία όπου παρατηρούνται απότομοι ελιγμοί ή υπέρβαση του ανώτατου ορίου ταχύτητας, σημαντικό ρόλο παίζει και η απότομη αλλαγή ταχύτητας από τον οδηγό, με τα μεγέθη να είναι και σε αυτήν την περίπτωση δυσανάλογα. Πιο συγκεκριμένα, στα σημεία όπου ο οδηγός πραγματοποιεί απότομες αλλαγές ταχύτητας είναι **λιγότερο πιθανό** να πραγματοποιήσει και κάποιον απότομο ελιγμό ή να βρίσκεται σε ταχύτητα άνω του επιτρεπτού.
- ✓ Η μεταβλητή που υποδηλώνει την οδήγηση κατά τις επικίνδυνες ώρες (RiskyHour) συμβάλλει ουσιαστικά **μόνο στην ταχύτητα του οχήματος (OBD_SPEED)**. Για όλες τις υπόλοιπες μεταβλητές, η ώρα οδήγησης δεν παίζει ουσιαστικό ρόλο.

- ✓ Οι υπόλοιπες μεταβλητές όπως ο τύπος οδού που βρίσκεται ο οδηγός (Type of Road) ή η ημέρα οδήγησης (Day) **δεν δείχνουν να επηρεάζουν** στο ελάχιστο την κυκλοφοριακή συμπεριφορά του οδηγού.

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

6.1 Σύνοψη Αποτελεσμάτων

Αντικείμενο της παρούσας Διπλωματικής Εργασίας αποτελεί η διερεύνηση της κυκλοφοριακής συμπεριφοράς του οδηγού με χρήση των διαγνωστικών στοιχείων του οχήματος. Δεδομένου ότι η δυνατότητα αξιοποίησης των στοιχείων αυτών είναι σχετικά πρόσφατη, πρόκειται για ένα θέμα το οποίο δεν έχει μελετηθεί ιδιαίτερα ούτε διεθνώς ούτε στην Ελλάδα, ωστόσο, όπως προκύπτει και από τη βιβλιογραφική ανασκόπηση, τα τελευταία χρόνια χάρη στην εξέλιξη της τεχνολογίας, παρατηρείται μία αυξανόμενη ενασχόληση του επιστημονικού κόσμου γύρω από το συγκεκριμένο ζήτημα.

Η **συλλογή των απαιτούμενων στοιχείων** για τη διερεύνηση της κυκλοφοριακής συμπεριφοράς του οδηγού πραγματοποιήθηκε με μεταφορά των σχετικών στοιχείων από το σύστημα διαγνωστικών του οχήματος, μέσω μίας ειδικής συσκευής που κατέγραφε δεδομένα ανά δευτερόλεπτο.

Μετά από κατάλληλη επεξεργασία και μία σειρά δοκιμών αναπτύχθηκαν μαθηματικά μοντέλα με τη μέθοδο της **διωνυμικής λογιστικής παλινδρόμησης** για τις παρακάτω εξαρτημένες μεταβλητές, από ένα μεγάλο αριθμό στοιχείων ανά δευτερόλεπτο που συλλέχθηκαν από τα διαγνωστικά του οχήματος:

- ✓ Απότομη Δεξιά Στροφή
- ✓ Απότομη Αριστερή Στροφή
- ✓ Απότομη Επιβράδυνση
- ✓ Απότομη Επιτάχυνση
- ✓ Παραβίαση Ανώτατου Ορίου Ταχύτητας

Αναπτύχθηκε και ένα μαθηματικό μοντέλο με τη μέθοδο της **γραμμικής παλινδρόμησης** για την ταχύτητα, καθότι πρόκειται για συνεχή μεταβλητή. Όλα τα μαθηματικά μοντέλα συνοψίζονται στον παρακάτω πίνακα.

Πίνακας 6.1: Συγκεντρωτικός πίνακας αποτελεσμάτων μαθηματικών μοντέλων

Ανεξάρτητες Μεταβλητές	Απότομη Δεξιά Στροφή	Απότομη Αριστερή Στροφή	Απότομη Επιβράδυνση	Απότομη Επιτάχυνση	Παραβίαση Ανώτατου Ορίου Ταχύτητας	Ταχύτητα Οχήματος
Σταθερό	-0.129	-0.046	-0.469	0.826	-3.575	38.022
Ημέρα Οδήγησης						
Επικίνδυνες Ώρες						2.375
Τύπος Οδού						
Απότομη Επιτάχυνση	-1.651	-2.158			-4.316	-16.032
Απότομη Επιβράδυνση	-1.714	-2.022			-5.216	-10.006
Απότομη Δεξιά Στροφή						
Απότομη Αριστερή Στροφή						
Παραβίαση Ανώτατου Ορίου Ταχύτητας	-5.281	-6.083	-5.549	-3.906		38.385
Ταχύτητα Οχήματος				-0.041	0.122	
Κατανάλωση Καυσίμων						
Στροφές ανά λεπτό						

6.2 Συνολικά Συμπεράσματα

Από τα διάφορα στάδια εκπόνησης της παρούσας Διπλωματικής Εργασίας προέκυψε μια σειρά συμπερασμάτων, όπως αυτά συνοψίζονται παρακάτω:

- 1) Τα δεδομένα τα οποία βρίσκονται στο σύστημα διαγνωστικών του οχήματος μπορούν να αποθηκευτούν ηλεκτρονικά σε υπολογιστή και περιέχουν ιδιαιτέρως σημαντικές πληροφορίες οι οποίες, μετά από κατάλληλη επεξεργασία και ανάπτυξη μαθηματικών μοντέλων, μπορούν να χρησιμεύσουν στην εξαγωγή χρήσιμων συμπερασμάτων για την **κυκλοφοριακή συμπεριφορά του οδηγού**.
- 2) Βασικός παράγοντας για την εμφάνιση κάποιου απότομου συμβάντος κατά τη διάρκεια της οδήγησης είναι η **παραβίαση του ορίου ταχύτητας**, όπως αποδείχθηκε από την ανάπτυξη των μαθηματικών μοντέλων. Η μαθηματική σχέση που προκύπτει για κάθε εξαρτημένη μεταβλητή αποδεικνύει την ισχυρή συσχέτιση μεταξύ της παραβίασης του ορίου ταχύτητας και των εξεταζομένων μεταβλητών. Στην περίπτωση της απότομης επιβράδυνσης, το over-speeding αποτελεί τη μοναδική μεταβλητή που επηρεάζει την εμφάνιση ενός τέτοιου συμβάντος.
- 3) Ωστόσο, στα σημεία όπου ο οδηγός βρίσκεται άνω των ορίων ταχύτητας, δείχνει να είναι αρκετά **προσεκτικός** όσον αφορά την οδήγησή του και να μην προβαίνει ούτε σε απότομη επιτάχυνση ή επιβράδυνση ούτε σε απότομους ελιγμούς. Αυτό φαίνεται μέσω των αρνητικών συντελεστών που προκύπτουν στις μαθηματικές σχέσεις για τις μεταβλητές που εξετάστηκαν. Το αποτέλεσμα, πιθανώς, καταδεικνύει την εμπειρία του οδηγού να διαχειριστεί κατάλληλα το όχημά του ακόμα και σε υψηλές ταχύτητες, χωρίς να προβαίνει σε επικίνδυνες ενέργειες.
- 4) Η μόνη περίπτωση όπου η παραβίαση του ορίου ταχύτητας επιδρά θετικά σε κάποια μεταβλητή είναι στην γραμμική ανάλυση παλινδρόμησης και στην **ταχύτητα του οχήματος**. Όταν ο οδηγός υπερβαίνει το όριο ταχύτητας, είναι λογικό η ταχύτητα του οχήματος να αυξάνεται. Στην περίπτωση της ταχύτητας του οχήματος αναπτύχθηκε η **γραμμική παλινδρόμηση** καθώς περιγράφει καλύτερα τις συνεχείς μεταβλητές.
- 5) Εξετάζοντας τους **απότομους ελιγμούς** του οδηγού, το συμπέρασμα που εξάγεται είναι ότι βρίσκονται σε άμεση συσχέτιση με τις απότομες εναλλαγές ταχύτητας και, φυσικά, με την υπέρβαση του ορίου ταχύτητας. Παρόλα αυτά, **το αντίστροφο δεν ισχύει**. Ειδικότερα, οι αναλύσεις έδειξαν ότι η εμφάνιση απότομων επιταχύνσεων ή επιβραδύνσεων κατά τη διάρκεια της οδήγησης δεν μπορούν να θεωρηθούν ως συνάρτηση των απότομων δεξιών ή αριστερών στροφών τις οποίες πραγματοποιεί ο οδηγός στη διαδρομή του.

- 6) Η ώρα οδήγησης δείχνει να επηρεάζει μόνο την ταχύτητα του οχήματος και όχι τη συνολική κυκλοφοριακή συμπεριφορά του οδηγού. Κατά τη διάρκεια **των επικίνδυνων ωρών** (00:00 - 4:00), παρατηρείται αύξηση της ταχύτητας του οχήματος, κάτι το οποίο συμβαίνει –πιθανότατα– λόγω μικρότερου κυκλοφοριακού φόρτου στις οδούς. Παρόλα αυτά, δεν προκύπτει ότι στο συγκεκριμένο χρονικό διάστημα ο οδηγός είτε είναι πιο απρόσεκτος είτε πιο επικίνδυνος καθώς στα υπόλοιπα μαθηματικά μοντέλα, το επίπεδο σημαντικότητας της μεταβλητής των επικίνδυνων ωρών ήταν ιδιαίτερα χαμηλό.
- 7) Οι υπόλοιπες μεταβλητές που υπήρχαν στη βάση δεδομένων και εξετάστηκαν όπως η ημέρα οδήγησης, ο τύπος της οδού, η κατανάλωση καυσίμων και οι στροφές ανά λεπτό **δεν φαίνεται να έχουν ουσιαστική επιρροή** στη συμπεριφορά του οδηγού. Σε όλα τα μαθηματικά μοντέλα που αναπτύχθηκαν, οι συγκεκριμένες μεταβλητές εμφάνιζαν ένα ιδιαίτερα χαμηλό επίπεδο σημαντικότητας, καταδεικνύοντας την μη επίδρασή τους στην επίτευξη του στόχου της παρούσας διπλωματικής εργασίας.
- 8) Η στατιστική επεξεργασία των περισσότερων στοιχείων πραγματοποιήθηκε με τη μέθοδο της **λογιστικής ανάλυσης παλινδρόμησης**, που αποδείχθηκε κατάλληλη για τέτοιου είδους ανάλυση. Η ανάλυση των στοιχείων με αυτή τη μέθοδο οδήγησε στην ανάπτυξη αξιόπιστων μαθηματικών μοντέλων διερεύνησης της κυκλοφοριακής συμπεριφοράς του οδηγού. Η επιλογή της συγκεκριμένης μεθόδου έγινε καθώς οι μεταβλητές προς εξέταση ήταν διακριτές, λαμβάνοντας τις τιμές 0 και 1.
- 9) Τα σαφή και ερμηνεύσιμα αποτελέσματα της παρούσας εργασίας καταδεικνύουν την **καταλληλότητα της μεθόδου ανάλυσης** σχετικά με την διερεύνηση της κυκλοφοριακής συμπεριφοράς του οδηγού μέσω των διαγνωστικών στοιχείων του οχήματος. Η αποτελεσματικότητα της μεθόδου επιτρέπει την εφαρμογή της σε έρευνες με παρόμοιο αντικείμενο έπειτα από κατάλληλη προσαρμογή. Βέβαια, οι μεταβλητές θα πρέπει να επιλέγονται ανάλογα με την εξεταζόμενη περίπτωση.

6.3 Προτάσεις για Βελτίωση της Κυκλοφοριακής Συμπεριφοράς

Σύμφωνα με τα αποτελέσματα και τα συνολικά συμπεράσματα που εξάχθησαν κατά την εκπόνηση της παρούσας Διπλωματικής Εργασίας, επιχειρείται η παράθεση μιας σειράς προτάσεων, οι οποίες ενδεχομένως να συμβάλλουν στη βελτίωση της κυκλοφοριακής συμπεριφοράς των οδηγών.

- ✓ Η παρούσα ανάλυση πρέπει να διευρυνθεί, εντάσσοντας συνεχώς όλο και περισσότερους οδηγούς, οι οποίοι θα ενημερώνονται για τα αποτελέσματά της και θα γνωρίζουν πως συμπεριφέρονται στο δρόμο έτσι ώστε να είναι σε θέση να βελτιωθούν.
- ✓ Απαιτείται ένα σχέδιο δράσης, μέσω **εκστρατειών ενημέρωσης** σε όλα τα μέσα ενημέρωσης και το διαδίκτυο, ώστε να επιτευχθεί η αλλαγή της νοοτροπίας των οδηγών και να αναδειχθούν οι κίνδυνοι που εγκυμονούν από την επικίνδυνη οδήγηση.
- ✓ Πρέπει να υπάρξει μια συντονισμένη αντιμετώπιση του προβλήματος από την πολιτεία αλλά πολύ περισσότερο από τους εκπαιδευτικούς φορείς, έτσι ώστε να δημιουργηθεί το υπόβαθρο για τη δημιουργία υπεύθυνων και υποδειγματικών οδηγών από μικρή ηλικία. Είναι, λοιπόν, απαραίτητη η **ανάπτυξη εκπαιδευτικών προγραμμάτων** αλλά και η ένταξη μαθημάτων οδικής ασφάλειας στα πλαίσια των σχολικών δραστηριοτήτων.
- ✓ Τέλος, θα πρέπει να **ελέγχεται ηλεκτρονικά** η ταχύτητα των οχημάτων (ιδίως στους δρόμους ταχείας κυκλοφορίας) μέσω ειδικών μηχανημάτων, ώστε να γνωρίζουν οι ειδικοί φορείς και η αστυνομία ποιος υπερβαίνει τα νομοθετημένα όρια και να προβαίνουν στις απαραίτητες ενέργειες.

6.4 Προτάσεις για Περαιτέρω Έρευνα

Για την περαιτέρω μελέτη του αντικειμένου της παρούσας Διπλωματικής Εργασίας, ενδιαφέρον θα παρουσίαζε η διερεύνηση των παρακάτω:

- Θα ήταν ακόμα πιο χρήσιμα τα αποτελέσματα μίας ανάλογης έρευνας αλλά έχοντας ένα **πιο ευρύ δείγμα οδηγών**, ώστε να είναι εφικτό να ομαδοποιηθούν οι συμμετέχοντες και να εξαχθούν αποτελέσματα με βάση διαφορετικά στοιχεία των οδηγών, όπως το φύλο, τα χρόνια οδήγησης, η ηλικία κτλ.
- Ιδιαίτερο ενδιαφέρον θα παρουσίαζε μία ανάλυση η οποία θα στηριζόταν σε **ακόμα περισσότερα δεδομένα**, όπως τα χαρακτηριστικά του οχήματος (κινητήριος δύναμη, ίπποι, ηλικία κτλ) ή η κατάσταση του οδοστρώματος στο οποίο κινείται το όχημα.
- Επιπλέον, επειδή η κυκλοφοριακή συμπεριφορά του κάθε οδηγού μεταβάλλεται ανάλογα και με την **ψυχολογική κατάσταση** του καθενός, θα μπορούσε να εξαχθεί η εν λόγω έρευνα αφού πρώτα έχουν εξεταστεί οι συμμετέχοντες από ειδικούς ψυχολόγους.

- Η παρούσα Διπλωματική Εργασία δεν εξέταζε αν ο οδηγός είχε καταναλώσει αλκοόλ πριν χρησιμοποιήσει το όχημά του. Καθώς αυτή η κακή συνήθεια προκαλεί αρκετά ατυχήματα στη χώρα μας καθημερινά, μία **συσχέτιση της κυκλοφοριακής συμπεριφοράς και της κατανάλωσης αλκοόλ** μέσω του παρόντος πλαισίου θα επιδρούσε σημαντικά και θετικά στην κατανόηση και καταπολέμηση αυτού του φαινομένου.
- Εκτός του αλκοόλ, άλλοι παράγοντες που πιθανόν να επηρεάζουν την κυκλοφοριακή συμπεριφορά του οδηγού και έχει νοήμα η περαιτέρω μελέτη και ανάλυσή τους είναι **ο κυκλοφοριακός φόρτος και η χρήση κινητού τηλεφώνου** κατά τη διάρκεια της οδήγησης.
- Τέλος, θα μπορούσε να διερευνηθεί η επιρροή που έχει στην συμπεριφορά του οδηγού **η παρουσία ή μη συνοδηγού** και άλλων ατόμων (ειδικότερα παιδιών) εντός οχήματος.

7. ΒΙΒΛΙΟΓΡΑΦΙΑ

- 1) **“Global Status Report on Road Safety”**, WHO (World Health Organization), 2013
- 2) **“Basic Road Fatalities Figures”**, European Commission, 2012
- 3) **“ PIN Report - Back on track to reach the 2020 target ”**, ETSC (European Transport Safety Council), Τομέας Μεταφορών και Συγκοινωνιακής Υποδομής ΕΜΠ, 2013
- 4) **“ Road fatalities per million population, European Union 2002 – 2011”** , ETSC (European Transport Safety Council), Τομέας Μεταφορών και Συγκοινωνιακής Υποδομής ΕΜΠ, 2013 (<http://www.nrso.ntua.gr/images/stories/data/nrso-eu51-2011.pdf>)
- 5) Φραντζεσκάκης Ι.Μ., Γκόλιας Ι.Κ., Πιτσιάβα-Λατινοπούλου Μ.Χ., **“Κυκλοφοριακή Τεχνική”**, Εκδόσεις Παπασωτηρίου, Αθήνα 2009
- 6) Φραντζεσκάκης Ι.Μ., Γκόλιας Ι.Κ., **“Οδική Ασφάλεια”**, Εκδόσεις Παπασωτηρίου, Αθήνα 1994
- 7) Κοκολάκης Γ., Σπηλιώτης Ι., **“Θεωρία Πιθανοτήτων και Στατιστική με Εφαρμογές”**, Εκδόσεις Συμεών, Αθήνα 2010
- 8) Αρβανίτη Δήμητρα, **“Συγκριτική Ανάλυση των παραγόντων επιρροής της οδήγησης υπό την επήρεια αλκοόλ των οδηγών αυτοκινήτων και μοτοσικλετών”**, Διπλωματική Εργασία, Σχολή Πολιτικών Μηχανικών, Εθνικό Μετσόβιο Πολυτεχνείο, Ιανουάριος 2014
- 9) Ohta, Tohru, and Shouji Nakajima. **"Development of a driving data recorder."** JSAE Review 15.3 (1994): 255-258.
- 10) Nadeem, Tamer, et al. **"TrafficView: traffic data dissemination using car-to-car communication."** ACM SIGMOBILE Mobile Computing and Communications Review 8.3 (2004): 6-19.
- 11) Huang, Yueng-Hsiang, et al. **"Feedback by technology: Attitudes and opinions of truck drivers."** Transportation Research Part F: Traffic Psychology and Behaviour 8.4 (2005): 277-297.
- 12) Toledo, Tomer, Oren Musicant, and Tsippy Lotan. **"In-vehicle data recorders for monitoring and feedback on drivers' behavior."** Transportation Research Part C: Emerging Technologies 16.3 (2008): 320-331.
- 13) Farmer, Charles M., Bevan B. Kirley, and Anne T. McCartt. **"Effects of in-vehicle monitoring on the driving behavior of teenagers."** Journal of Safety Research 41.1 (2010): 39-45.
- 14) Prato, Carlo Giacomo, et al. **"Modeling the behavior of novice young drivers during the first year after licensure."** Accident Analysis & Prevention 42.2 (2010): 480-486.
- 15) Zaldivar, Jorge, et al. **"Providing accident detection in vehicular networks through OBD-II devices and Android-based smartphones."** Local Computer Networks (LCN), 2011 IEEE 36th Conference on. IEEE, 2011.

- 16) Vaiana, Rosolino, et al. "**Driving behavior and traffic safety: an acceleration-based safety evaluation procedure for smartphones.**" Modern Applied Science 8.1 (2014): p88.
- 17) Shichrur, Rachel, Adi Sarid, and Navah Z. Ratzon. "**Determining the sampling time frame for In-Vehicle Data Recorder measurement in assessing drivers.**" Transportation research part C: emerging technologies 42 (2014): 99-106.
- 18) Τομέας Μεταφορών και Συγκοινωνιακής Υποδομής, Ε.Μ.Π., "**Ανάπτυξη στρατηγικού σχεδίου για τη βελτίωση της οδικής ασφάλειας στην Ελλάδα 2011-2020**", Αθήνα 2011