



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ ΚΑΙ ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ ΥΠΟΔΟΜΗΣ

Χωρική Ανάλυση των Οδικών Ατυχημάτων στους Δήμους της Ελλάδας

Διπλωματική Εργασία



Δημήτρης Ανδρέικος

Επιβλέπων: Γιώργος Γιαννής, Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2024

Ευχαριστίες

Ευχαριστώ θερμά τον κ. Γιώργο Γιαννή, Καθηγητή της Σχολής Πολιτικών Μηχανικών Ε.Μ.Π. για την ανάθεση του θέματος, για την πολύτιμη καθοδήγησή του και για την άριστη συνεργασία που είχαμε σε όλα τα στάδια της Διπλωματικής Εργασίας.

Ένα μεγάλο ευχαριστώ οφείλω και στους κ. Δημήτρη Νικολάου και Απόστολο Ζιακόπουλο, ερευνητές και Διδάκτορες Ε.Μ.Π., για τις χρήσιμες παρατηρήσεις και συμβουλές τους και για την άμεση ανταπόκριση σε απορίες που προέκυψαν κατά την διάρκεια εκπόνησης της εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου για την μεγάλη στήριξη που μου παρείχαν, τόσο σε οικονομικό όσο και σε ψυχολογικό επίπεδο, καθ' όλη τη διάρκεια των σπουδών μου.

Αθήνα, Ιούλιος 2024

Δημήτρης Ανδρέικος

Χωρική Ανάλυση των Οδικών Ατυχημάτων στους Δήμους της Ελλάδας

Δημήτρης Ανδρέικος

Επιβλέπων: Γιώργος Γιαννής, Καθηγητής Ε.Μ.Π.

Σύνοψη

Στόχο της παρούσας Διπλωματικής Εργασίας αποτελεί η διερεύνηση της επιρροής του παράγοντα του χώρου στην ανάλυση των οδικών ατυχημάτων, και κατά πόσον οι μέθοδοι ανάλυσης βελτιώνονται με χρήση γεωγραφικά σταθμισμένης παλινδρόμησης. Για τον σκοπό αυτό, συλλέχθηκαν πληθυσμιακά, κοινωνικά, οικονομικά και συγκοινωνιακά δεδομένα και εξετάστηκε η επίδρασή τους στον αριθμό των θανάτων σε οδικά ατυχήματα. Πραγματοποιήθηκε πρωταρχική στατιστική επεξεργασία των δεδομένων καθώς και ομαδοποίησή τους από την οποία προέκυψαν δύο συστάδες Δήμων: μία συστάδα με τους αστικούς, μεγάλους πληθυσμιακά Δήμους, και μία συστάδα με τους περιφερειακούς και επαρχιακούς Δήμους. Στη συνέχεια, και μετά από δοκιμές, αναπτύχθηκαν έξι τελικά μαθηματικά μοντέλα: τρία πρότυπα γραμμικής παλινδρόμησης, για το σύνολο των Δήμων και για κάθε συστάδα ξεχωριστά, και τρία πρότυπα γεωγραφικά σταθμισμένης παλινδρόμησης, για τις αντίστοιχες περιπτώσεις. Με την εφαρμογή και αξιολόγηση των μοντέλων προκύπτει ότι οι σημαντικότεροι παράγοντες που επηρεάζουν τους θανάτους σε οδικά ατυχήματα είναι η ιδιοκτησία ΙΧ και το Ακαθάριστο Εθνικό Προϊόν. Η ιδιοκτησία ΙΧ συσχετίζεται θετικά με τους θανάτους σε οδικά ατυχήματα, καθώς η αύξηση του στόλου οχημάτων οδηγεί σε αυξημένη έκθεση στον κίνδυνο, ενώ το ΑΕΠ συσχετίζεται αρνητικά, καθώς η αύξησή του σχετίζεται με καλύτερο βιοτικό επίπεδο και βελτιωμένα μέτρα οδικής ασφάλειας.

Συμπερασματικά, προκύπτει ότι τα χωρικά σταθμισμένα πρότυπα παράγουν καλύτερα αποτελέσματα σε σχέση με την κλασική γραμμική παλινδρόμηση, όπως αποδεικνύεται από τους στατιστικούς δείκτες R^2 και AIC. Συνεπώς, η εφαρμογή χωρικής μεθόδου ανάλυσης φαίνεται να βελτιώνει την ακρίβεια της εκτίμησης των μαθηματικών προτύπων.

Λέξεις κλειδιά: οδικά ατυχήματα, παράγοντες επιρροής, ομαδοποίηση δεδομένων, γραμμική παλινδρόμηση, χωρική ανάλυση, γεωγραφικά σταθμισμένη παλινδρόμηση, ακρίβεια μαθηματικών προτύπων.

Spatial Analysis of Road Accidents in the Municipalities of Greece

Dimitris Andreikos

Supervisor: George Yannis, Professor N.T.U.A.

Abstract

The objective of the present Diploma Thesis is the examination of the influence of the spatial factor in the analysis of road accidents, as well as the degree of improvement of the analysis methods with the use of a geographically weighted regression. To that end, population, social, economic and transportation data was collected, and its influence on the number of road fatalities was examined. Following an initial processing of the data, their clustering procedure produced two clusters: one with the urban, densely populated Municipalities, and one with the regional and rural Municipalities. Subsequently, and following many tests, six final mathematical models were developed: three linear regression models for all the Municipalities and for each cluster separately, as well as three geographically weighted regression models, respectively. The application and evaluation of the models showed that the most important factors that influence the number of road fatalities are the possession of private vehicles and the Gross Domestic Product. Private vehicle possession correlates positively with the number of road fatalities, as the increase in vehicles leads to an increased exposure to danger, while GDP correlates negatively, as its increase is linked to better life quality and improved road safety measures.

In conclusion, it appears that the spatial models produce improved results when compared to the classic linear regression, as demonstrated by the R^2 and AIC indexes. Therefore, the application of a spatial method of analysis improves the accuracy of the mathematical model estimation.

Keywords: road accidents, influencing factors, data clustering, linear regression, spatial analysis, geographically weighted regression, mathematical model accuracy.

Περίληψη

Αντικείμενο της παρούσας Διπλωματικής Εργασίας είναι η **ανάλυση των οδικών ατυχημάτων στους Δήμους της Ελλάδας με χωρικά στατιστικά μοντέλα**. Πιο συγκεκριμένα, επιδιώκεται η σύγκριση των αποτελεσμάτων των χωρικών μοντέλων με αυτά των πιο κλασικών μεθόδων παλινδρόμησης. Σε κάθε περίπτωση, πραγματοποιείται διερεύνηση σχετικά με τον συνδυασμό παραγόντων που επηρεάζουν τα οδικά ατυχήματα. Στη συνέχεια, με τον στόχο της Διπλωματικής Εργασίας καθορισμένο, ακολούθησε η **βιβλιογραφική ανασκόπηση** προς αναζήτηση ερευνών με παρόμοιο αντικείμενο.

Ακολούθως, χρησιμοποιώντας τις βάσεις δεδομένων της Ελληνικής Στατιστικής Αρχής, πραγματοποιήθηκε η **συλλογή των απαραίτητων στοιχείων** για την στατιστική ανάλυση. Η βάση δεδομένων της Διπλωματικής Εργασίας ξεκινά με την καταγραφή του αριθμού θανάτων σε οδικά ατυχήματα κατά την περίοδο 2016 – 2020, ανά Δήμο. Τα υπόλοιπα δεδομένα περιλαμβάνουν οικονομικά, κοινωνικά και δημογραφικά χαρακτηριστικά για τους Δήμους της Ελλάδας, καθώς και τις γεωγραφικές συντεταγμένες του κέντρου κάθε Δήμου που προέκυψαν από τους διαδικτυακούς χάρτες της Google.

Σε αρχικό στάδιο, η βάση δεδομένων της Διπλωματικής Εργασίας αναπτύχθηκε στο περιβάλλον του Microsoft Excel και στη συνέχεια εισήχθη στο ειδικό στατιστικό λογισμικό R. Με εφαρμογή ενός **αλγορίθμου ιεραρχικής ομαδοποίησης**, οι Δήμοι κατατάχθηκαν σε δύο συστάδες, οι οποίες υπέστησαν ξεχωριστή επεξεργασία. Για κάθε συστάδα, προέκυψαν ορισμένα βασικά στατιστικά αποτελέσματα καθώς και γραφήματα, με σκοπό την καλύτερη κατανόηση των δεδομένων.

Η ουσιαστική **στατιστική ανάλυση** στο πλαίσιο της Διπλωματικής Εργασίας έγινε με το λογισμικό R στο οποίο αναπτύχθηκαν συνολικά έξι μαθηματικά πρότυπα, τα οποία έχουν όλα την ίδια εξαρτημένη μεταβλητή: τον δεκαδικό λογάριθμο του αριθμού των θανάτων σε οδικά ατυχήματα ανά τετραγωνικό χιλιόμετρο. Συγκεκριμένα, αναπτύχθηκαν τρία κλασικά μοντέλα λογαριθμοκανονικής παλινδρόμησης – ένα για το σύνολο των Δήμων, και ένα για κάθε μία από τις συστάδες Δήμων όπως αυτές προέκυψαν από την ομαδοποίηση.

Επιπλέον, αναπτύχθηκαν τρία μοντέλα **γεωγραφικά σταθμισμένης παλινδρόμησης** που λαμβάνουν υπόψη την χωρική γειτονία των Δήμων για την ανάλυσή τους. Επισημαίνεται ότι τα έξι παραπάνω μοντέλα προέκυψαν μετά από μεγάλο αριθμό δοκιμών για διάφορους συνδυασμούς των ανεξάρτητων μεταβλητών.

Αναφορικά με την **σχετική επιρροή** των ανεξάρτητων μεταβλητών στην εξαρτημένη, οι υπολογισμοί έγιναν μόνο για τα τρία κλασικά μοντέλα παλινδρόμησης. Στα πρότυπα αυτά, η σχετική επιρροή προσδιορίζεται μέσω του μεγέθους της ελαστικότητας, και οδηγεί σε μια καλύτερη αντίληψη της σχέσης μεταξύ του αριθμού θανάτων σε οδικά ατυχήματα ανά τετραγωνικό χιλιόμετρο, και των διάφορων ανεξάρτητων μεταβλητών.

Στον παρακάτω **συγκεντρωτικό πίνακα** παρουσιάζονται τα αποτελέσματα των έξι τελικών μαθηματικών μοντέλων της Διπλωματικής Εργασίας. Επισημαίνεται ότι όλες οι τιμές του συντελεστή R^2 αναφέρονται στον προσαρμοσμένο δείκτη (adjusted R^2). Επίσης, οι εκτιμήσεις των συντελεστών στα χωρικά πρότυπα είναι οι διάμεσοι (median) του εκάστοτε εύρους τιμών.

Πρότυπα Λογαριθμοκανονικής Παλινδρόμησης															
Ανεξάρτητες μεταβλητές	Σύνολο Δήμων					Συστάδα 1					Συστάδα 2				
	R ²	AIC	Estimate	t	ei*	R ²	AIC	Estimate	t	ei*	R ²	AIC	Estimate	t	ei*
Ποσοστό απόφοιτων πανεπιστημίου	0.529	524.403	-0.019	-2.806	1.00	0.811	59.663	0.039	3.981	1.00	0.396	396.276	-0.021	-2.404	2.89
Ποσοστό γυναικών			-0.065	-2.709	-1.91			-0.239	-4.488	-3.69			-0.017	-0.685	1.00
Ποσοστό εργαζόμενων			-0.057	-6.903	-2.26			-0.087	-4.448	-1.78			-0.021	-2.152	1.92
Αριθμός νοικοκυριών με ιδιοκτησία ΙΧ			0.00002	6.124	10.83			0.0000093	3.877	7.50			0.000065	4.442	57.54
Ακαθάριστο Εθνικό Προϊόν			-0.00008	-6.955	-11.13			-0.00011	-7.563	-14.09			-0.000066	-3.917	38.32
Ποσοστό ηλικιωμένων									0.04	1.925			1.09	0.036	4.651
Πρότυπα Χωρικής Παλινδρόμησης GWR															
Ανεξάρτητες μεταβλητές	Σύνολο Δήμων				Συστάδα 1				Συστάδα 2						
	R ²	AIC	Estimate (Median)	Εύρος ζώνης	R ²	AIC	Εύρος ζώνης	Estimate (Median)	R ²	AIC	Εύρος ζώνης	Estimate (Median)			
Ποσοστό απόφοιτων πανεπιστημίου	0.681	381.628	0.003	115	0.841	34.461	66	0.033	0.537	311.699	126	0.008			
Ποσοστό γυναικών			-0.081					-0.189				-0.039			
Ποσοστό εργαζόμενων			-0.053					-0.052				-0.04			
Αριθμός νοικοκυριών με ιδιοκτησία ΙΧ			0.000024					0.0000071				0.000084			
Ακαθάριστο Εθνικό Προϊόν			-0.000099					-0.00011				-0.001			
Ποσοστό ηλικιωμένων												0.086	0.013		

Πίνακας: Συγκεντρωτικός πίνακας αποτελεσμάτων μαθηματικών προτύπων

Σε όλα τα στάδια εκπόνησης της παρούσας Διπλωματικής Εργασίας προέκυψε μια σειρά από συμπεράσματα που σχετίζονται με τους στόχους της. Σε αυτό το σημείο παρατίθενται οι απαντήσεις στα διάφορα ερωτήματα που προέκυψαν, χρησιμοποιώντας τα αποτελέσματα των προηγούμενων κεφαλαίων. Συνοπτικά, τα **γενικά συμπεράσματα** της Διπλωματικής Εργασίας είναι:

- Υπάρχει μια σειρά από αρκετούς παράγοντες που μπορούν να επηρεάσουν τον αριθμό των θανάτων σε οδικά ατυχήματα. Ωστόσο, προέκυψε ότι η ιδιοκτησία ΙΧ και το Ακαθάριστο Εθνικό Προϊόν είναι πιο **καθοριστικοί παράγοντες** σε σχέση με δημογραφικά ή εκπαιδευτικά χαρακτηριστικά.
- Για το **σύνολο των Δήμων**, όπως και για τους **Δήμους με τον μεγαλύτερο πληθυσμό**, προέκυψε το αποτέλεσμα ότι η πιο κρίσιμη μεταβλητή είναι το **Ακαθάριστο Εθνικό Προϊόν**. Συγκεκριμένα, εάν αυξηθεί, παρατηρείται μείωση των θανάτων σε οδικά ατυχήματα. Η πιο πιθανή εξήγηση είναι ότι ένα υψηλότερο ΑΕΠ συνδέεται με καλύτερη ποιότητα ζωής, και επιτρέπει την ανάπτυξη πιο αποτελεσματικών μέτρων οδικής ασφάλειας.
- Παρατηρήθηκε ότι το **ποσοστό απόφοιτων πανεπιστημίου** συσχετίζεται θετικά με την εξαρτημένη μεταβλητή στους Δήμους μεγαλύτερου πληθυσμού, ενώ συσχετίζεται αρνητικά στους περιφερειακούς και επαρχιακούς Δήμους, όπως και στο σύνολο των Δήμων. Αυτή η διαφορά μπορεί να εξηγηθεί ως εξής:

Γενικά, έγινε η παραδοχή ότι οι απόφοιτοι πανεπιστημίου βρίσκονται σε ηλικία άνω των 24 ετών και έχουν αναπτύξει πιο ώριμη οδική συμπεριφορά, κάτι που εξηγεί το αρνητικό πρόσημο του συντελεστή. Ωστόσο, στους μεγάλους αστικούς Δήμους οι μετακινήσεις είναι αυξημένες, κάτι που αυξάνει την έκθεση στον κίνδυνο, και κατά

συνέπεια και τους θανάτους σε οδικά ατυχήματα. Επομένως, υπάρχει μια θετική συσχέτιση των απόφοιτων πανεπιστημίου με την εξαρτημένη μεταβλητή σε αυτούς τους Δήμους.

4. Για τους **περιφερειακούς Δήμους**, βρέθηκε ότι ο πιο καθοριστικός παράγοντας είναι ο **αριθμός νοικοκυριών με ιδιοκτησία ΙΧ**. Συγκεκριμένα, η αύξηση της εν λόγω μεταβλητής προκαλεί αύξηση των θανάτων σε οδικά ατυχήματα. Το αποτέλεσμα είναι λογικό καθώς η αυξημένη ιδιοκτησία ΙΧ, οδηγεί σε μεγαλύτερο στόλο οχημάτων και κατά συνέπεια μεγαλύτερη έκθεση στον κίνδυνο.
5. Παρατηρήθηκε ότι το **ποσοστό του γυναικείου πληθυσμού** γενικά δεν έχει μεγάλη επιρροή στον αριθμό των θανάτων σε οδικά ατυχήματα, ενώ στην περίπτωση των περιφερειακών Δήμων είναι η μεταβλητή με την χαμηλότερη επιρροή. Σε όλες τις περιπτώσεις, η συσχέτιση με την εξαρτημένη μεταβλητή είναι αρνητική, κάτι που είναι λογικό καθώς οι γυναίκες οδηγούν πιο υπεύθυνα σε σχέση με τους άνδρες.
6. Το **ποσοστό των εργαζομένων** είναι μια μεταβλητή με χαμηλή επιρροή στην εξαρτημένη, ενώ τόσο στο σύνολο των Δήμων, όσο και χωριστά στους αστικούς και περιφερειακούς Δήμους, η συσχέτιση είναι αρνητική. Η εξήγηση για αυτό το αποτέλεσμα είναι ότι η αύξηση των απασχολούμενων σε εργασία οδηγεί σε καλύτερο βιοτικό επίπεδο, και συνεπώς σε πιο ενημερωμένους και υπεύθυνους οδηγούς, κάτι που προκαλεί μείωση των θανάτων σε οδικά ατυχήματα.
7. Γενικά, φαίνεται ότι σε κάθε περίπτωση η χωρική παλινδρόμηση παράγει ένα πρότυπο με υψηλότερο R^2 και χαμηλότερο AIC από τους αντίστοιχους δείκτες της λογαριθμοκανονικής. Συνεπώς, η **εφαρμογή της χωρικής μεθόδου** στην ανάλυση των οδικών ατυχημάτων βελτίωσε την ακρίβεια των μαθηματικών μοντέλων.
8. Ανάμεσα στο πρότυπο λογαριθμοκανονικής παλινδρόμησης για τους Δήμους με τον μεγαλύτερο πληθυσμό και σε αυτό για τους περιφερειακούς Δήμους, παρατηρείται μια σημαντική απόκλιση στον **συντελεστή R^2** , δηλαδή στην προσαρμογή του προτύπου στα δεδομένα. Μια πιθανή εξήγηση είναι η μεγάλη διαφορά στο πλήθος των Δήμων μεταξύ των δύο περιπτώσεων.

Οι πιο αστικοποιημένοι Δήμοι είναι λιγότεροι, οπότε το μοντέλο προσαρμόζεται πιο εύκολα στα δεδομένα, οδηγώντας σε πολύ ικανοποιητικό δείκτη R^2 . Αντιθέτως, η συστάδα των επαρχιακών και περιφερειακών Δήμων περιέχει πολύ περισσότερους, που πιθανώς παρουσιάζουν αποκλίσεις στα δεδομένα, κάτι που οδηγεί σε πιο δύσκολη προσαρμογή του προτύπου – δηλαδή χαμηλότερο δείκτη R^2 .

9. Οι στατιστικές μέθοδοι που χρησιμοποιήθηκαν στην ανάλυση είναι η **λογαριθμοκανονική παλινδρόμηση** και η **γεωγραφικά σταθμισμένη παλινδρόμηση**, οι οποίες κρίθηκαν κατάλληλες για την παρούσα Διπλωματική Εργασία. Τα μαθηματικά μοντέλα που προέκυψαν είναι αξιόπιστα και περιγράφουν ικανοποιητικά την εξάρτηση των θανάτων σε οδικά ατυχήματα από τις ανεξάρτητες μεταβλητές.

Περιεχόμενα

Κεφάλαιο 1: Εισαγωγή	1
1.1 Γενική Ανασκόπηση	1
1.2 Στόχος Διπλωματικής Εργασίας	4
1.3 Μεθοδολογία	4
1.4 Δομή Διπλωματικής Εργασίας	5
Κεφάλαιο 2: Βιβλιογραφική ανασκόπηση	7
2.1 Εισαγωγή	7
2.2 Έρευνες	7
2.3 Σύνοψη	17
Κεφάλαιο 3: Θεωρητικό Υπόβαθρο	18
3.1 Εισαγωγή	18
3.2 Βασικές Έννοιες Στατιστικής	18
3.3 Συσχέτιση μεταβλητών & συντελεστής συσχέτισης	19
3.4 Βασικές Στατιστικές Κατανομές	20
3.4.1 Κανονική Κατανομή (Gauss)	20
3.4.2 Κατανομή Poisson	20
3.4.3 Αρνητική Διωνυμική Κατανομή (Negative Binomial)	21
3.5 Βασικά Μαθηματικά Πρότυπα	21
3.5.1 Γραμμική Παλινδρόμηση (Linear Regression)	21
3.5.2 Λογαριθμοκανονική Παλινδρόμηση (Log-normal Regression)	22
3.6 Γεωγραφικά Σταθμισμένη Παλινδρόμηση (GWR)	23
3.7 Ομαδοποίηση δεδομένων (clustering)	24
3.8 Στατιστικοί έλεγχοι & κριτήρια επιλογής	24
Κεφάλαιο 4: Συλλογή & Επεξεργασία Στοιχείων	28
4.1 Εισαγωγή	28
4.2 Συλλογή Δεδομένων	28
4.3 Βάση δεδομένων Διπλωματικής Εργασίας	29
4.4 Ομαδοποίηση δεδομένων (clustering)	31
4.5 Περιγραφική Στατιστική	35
4.5.1 Στατιστικοί υπολογισμοί & γραφήματα	35
4.5.2 Παρατηρήσεις	39
Κεφάλαιο 5: Εφαρμογή Μεθοδολογίας	41
5.1 Εισαγωγή	41
5.2 Συνολική επεξεργασία	41
5.2.1 Συσχέτιση μεταβλητών	41
5.2.2 Ανάπτυξη μαθηματικού προτύπου για το σύνολο των Δήμων	43

5.2.3 Ποιότητα προτύπου & σχολιασμός.....	44
5.2.4 Σχετική επιρροή μεταβλητών.....	45
5.3 Επεξεργασία ανά συστάδα	46
5.3.1 Πρότυπο 1 ^{ης} Συστάδας Δήμων	46
5.3.1.1 Ανάπτυξη μαθηματικού προτύπου	46
5.3.1.2 Ποιότητα προτύπου & σχολιασμός.....	47
5.3.1.3 Σχετική επιρροή μεταβλητών.....	48
5.3.2 Πρότυπο 2 ^{ης} Συστάδας Δήμων	49
5.3.2.1 Ανάπτυξη μαθηματικού προτύπου	49
5.3.2.2 Ποιότητα προτύπου & σχολιασμός.....	50
5.3.2.3 Σχετική επιρροή μεταβλητών.....	51
5.4 Χωρική Ανάλυση	51
5.4.1 Ανάπτυξη χωρικού μοντέλου για το σύνολο των Δήμων.....	52
5.4.2 Ποιότητα & σχολιασμός συνολικού χωρικού προτύπου.....	53
5.4.3 Ανάπτυξη χωρικού μοντέλου για τους Δήμους της 1 ^{ης} συστάδας.....	54
5.4.4 Ποιότητα & σχολιασμός χωρικού προτύπου 1 ^{ης} συστάδας.....	55
5.4.5 Ανάπτυξη χωρικού μοντέλου για τους Δήμους της 2 ^{ης} συστάδας.....	55
5.4.6 Ποιότητα & σχολιασμός χωρικού προτύπου 2 ^{ης} συστάδας.....	56
Κεφάλαιο 6: Συμπεράσματα	57
6.1 Σύνοψη αποτελεσμάτων	57
6.2 Συνολικά συμπεράσματα	58
6.3 Προτάσεις για την μείωση των θανάτων σε οδικά ατυχήματα.....	60
6.4 Προτάσεις για περαιτέρω έρευνα	60
Κεφάλαιο 7: Βιβλιογραφικές Αναφορές	62

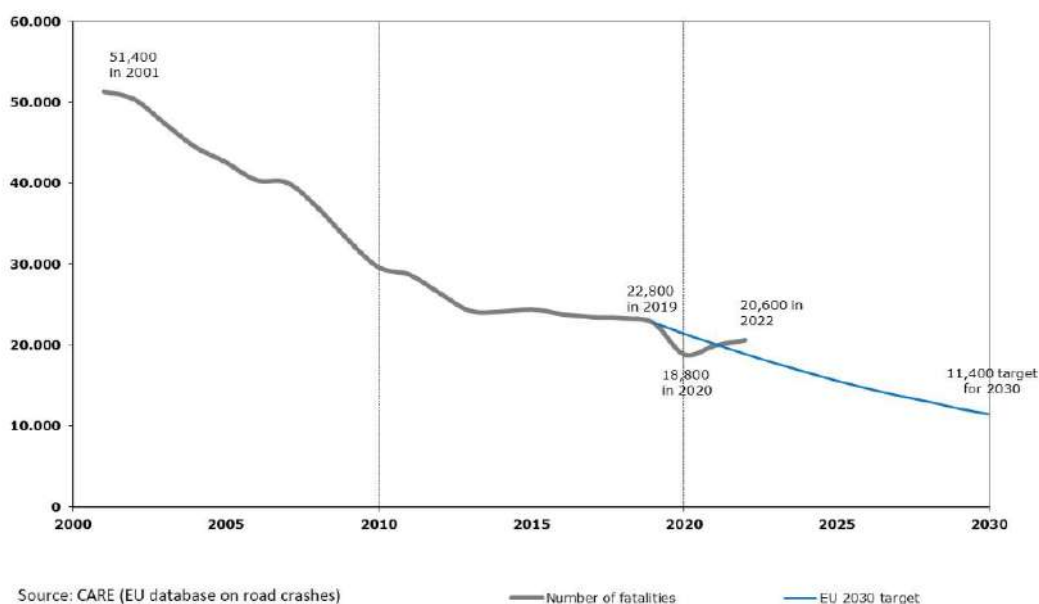
Κεφάλαιο 1: Εισαγωγή

1.1 Γενική Ανασκόπηση

Σε μία σύγχρονη κοινωνία, ο **περιορισμός των οδικών ατυχημάτων** και η μείωση των επιπτώσεών τους αποτελούν δείκτες προόδου και υποστήριξης ενός συγχωρητικού οδικού περιβάλλοντος που έχει ως προτεραιότητα την προστασία των πολιτών. Τα οδικά ατυχήματα επηρεάζονται σε διαφορετικό βαθμό από μια σειρά εξωγενών παραγόντων όπως κοινωνικά και οικονομικά χαρακτηριστικά, κυκλοφοριακές συνθήκες, καθώς και αλλαγές στην συγκοινωνιακή πολιτική. Η συνεχώς αυξανόμενη ανάγκη αποτροπής των οδικών ατυχημάτων, και μείωσης των θανάτων εξαιτίας τους, απαιτούν την λεπτομερή μελέτη των αιτίων των ατυχημάτων, της συσχέτισής τους με διάφορα χαρακτηριστικά, και την επιλογή των κατάλληλων μέτρων οδικής ασφάλειας.

Στη σημερινή εποχή, σε κάθε κράτος τα οδικά ατυχήματα αποτελούν ένα από τα **μεγαλύτερα αίτια θανάτου** και συνεπώς παραμένουν ένα σοβαρό πρόβλημα. Σύμφωνα με την Ευρωπαϊκή Επιτροπή, 20640 άνθρωποι έχασαν τη ζωή τους σε οδικά ατυχήματα στην Ευρωπαϊκή Ένωση το 2022, μια αύξηση της τάξης του 4% σε σχέση με το 2021, καθώς η κυκλοφορία επανήλθε σε φυσιολογικά επίπεδα μετά την πανδημία της Covid-19. Σε σχέση με το 2019 (τελευταίο έτος πριν την πανδημία) καταγράφηκε μια μείωση των θανάτων κατά 9% που, ωστόσο, δεν θεωρείται επαρκής. Στόχος της Ευρωπαϊκής Ένωσης είναι έως το 2030 ο αριθμός των θανάτων σε οδικά ατυχήματα να έχει μειωθεί στο μισό.

Στις αστικές περιοχές, παρατηρήθηκε ότι οι **ευάλωτοι χρήστες της οδού** (πεζοί, ποδηλάτες και μοτοσικλετιστές) αντιπροσωπεύουν σχεδόν το 70% των θανάτων σε οδικά ατυχήματα. Ανησυχητικό είναι το γεγονός ότι οι ποδηλάτες παραμένουν η κατηγορία όπου οι θάνατοι δεν έχουν μειωθεί σημαντικά μέσα στην τελευταία δεκαετία, παρά τη βελτίωση των αστικών υποδομών και την προώθηση της βιώσιμης κινητικότητας.

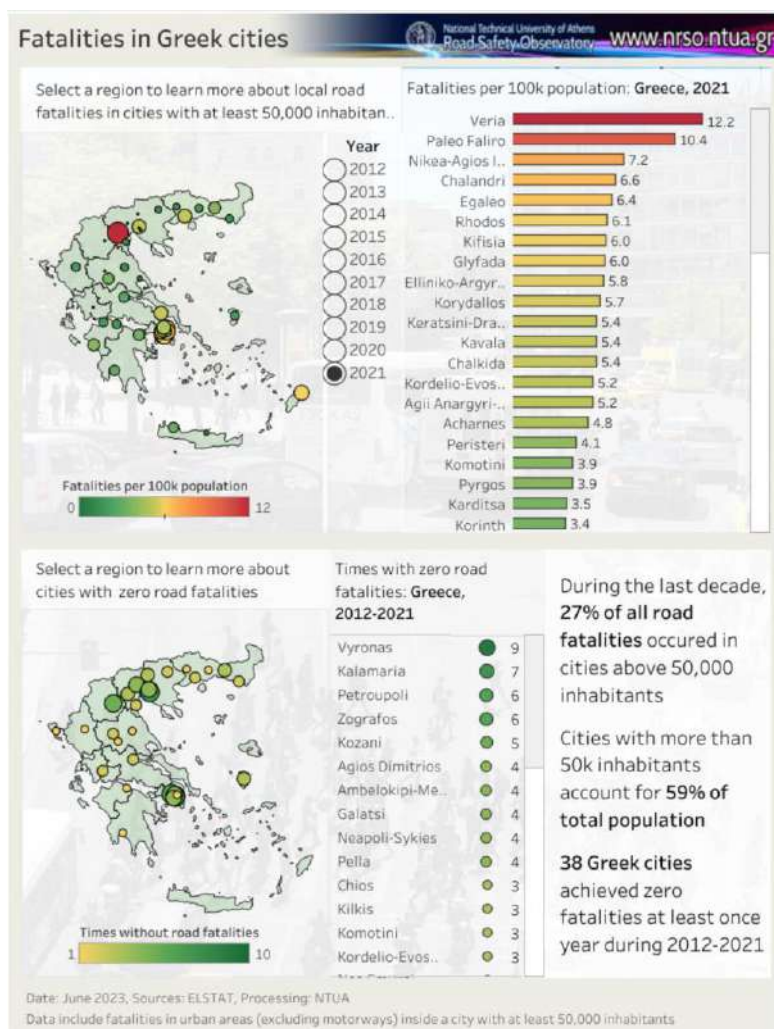


Διάγραμμα 1.1: Θάνατοι σε οδικά ατυχήματα στην Ευρωπαϊκή Ένωση

Απώτερος στόχος της Ευρωπαϊκής Ένωσης είναι ο μηδενισμός των θανάτων εξαιτίας οδικών ατυχημάτων μέχρι το 2050 (“Vision Zero”).

Σύμφωνα με την **Ελληνική Στατιστική Αρχή**, για την Ελλάδα συγκεκριμένα, κατά το έτος 2021, καταγράφηκαν 624 νεκροί σε οδικά ατυχήματα με περίπου ίση κατανομή ανάμεσα σε κατοικημένες και μη κατοικημένες περιοχές. Καταγράφηκαν επίσης 610 βαριά τραυματίες με τα δύο τρίτα αυτών να είναι σε κατοικημένες περιοχές. Για την ίδια χρονιά, σε όλη την έκταση της χώρας καταγράφηκαν 74 θανατηφόρα ατυχήματα που περιλάμβαναν πεζούς. Παρατηρείται λοιπόν ότι οι ευάλωτοι χρήστες του οδικού περιβάλλοντος παραμένουν γενικά εκτεθειμένοι σε κίνδυνο, παρά τη συνολική βελτίωση των επιπέδων οδικής ασφάλειας.

Ο παρακάτω χάρτης κατασκευάστηκε από το **Παρατηρητήριο Οδικής Ασφάλειας** του Εθνικού Μετσόβιου Πολυτεχνείου (NRSO – NTUA) και απεικονίζει μια βαθμολόγηση ελληνικών Δήμων, ως προς τα οδικά ατυχήματα σε αστικές περιοχές, κατά την δεκαετία 2012 – 2021.



Εικόνα 1.1: Βαθμολόγηση ελληνικών Δήμων ως προς τα οδικά ατυχήματα

Όπως φαίνεται και στην παραπάνω εικόνα, η κατάσταση έχει παρουσιάσει βελτίωση, ενώ υπήρξαν Δήμοι με πολλαπλές χρονιές μηδενικών θανάτων σε οδικά ατυχήματα, ακόμα και σε αστικές περιοχές της μητροπολιτικής περιοχής της Αθήνας, όπως ο Βύρωνας και

ο Ζωγράφος. Ωστόσο, τα οδικά ατυχήματα παραμένουν ένα σημαντικό πρόβλημα, ενώ παρατηρήθηκε ότι τουλάχιστον το ένα τέταρτο των θανάτων καταγράφηκε σε περιοχές με πάνω από 50000 κατοίκους. Το Παρατηρητήριο Οδικής Ασφάλειας, σύμφωνα με δεδομένα της **Ευρωπαϊκής Επιτροπής**, προχώρησε και στην κατασκευή διαγράμματος κατάταξης των κρατών – μελών της Ευρωπαϊκής Ένωσης ως προς τους θανάτους ανά τύπο ατυχήματος για το 2021.

Fatalities per million population: All EU countries



Διάγραμμα 1.2: Κατάταξη κρατών – μελών της Ε.Ε. για τα οδικά ατυχήματα

Παρατηρείται ότι, στην πλειοψηφία των περιπτώσεων, η Ελλάδα βρίσκεται **κάτω από τον ευρωπαϊκό μέσο όρο** και συνήθως στις πέντε με έξι τελευταίες θέσεις. Είναι λοιπόν προφανές ότι ο τομέας της οδικής ασφάλειας έχει μεγάλο περιθώριο ανάπτυξης στην χώρα μας.

Τα οικονομικά χαρακτηριστικά ενός κράτους, όπως το εισόδημα και το ΑΕΠ (GDP), επηρεάζουν σημαντικά την οδική ασφάλεια (Yannis et al., 2014). Παρομοίως, κοινωνικές συνθήκες (Nikolaou et al., 2021), μήκος διαδρομών και σύνθεση κυκλοφορίας αποτελούν εξίσου σημαντικούς παράγοντες. Η επιρροή των **γειτονικών γεωγραφικά περιοχών** στα οδικά ατυχήματα, ωστόσο, παραμένει ένα αντικείμενο με περιορισμένο ερευνητικό ιστορικό. Στο πλαίσιο ενός ευρύτερου συστήματος μελέτης και ελαχιστοποίησης των οδικών ατυχημάτων, και δημιουργίας ενός συγχωρητικού οδικού περιβάλλοντος, η χρήση μεθόδων χωρικής ανάλυσης γίνεται όλο και πιο απαιτητική.

1.2 Στόχος Διπλωματικής Εργασίας

Στόχος της παρούσας Διπλωματικής Εργασίας είναι η **χωρική ανάλυση των οδικών ατυχημάτων στους Δήμους της Ελλάδας**.

Συγκεκριμένα, επιδιώκεται να εξεταστεί ο βαθμός στον οποίο η γεωγραφική γειτονία δύο ή περισσότερων Δήμων επηρεάζει άμεσα τα οδικά ατυχήματα και τη συσχέτισή τους με κοινωνικούς, οικονομικούς και συγκοινωνιακούς δείκτες. Επιπλέον, εξετάζεται το κατά πόσον βελτιώνονται τα στατιστικά μοντέλα μετά την εφαρμογή της χωρικής μεθόδου ανάλυσης.

Προκειμένου να επιτευχθεί ο παραπάνω στόχος, χρησιμοποιήθηκαν οι βάσεις δεδομένων της Ελληνικής Στατιστικής Αρχής (ΕΛΣΤΑΤ) για την καταγραφή των αρχικών δεδομένων όπως το φύλο, η ηλικία, το επίπεδο εκπαίδευσης, η ιδιοκτησία ΙΧ και το ΑΕΠ. Σε πρώτο στάδιο, η ανάλυση γίνεται με τα παραδοσιακά στατιστικά μοντέλα παλινδρόμησης προκειμένου να εντοπιστούν οι ανεξάρτητες μεταβλητές που επηρεάζουν πιο ενεργά τα οδικά ατυχήματα. Τα μοντέλα αξιολογούνται και κατά συνέπεια οριστικοποιούνται τα καταλληλότερα. Σε δεύτερο στάδιο, αναπτύσσονται τα μοντέλα χωρικής ανάλυσης και συγκρίνονται με τα αρχικά. Η μέθοδος που χρησιμοποιείται για την χωρική ανάλυση είναι αυτή της **Γεωγραφικά Σταθμισμένης Παλινδρόμησης** (Geographically Weighted Regression – GWR).

Ο απώτερος σκοπός είναι να προκύψουν **μαθηματικά πρότυπα** που συσχετίζουν ικανοποιητικά ορισμένες ανεξάρτητες μεταβλητές με τα δεδομένα των οδικών ατυχημάτων, καθώς και αντίστοιχα πρότυπα χωρικής ανάλυσης. Τα μοντέλα που θα αναπτυχθούν, καθώς και τα συμπεράσματα που θα προκύψουν από αυτά, αναμένεται να οδηγήσουν σε μια καλύτερη κατανόηση της επιρροής των χωρικών μεθόδων στην ανάλυση των οδικών ατυχημάτων.

1.3 Μεθοδολογία

Σε αυτή την ενότητα παρουσιάζεται επιγραμματικά η μεθοδολογία που ακολουθήθηκε για την εκπόνηση της Διπλωματικής Εργασίας.

Αρχικά, καθορίστηκε ο **στόχος** έτσι ώστε να προσδιοριστεί η πορεία της εργασίας. Ύστερα, ακολουθεί η **βιβλιογραφική ανασκόπηση**, όπου αναζητήθηκαν έρευνες με παρόμοιο αντικείμενο – αυτό δηλαδή της χωρικής ανάλυσης οδικών ατυχημάτων – με σκοπό να αντληθούν ορισμένες χρήσιμες παρατηρήσεις που προκύπτουν από τη διεθνή βιβλιογραφία.

Στη συνέχεια, πραγματοποιήθηκε **συλλογή και επεξεργασία των στοιχείων**. Τα στατιστικά δεδομένα της Διπλωματικής Εργασίας αντλήθηκαν από την ΕΛΣΤΑΤ, ενώ οι γεωγραφικές συντεταγμένες των δήμων προέρχονται από τους διαδικτυακούς χάρτες Google Maps που παρέχει η Google. Το τμήμα αυτό της Διπλωματικής Εργασίας μπορεί να διακριθεί στις παρακάτω φάσεις:

Αρχικά, δημιουργήθηκε μια **βάση δεδομένων** σε ένα αρχείο Microsoft Excel προκειμένου να καταγραφούν όλες οι μεταβλητές που θα χρησιμοποιηθούν στην ανάλυση. Με απλές μαθηματικές πράξεις προκύπτουν ορισμένες δευτερογενείς μεταβλητές που είναι

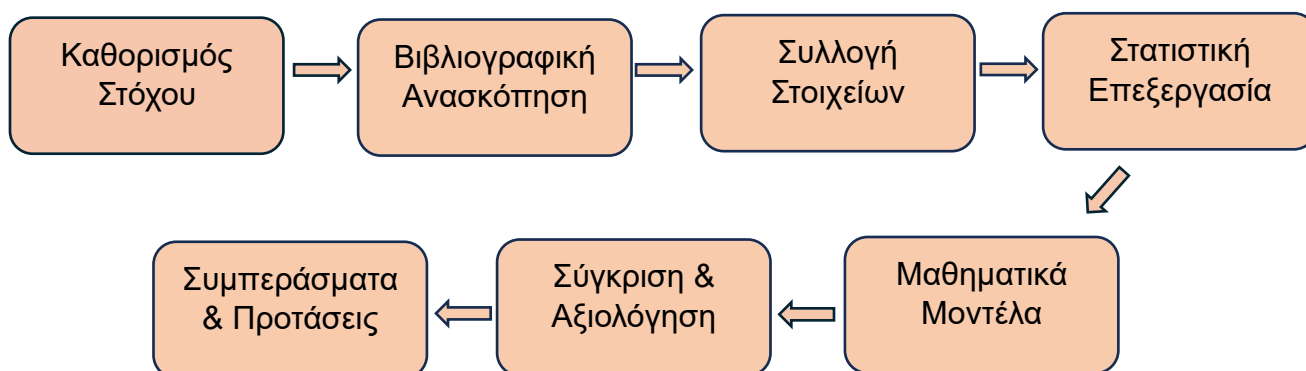
απαραίτητες για την προσαρμογή των στατιστικών μοντέλων. Η βάση δεδομένων στη συνέχεια εισάγεται στο περιβάλλον του **λογισμικού στατιστικής ανάλυσης R**, προκειμένου να υποστεί περαιτέρω επεξεργασία.

Σε δεύτερη φάση, τα δεδομένα εισάγονται σε έναν αλγόριθμο **ιεραρχικής ομαδοποίησης** (hierarchical clustering) και διαχωρίζονται σε δύο διακριτές συστάδες, οι οποίες διατηρούν εσωτερική ομοιομορφία ως προς τα χαρακτηριστικά τους. Τέλος, για κάθε συστάδα ξεχωριστά, εξάγονται με τη βοήθεια του Excel ορισμένα **περιγραφικά στατιστικά στοιχεία** που οδηγούν σε μια καλύτερη κατανόηση των δεδομένων, προτού ακολουθήσει η πιο προηγμένη ανάλυσή τους.

Στο επόμενο στάδιο της Διπλωματικής Εργασίας, αναπτύσσονται τα **μαθηματικά μοντέλα** παλινδρόμησης στο λογισμικό R, και με χρήση ορισμένων κριτηρίων επιλέγονται τα καταλληλότερα. Στη συνέχεια, εφαρμόζεται η **χωρική μέθοδος GWR** και τα μοντέλα που προκύπτουν αξιολογούνται ως προς τη βελτίωση των αποτελεσμάτων σε σχέση με την παραδοσιακή μέθοδο παλινδρόμησης.

Τέλος, παρατίθενται τα **συμπεράσματα** που προέκυψαν από το σύνολο της διαδικασίας της Διπλωματικής Εργασίας και διατυπώνονται προτάσεις για περαιτέρω έρευνα πάνω στο αντικείμενο της χωρικής ανάλυσης οδικών ατυχημάτων.

Παρακάτω φαίνεται το διάγραμμα ροής που απεικονίζει τη μεθοδολογία της Διπλωματικής Εργασίας.



1.4 Δομή Διπλωματικής Εργασίας

Παρακάτω παρουσιάζεται η δομή της Διπλωματικής Εργασίας με τα επιμέρους κεφάλαια που την απαρτίζουν.

Το **Κεφάλαιο 1** αποτελεί την εισαγωγή στο επιστημονικό πεδίο της οδικής ασφάλειας και των ατυχημάτων. Επιπλέον, καθορίζεται ο στόχος της εργασίας και παρουσιάζεται η μεθοδολογία που ακολουθήθηκε.

Το **Κεφάλαιο 2** αποτελεί τη βιβλιογραφική ανασκόπηση όπου παρουσιάζονται διεθνείς μελέτες με παρόμοιο αντικείμενο, και καταγράφονται ορισμένα χρήσιμα συμπεράσματα.

Το **Κεφάλαιο 3** αποτελεί το θεωρητικό υπόβαθρο της Διπλωματικής Εργασίας και βασίζεται στην ανάλυση δεδομένων. Συγκεκριμένα, αναφέρονται τα χαρακτηριστικά των

μαθηματικών προτύπων, και επιπλέον παρουσιάζονται οι απαραίτητοι στατιστικοί έλεγχοι που οδήγησαν στην επιλογή των κατάλληλων μοντέλων. Επίσης, επεξηγείται το θεωρητικό υπόβαθρο της μεθόδου GWR που χρησιμοποιήθηκε για τη χωρική ανάλυση των οδικών ατυχημάτων.

Στο **Κεφάλαιο 4** παρουσιάζεται η διαδικασία συλλογής και επεξεργασίας των δεδομένων που χρησιμοποιήθηκαν στα μαθηματικά μοντέλα. Αρχικά, περιγράφεται η μέθοδος συλλογής στοιχείων και έπειτα η ανάλυσή τους με χρήση του προγράμματος Microsoft Excel και του στατιστικού λογισμικού R.

Το **Κεφάλαιο 5** αποτελεί την εφαρμογή της μεθοδολογίας με διαδοχικές δοκιμές και στατιστικούς ελέγχους. Επεξηγείται η σχέση της εξαρτημένης μεταβλητής με τις ανεξάρτητες και επιπλέον, παρουσιάζονται αναλυτικά τα αποτελέσματα των τελικών μαθηματικών μοντέλων σε μορφή πινάκων.

Στο **Κεφάλαιο 6** παρουσιάζονται τα τελικά συμπεράσματα που προέκυψαν από τα προηγούμενα κεφάλαια, και συγκεκριμένα από την ανάλυση των στοιχείων και την εξαγωγή των τελικών μαθηματικών προτύπων. Αναφέρονται, επίσης, προτάσεις για την μείωση των θανάτων σε οδικά ατυχήματα, καθώς και προτάσεις για περαιτέρω έρευνα.

Στο **Κεφάλαιο 7** παρατίθενται οι βιβλιογραφικές αναφορές οι οποίες χρησιμοποιήθηκαν για την εκπόνηση της Διπλωματικής Εργασίας.

Κεφάλαιο 2: Βιβλιογραφική ανασκόπηση

2.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζεται η βιβλιογραφική ανασκόπηση που πραγματοποιήθηκε στο πλαίσιο της Διπλωματικής Εργασίας. Συγκεκριμένα, παρατίθενται έρευνες που αφορούν στη χωρική ανάλυση οδικών ατυχημάτων, παρουσιάζεται η μεθοδολογία που η κάθε έρευνα χρησιμοποίησε, καθώς και τα συμπεράσματα στα οποία κατέληξε. Μέσω αυτής της ανασκόπησης καθορίστηκε και το ακριβές αντικείμενο της Διπλωματικής Εργασίας και η κατάλληλη μέθοδος για την πραγματοποίησή της.

Παρακάτω φαίνονται οι έρευνες, και αναλύονται με αυτή τη σειρά:

- A review of spatial approaches in road safety (Ziakopoulos & Yannis, 2020)
- Spatial analysis of fatal and injury crashes in Pennsylvania (Aguero-Valverde & Jovanis, 2006)
- Spatial analysis of road crashes involving vulnerable road users in support of road safety management strategies (Bassani et al., 2020)
- Congestion and safety: A spatial analysis of London (Noland & Quddus, 2005)
- Investigating articulated heavy – vehicle crashes in Western Australia using a spatial approach (Gudes et al., 2017)
- Spatial and temporal visualisation techniques for crash analysis (Plug et al., 2011)
- Bayesian spatial joint modelling of traffic crashes on an urban road network (Zeng & Huang, 2014)
- Spatial analysis of harsh driving behaviour events in urban networks using high – resolution smartphone and geometric data (Ziakopoulos, 2021)
- Mortality from road traffic accidents in Switzerland: longitudinal and spatial analyses (Spoerri et al., 2011)
- Multivariate spatial models of excess crash frequency at area level: case of Costa Rica (Aguero-Valverde, 2013)

2.2 Έρευνες

Σε αυτή την ενότητα παρουσιάζονται οι παραπάνω έρευνες αναλυτικά με σκοπό τον έλεγχο εάν τα αποτελέσματα της Διπλωματικής Εργασίας συμφωνούν με παρόμοια της διεθνούς βιβλιογραφίας.

Ο καθηγητής του Εθνικού Μετσόβιου Πολυτεχνείου κ. Γιαννής και ο ερευνητής κ. Ζιακόπουλος πραγματοποίησαν μια έρευνα το (2020) **σχετικά με τις μεθόδους χωρικής ανάλυσης** στην μελέτη της οδικής ασφάλειας, εξετάζοντας την υπάρχουσα βιβλιογραφία που αφορά τις μεθόδους αυτές (Ziakopoulos & Yannis, 2020). Συγκεκριμένα, αναλύονται οι διαφορετικές προσεγγίσεις μοντελοποίησης και παρουσιάζονται σχετικές παράμετροι σχεδιασμού, όπως η κυκλοφορία και το οδικό περιβάλλον. Τέλος γίνεται μια σύγκριση των λειτουργικών και οικονομετρικών προσεγγίσεων πάνω στην εφαρμογή των χωρικών μοντέλων.

Αρχικά, πραγματοποιήθηκε μια ανασκόπηση μελετών που αφορούν την οδική ασφάλεια υπό όρους χωρικής ανάλυσης σε επίπεδο διασταύρωσης, ζώνης και περιφέρειας. Οι ανεξάρτητες παράμετροι που χρησιμοποιήθηκαν περιλαμβάνουν ταχύτητα, κυκλοφοριακό φόρτο, μήκος διαδρομής, καμπυλότητα χάραξης, αριθμό και πλάτος λωρίδων, μήκος οδοστρώματος, πυκνότητα πληθυσμού, ηλικία, εισόδημα, ποσοστό απασχόλησης και συντελεστή χρήσης γης. Οι εξαρτημένες μεταβλητές των μοντέλων είναι ο τύπος ατυχήματος, η συχνότητα και το ποσοστό σφαλμάτων καθώς και η σοβαρότητα του ατυχήματος.

Οι **περιοχές ανάλυσης** που εξετάζονται στις έρευνες είναι:

Διασταυρώσεις, αστικά & υπεραστικά οδικά τμήματα, τμήματα αυτοκινητοδρόμων (τοπικό επίπεδο / διασταύρωσης), γειτονίες, περιοχές hotspots, γεωγραφικές ενότητες (επίπεδο ζώνης), πόλεις, κομητείες, νομοί, δήμοι, μητροπολιτικές περιοχές καθώς και ολόκληρα κράτη (περιφερειακό επίπεδο).

Οι **μέθοδοι ανάλυσης** που οι εξεταζόμενες έρευνες χρησιμοποίησαν είναι μεταξύ άλλων:

Αρνητική Διωνυμική Παλινδρόμηση, Μπεϋζιανή πολυμεταβλητή Poisson λογαριθμική παλινδρόμηση, Πλήρη Μπεϋζιανά πολυμεταβλητά μοντέλα τυχαίων παραμέτρων, Πολυωνυμικό – γενικευμένο Poisson με συνιστώσες σφάλματος, Μπεϋζιανό χωρικό μοντέλο, Λογιστική παλινδρόμηση με και χωρίς χωρική αυτοσυσχέτιση, Γεωγραφικά σταθμισμένη αρνητική διωνυμική παλινδρόμηση, Μονομεταβλητή και διμεταβλητή περιγραφική στατιστική, Γενικευμένες Εξισώσεις Εκτίμησης με Συνάρτηση Αρνητικού Διωνυμικού Συνδέσμου, Λανθάνουσα κατανομή Dirichlet, Κλασματικό Πολυωνυμικό Μοντέλο Διαχωρισμού.

Συμπερασματικά, φαίνεται ότι κατά την τελευταία δεκαετία έχουν υιοθετηθεί πολλές διαφορετικές μέθοδοι χωρικής ανάλυσης στη βιβλιογραφία, με μια τάση προς προηγμένα μοντέλα και Μπεϋζιανές μεθόδους, κάτι που οδήγησε σε ακριβείς προβλέψεις και ελαχιστοποίηση των σφαλμάτων. Ωστόσο, οι παραδοσιακές λειτουργικές και οικονομικές μέθοδοι εξακολουθούν να χρησιμοποιούνται, τουλάχιστον για λόγους σύγκρισης και αξιολόγησης. Επίσης, οι μέθοδοι μηχανικής μάθησης που πρόσφατα άρχισαν να εφαρμόζονται, σταδιακά αντικαθιστούν τα Μπεϋζιανά μοντέλα.

Ένα σημαντικό εύρημα της έρευνας είναι η **ευαισθησία** των hotspots ατυχημάτων. Τα αποτελέσματα έδειξαν ότι τα hotspots είναι διαφορετικά μεταξύ χρηστών διαφορετικών οχημάτων και ηλικιών, ενώ εμφανίζουν σημαντικές διακυμάνσεις κατά τη διάρκεια της μέρας. Επίσης, ένα κρίσιμο ζήτημα είναι η διαθεσιμότητα δεδομένων. Οι περισσότερες χωρικές αναλύσεις οδικής ασφάλειας βασίζονται σε προσπάθειες των ερευνητών να αποδείξουν μια ιδέα, χωρίς οι μεταβλητές να έχουν δοκιμαστεί εκτενώς λόγω έλλειψης δεδομένων.

Τέλος, η έρευνα καταλήγει στο συμπέρασμα ότι οι παράμετροι της ταχύτητας κυκλοφορίας και της γεωμετρίας του δικτύου δεν χρησιμοποιούνται τόσο συχνά όσο σε αναλύσεις των προηγούμενων δεκαετιών.

Το (2006) οι Agüero-Valverde & Jovanis δημοσίευσαν μια έρευνα **χωρικής ανάλυσης θανατηφόρων οδικών ατυχημάτων και τραυματισμών** για την πολιτεία της Pennsylvania των ΗΠΑ. Η μελέτη χρησιμοποιεί δεδομένα τραυματισμών και θανατηφόρων ατυχημάτων για την περίοδο 1996 – 2000 και συγκρίνει τα αποτελέσματα με παραδοσιακές εκτιμήσεις της ετήσιας συχνότητας ατυχημάτων σε επίπεδο κομητείας.

Οι παράγοντες που λαμβάνονται υπόψη είναι κοινωνικοοικονομικοί (ηλικιακές ομάδες, φύλο, ποσοστό φτώχειας), σχετιζόμενοι με τις μεταφορές (μίλια οδού, Ομοσπονδιακή οδική βοήθεια, πυκνότητα οδικού δικτύου), καθώς και περιβαλλοντικοί (βροχόπτωση, χιονόπτωση).

Οι **μέθοδοι ανάλυσης** που χρησιμοποιήθηκαν στη μελέτη περιλαμβάνουν κατανομή Poisson και αρνητικές διωνυμικές κατανομές που αποτελούν τις πλέον συνηθέστερες στον τομέα της μοντελοποίησης μετρήσεων για τα οδικά ατυχήματα. Επίσης, χρησιμοποιήθηκε χωρική μοντελοποίηση με πλήρη ιεραρχική Μπεϋζιανή προσέγγιση, μοντέλο αυτόματης παλινδρόμησης υπό όρους, καθώς και χωροχρονικά μοντέλα.

Στην αρχική ανάλυση δεδομένων χρησιμοποιείται μια σειρά αρνητικών διωνυμικών μοντέλων παλινδρόμησης, ενώ απορρίπτονται οι μεταβλητές με επίπεδο σημαντικότητας άνω του 0.10. Για αυτό το μοντέλο σημαντικά είναι τα μίλια υποδομής και το ποσοστό ταξιδιού σε Ομοσπονδιακές οδούς. Τα υψηλά πρότυπα σχεδιασμού σε αυτές τις οδούς μπορεί να ευθύνονται για τη μείωση του κινδύνου θανατηφόρου ατυχήματος. Επίσης σημαντικοί παράγοντες είναι το ποσοστό νέων οδηγών και η συνολική βροχόπτωση. Αντίστοιχα αποτελέσματα προέκυψαν και στο μοντέλο τραυματισμών.

Αναφορικά με τα πλήρη ιεραρχικά Μπεϋζιανά μοντέλα, οι σημαντικοί παράγοντες είναι παρόμοιοι με αυτούς του αρνητικού διωνυμικού μοντέλου.

Η έρευνα **συμπεραίνει** ότι δεν υπάρχουν ενδείξεις χωρικής συσχέτισης σε θανατηφόρα οδικά ατυχήματα. Ωστόσο, η χωρική συσχέτιση βρέθηκε να είναι σημαντική σε ατυχήματα που επέφεραν τραυματισμό. Τα μοντέλα θανατηφόρων ατυχημάτων έχουν λιγότερες σημαντικές μεταβλητές σε σύγκριση με τα μοντέλα τραυματισμών. Ο κύριος λόγος για αυτό είναι ο μικρός αριθμός συμβάντων για θανατηφόρα ατυχήματα - περίπου ένα θανατηφόρο ατύχημα για κάθε χίλια ατυχήματα με τραυματισμό.

Οι κομητείες με υψηλότερο ποσοστό του πληθυσμού κάτω από το επίπεδο φτώχειας, υψηλότερο ποσοστό του πληθυσμού τους στις ηλικιακές ομάδες 15 – 24 και άνω των 64 ετών και εκείνες με αυξημένη χιλιομετρική απόσταση και πυκνότητα οδικού δικτύου έχουν αυξήσει σημαντικά τον κίνδυνο ατυχήματος. Τέλος, η βροχόπτωση είναι σημαντική στα αρνητικά διωνυμικά μοντέλα αλλά όχι στα Μπεϋζιανά.

Μία πρόσφατη έρευνα δημοσιεύτηκε από τους Bassani et al. (2020) σχετικά με **χωρική ανάλυση οδικών ατυχημάτων** που περιλαμβάνουν ευάλωτους χρήστες της οδού.

Στα αστικά οδικά δίκτυα, περίπου 2 στους 3 θανάτους αφορούν πεζούς, ποδηλάτες και μοτοσικλετιστές, που αναφέρονται συλλογικά ως “**ευάλωτοι χρήστες της οδού**” λόγω της ανεπαρκούς φυσικής προστασίας τους σε περίπτωση ατυχήματος. Η μελέτη αυτή αναφέρει τα αποτελέσματα μιας ανάλυσης χωρικής κατανομής οδικών ατυχημάτων στις οποίες εμπλέκονται ευάλωτοι χρήστες στο Τορίνο της Ιταλίας κατά την περίοδο 2006 – 2016.

Αρχικά, εντοπίστηκαν επικίνδυνες θέσεις σε μητροπολιτική κλίμακα. Οι εσφαλμένες εκτιμήσεις της πραγματικής συχνότητας ατυχήματος, οι οποίες είναι χαρακτηριστικές των μελετών που διεξήχθησαν σε σύντομες περιόδους, αποφεύχθηκαν λαμβάνοντας υπόψη μόνο τις τοποθεσίες όπου τα ποσοστά ατυχήματος παρέμειναν σημαντικά υψηλά καθ' όλη τη διάρκεια της περιόδου παρατήρησης (έντεκα έτη). Τα δεδομένα ατυχημάτων αναλύθηκαν χρησιμοποιώντας τεχνολογίες γεωγραφικών συστημάτων πληροφοριών (GIS).

Η παρουσία **συστάδων** στα συμβάντα δείχνει μια χωρική εξάρτηση στα ατυχήματα, κάτι που σημαίνει ότι δύο ή περισσότερα ατυχήματα που έγιναν στο ίδιο σημείο μπορεί να οφείλονται στον ίδιο παράγοντα.

Το κύριο εργαλείο για την ανάλυση με μεθόδους πυκνότητας είναι η **Εκτίμηση Πυκνότητας Πυρήνα** (Kernel Density Estimation - KDE) που συνήθως χρησιμοποιείται στην εγκληματολογία, στην επιδημιολογία, στις οικονομικές αναλύσεις και στην οδική ασφάλεια. Ο μαθηματικός τύπος που χρησιμοποιεί η μέθοδος είναι:

$$\hat{f}(u, v) = \frac{1}{n \cdot h} \cdot \sum_{i=1}^n K\left(\frac{d_i}{h}\right)$$

όπου f είναι η εκτίμηση πυκνότητας ατυχημάτων στη θέση (u, v) , n ο αριθμός μετρήσεων, h το εύρος της μεθόδου KDE, K η επιλεγθείσα συνάρτηση, και d_i η απόσταση ανάμεσα στη θέση (u, v) και στη θέση της i μέτρησης.

Η επιλογή της συνάρτησης δεν επηρεάζει σημαντικά το αποτέλεσμα, ενώ ένα προτεινόμενο διάστημα τιμών για το εύρος είναι μεταξύ 100 και 300 m για αστικές περιοχές.

Η απεικόνιση των αποτελεσμάτων της ανάλυσης πυκνότητας σε χάρτες της αστικής περιοχής του Τορίνο έδειξε ότι οι μεγάλοι οδικοί άξονες με αυξημένους κυκλοφοριακούς φόρτους αναμενόμενα αυξάνουν και τον αριθμό των ατυχημάτων ανάμεσα σε οχήματα και ευάλωτους χρήστες. Επίσης, οι υψηλές ταχύτητες που αναπτύσσουν τα οχήματα σε αυτές τις οδούς έρχονται σε αντίθεση με αυτές των ευάλωτων χρηστών κάτι που πιθανότατα οδηγεί σε σοβαρότερα οδικά ατυχήματα.

Το **συμπέρασμα** είναι ότι τα ατυχήματα που περιλαμβάνουν τουλάχιστον έναν ευάλωτο χρήστη αποτελούν περίπου το 40% των συνολικών ατυχημάτων, ενώ τα περισσότερα συμβαίνουν σε διασταυρώσεις κατά μήκος μεγάλων οδικών αξόνων μέσα σε αστική περιοχή. Οι θέσεις αυτές έχουν τρία κρίσιμα χαρακτηριστικά: (1) αυξημένος φόρτος στην κύρια οδό, (2) διαφορά ταχυτήτων ανάμεσα στους χρήστες και (3) γεωμετρία της διασταύρωσης που εκθέτει τους ευάλωτους χρήστες σε κίνδυνο.

Το (2005) δημοσιεύτηκε μια μελέτη **χωρικής ανάλυσης της συμφόρησης και οδικής ασφάλειας** στο Λονδίνο από τους Noland & Quddus. Σκοπός της ήταν να εξεταστεί πως η κυκλοφοριακή συμφόρηση μπορεί να επηρεάσει την ασφάλεια της κυκλοφορίας.

Η μείωση της συμφόρησης πάντα αποτελούσε πρωταρχικό στόχο της πολιτικής των μεταφορών καθώς περιορίζει τις καθυστερήσεις και το οικονομικό κόστος. Ωστόσο, μπορεί να υπάρχει αντίστροφη σχέση μεταξύ της μείωσης της συμφόρησης και της βελτίωσης της ασφάλειας. Αυτό οφείλεται κυρίως στις υψηλές ταχύτητες που μπορούν να αναπτύξουν τα οχήματα όσο οι συνθήκες πλησιάζουν ελεύθερη ροή.

Η έρευνα χρησιμοποίησε την εθνική βάση δεδομένων του Ηνωμένου Βασιλείου STATS19 για τη συλλογή στοιχείων ατυχημάτων ανά επίπεδο σοβαρότητας, ώρα της ημέρας και τοποθεσία στη μητροπολιτική περιοχή του Λονδίνου. Παρατηρήθηκε όμοια κατάσταση τόσο στο εσωτερικό όσο και στο εξωτερικό Λονδίνο, ενώ οι διαφορές ανάμεσα σε συμφορημένες και μη συμφορημένες χρονικές περιόδους είναι γενικά χαμηλές καθώς οι υποδομές για τη βελτίωση της ασφάλειας σε κυκλοφορία υψηλής ταχύτητας είναι ήδη υπαρκτές.

Η σοβαρότητα των οδικών ατυχημάτων κατατάσσεται ανάλογα με ελαφριούς ή σοβαρούς τραυματισμούς, θανάτους, καθώς και τα αντίστοιχα κριτήρια για τους πεζούς. Οι **ανεξάρτητες μεταβλητές** που χρησιμοποιήθηκαν στην ανάλυση είναι η απασχόληση, το μήκος διαδρομής, η γεωμετρία των κόμβων, η ύπαρξη υπόγειων σταθμών και στάσεων λεωφορείων και η ιδιοκτησία ΙΧ.

Η **στατιστική επεξεργασία** ξεκινά με ένα μοντέλο Poisson, που αποτελεί τη συνηθέστερη προσέγγιση για διακριτά, τυχαία γεγονότα όπως οδικά ατυχήματα. Ωστόσο, παρατηρήθηκε το πρόβλημα ότι ο ρυθμός με τον οποίο συμβαίνουν τα γεγονότα σε μια χρονική περίοδο μπορεί να μην είναι σταθερός, δηλαδή η διακύμανση να μην ισούται με τον μέσο όρο. Συνεπώς, οι μελετητές προχώρησαν με ένα αρνητικό διωνυμικό μοντέλο.

Τα **αποτελέσματα** της έρευνας έδειξαν αρχικά ότι οι θάνατοι πεζών είναι περίπου το 50% όλων των θανάτων από οδικά ατυχήματα στο Λονδίνο. Διαπιστώθηκε ότι οι σταθμοί δημόσιων συγκοινωνιών τείνουν να συνδέονται με περισσότερα θύματα, αλλά λιγότερο κατά τη διάρκεια χρονικών περιόδων χωρίς συμφόρηση. Τα αποτελέσματα για την πυκνότητα των αυτοκινητοδρόμων επιβεβαιώνουν την αρχική υπόθεση ότι η συμφόρηση βελτιώνει την οδική ασφάλεια. Αντίθετα, οι κυκλικοί κόμβοι φαίνεται να σχετίζονται με ατυχήματα που περιλαμβάνουν πεζούς ακόμα και σε συνθήκες συμφόρησης, κυρίως κατά τη διάρκεια της νύχτας.

Μία ακόμη έρευνα πάνω στις **χωρικές αναλύσεις** στην οδική ασφάλεια – συγκεκριμένα πάνω στα **ατυχήματα βαρέων οχημάτων** – έγινε το (2017) από τους Gudes et al. στην Αυστραλία.

Σκοπός της μελέτης είναι ο προσδιορισμός των hotspots ατυχημάτων βαρέων οχημάτων στη Δυτική Αυστραλία μεταξύ των ετών 2001 – 2013, με τη χρήση μιας προσέγγισης χωρικής ανάλυσης. Προκειμένου να επιτευχθεί ο στόχος της μελέτης εντοπίστηκαν τρία ερευνητικά ερωτήματα:

- Ποιες κρατικές οδοί έχουν τα υψηλότερα ποσοστά ατυχημάτων;
- Πού είναι οι υψηλές συγκεντρώσεις (hotspots) ατυχημάτων βαρέων οχημάτων σε κρατικές και μη κρατικές οδούς;
- Ποιες είναι οι τάσεις στο χρόνο και στο χώρο των ατυχημάτων βαρέων οχημάτων (χωροχρονικά μοτίβα);

Η αρχική μέθοδος που χρησιμοποιήθηκε για την ανάλυση των ατυχημάτων είναι μια προσέγγιση **GIS**. Δεδομένα ατυχημάτων και σχετικά χωρικά δεδομένα από διάφορες πηγές εισήχθησαν σε μια βάση δεδομένων ArcGIS για διαχείριση και ανάλυση. Τα πρωτογενή δεδομένα ελήφθησαν από τις τοπικές οδικές αρχές. Πρόσθετες πληροφορίες που παρασχέθηκαν από τη βάση δεδομένων IRIS περιλάμβαναν την τοποθεσία, τον μήνα, την ημερομηνία, την ημέρα, την ώρα, τον τύπο του οχήματος, τη σοβαρότητα και τον τύπο του ατυχήματος.

Τα οδικά ατυχήματα κατά την περίοδο μελέτης (13 έτη) υπολογίστηκαν για κάθε τμήμα του συνολικού μήκους της οδού με βάση την παρακάτω εξίσωση:

Οδικά ατυχήματα / MVKT = οδικά ατυχήματα ανά έτος/(μήκος τμήματος (km) × AADT × 365 ημέρες × 13 έτη/εκατομμύριο).

όπου MVKT τα εκατομμύρια οχηματο – χιλιόμετρα, και AADT η ετήσια μέση ημερήσια κυκλοφορία (ΕΜΗΚ).

Αυτή η προσέγγιση, ωστόσο, κρίθηκε ακατάλληλη, καθώς η μονάδα MVKT κάνει την υπόθεση ότι το ποσοστό ατυχήματος έχει γραμμική σχέση με τα ατυχήματα, δηλαδή καθώς αυξάνεται η κυκλοφορία σε ένα οδικό τμήμα, τα ατυχήματα θα αυξηθούν επίσης. Έτσι, χρησιμοποιήθηκε τελικά η ακόλουθη μέθοδος:

Οδικά ατυχήματα / MVKT = οδικά ατυχήματα ανά έτος / (άθροισμα όλων των μηκών τμημάτων οδού (km) για μία συγκεκριμένη οδό) × (AADT × 365 ημέρες × 13 χρόνια / εκατομμύριο).

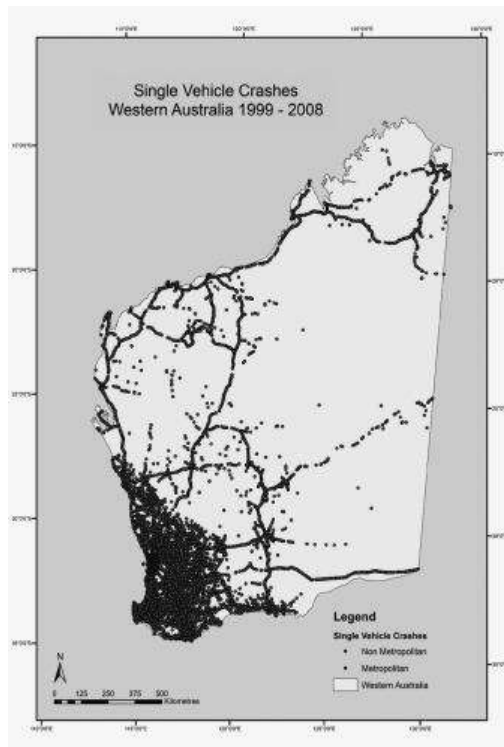
Συμπερασματικά, η πλειονότητα των οδικών ατυχημάτων βαρέων οχημάτων συνέβησαν στην μητροπολιτική περιοχή του Περθ. Στο Διάγραμμα 2.1. φαίνεται ότι οι περιοχές με τα υψηλότερα ποσοστά ατυχημάτων (πορτοκαλί και κόκκινο), παρατηρούνται κοντά στο λιμάνι του Fremantle, το Διεθνές Αεροδρόμιο του Περθ και ορισμένες βιομηχανικές ζώνες.



Διάγραμμα 2.1.: Ποσοστά ατυχημάτων βαρέων οχημάτων στο Περθ (Gudes et al., 2017)

Μια έρευνα πάνω στις **χωροχρονικές μεθόδους ανάλυσης στην οδική ασφάλεια** πραγματοποιήθηκε το (2011) από τους Plug et al. σχετικά με οδικά ατυχήματα μεμονωμένων οχημάτων. Σκοπός της ήταν η διερεύνηση της αλληλεπίδρασης μεταξύ της θέσης των ατυχημάτων και του χρόνου εμφάνισής τους.

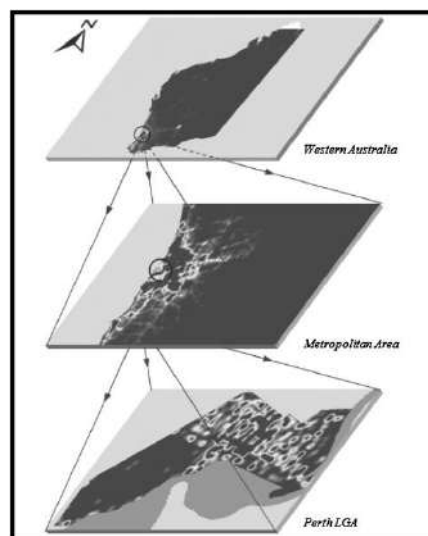
Η μελέτη αφορούσε μοτίβα ατυχημάτων στη Δυτική Αυστραλία για την περίοδο 1999 – 2008, σε καθημερινό και εβδομαδιαίο επίπεδο, και σε τρεις διαφορετικές κλίμακες: Δυτική Αυστραλία, Μητροπολιτική περιοχή και Περιοχή τοπικής αυτοδιοίκησης του Περθ. Η περιοχή μελέτης και η κατανομή των ατυχημάτων μεμονωμένου οχήματος παρουσιάζονται στο Διάγραμμα 2.2..



**Διάγραμμα 2.2.: Κατανομή ατυχημάτων στην περιοχή μελέτης
(Plug et al., 2011)**

Οι μέθοδοι που χρησιμοποιήθηκαν είναι τα χρονικά διαγράμματα αράχνης (spider graphs), η Εκτίμηση Πυκνότητας Πυρήνα (KDE) και η τεχνική ανάλυσης Comap που επιτρέπει τη διερεύνηση ενός μοτίβου σε πολυμεταβλητά δεδομένα πριν την επίσημη στατιστική μοντελοποίηση.

Ως προς τη χωρική ανάλυση, χρησιμοποιήθηκε η Εκτίμηση Πυκνότητας Πυρήνα στις τρεις χωρικές κλίμακες που αναφέρθηκαν παραπάνω. Στο Διάγραμμα 2.3. φαίνεται η ομαδοποίησή τους.



**Διάγραμμα 2.3.: Χωρική ομαδοποίηση αποτελεσμάτων
(Plug et al., 2011)**

Συμπερασματικά, η κλίμακα παίζει κεντρικό ρόλο στον επηρεασμό της αντίληψης του χώρου και του χρόνου. Οι τεχνικές που εφαρμόστηκαν σε αυτή τη μελέτη οδηγούν σε μοτίβα ατυχημάτων οχημάτων σε συγκεκριμένη περιοχή και χρονική περίοδο. Η διασταύρωση των μοτίβων με γνωστές παραμέτρους κινδύνου όπως η ταχύτητα και η κόπωση του οδηγού, μπορούν να βοηθήσουν σε καλύτερη κατανόηση των λόγων πίσω από τα οδικά ατυχήματα.

Οι Zeng & Huang το (2014) δημοσίευσαν μια μελέτη που προτείνει ένα Μπεύζιανό **χωρικό κοινό μοντέλο πρόβλεψης οδικών ατυχημάτων** που περιλαμβάνει τόσο οδικά τμήματα όσο και διασταυρώσεις που βρίσκονται σε ένα αστικό οδικό δίκτυο, μέσω του οποίου θα μπορούσαν να εξεταστούν οι χωρικές συσχετίσεις μεταξύ ετερογενών τύπων οντοτήτων.

Ως περιοχή μελέτης επιλέχθηκε ένα οδικό δίκτυο στο Hillsborough, Florida των ΗΠΑ με δεδομένα ατυχημάτων, οδών και κυκλοφοριακών χαρακτηριστικών για μια τριετή περίοδο (2005 – 2007). Απώτερος σκοπός των ερευνητών ήταν να συγκριθεί το προτεινόμενο κοινό μοντέλο (Joint) με τρία μοντέλα πρόβλεψης ατυχήματος σε επίπεδο τοποθεσίας – ένα Poisson, ένα αρνητικό διωνυμικό και ένα υπό όρους αυτοσυσχετιζόμενο (CAR).

Οι ανεξάρτητες μεταβλητές που χρησιμοποιήθηκαν για την ανάλυση είναι το μήκος οδικού τμήματος, η Ετήσια Μέση Ημερήσια Κυκλοφορία (ΕΜΗΚ), το όριο ταχύτητας, ο αριθμός λωρίδων και οδών πρόσβασης, η επιφανειακή κατάσταση του οδοστρώματος και η παρουσία ή όχι διαχωριστικής νησίδας.

Η εφαρμογή των μοντέλων έδειξε ότι στα αρνητικά διωνυμικά μοντέλα υπήρχε υπερδιασπορά στα σύνολα δεδομένων που αναλύονται. Επίσης, τα μοντέλα CAR, Joint καθώς και τα αρνητικά διωνυμικά ξεπερνούν το μοντέλο Poisson σε σημαντικό βαθμό ως προς την προσαρμογή στα δεδομένα.

Η έρευνα καταλήγει στο **συμπέρασμα** ότι το υπό όρους αυτοσυσχετιζόμενο μοντέλο (CAR) και το κοινό μοντέλο Joint υπερτερούν των αρνητικών διωνυμικών και Poisson, κάτι που δείχνει ότι η εξέταση των χωρικών συσχετίσεων μεταξύ γειτονικών οντοτήτων είναι λογική. Σε σύγκριση με τα παραδοσιακά μοντέλα Poisson, το προτεινόμενο αυτής της μελέτης λαμβάνει υπόψη διάφορα χαρακτηριστικά του οδικού περιβάλλοντος, καθώς και τη χωρική τους συσχέτιση στο πλαίσιο ενός ολοκληρωμένου οδικού δικτύου.

Σε μία μελέτη του το (2021) ο ερευνητής του ΕΜΠ κ. Ζιακόπουλος ασχολήθηκε με τη **χωρική ανάλυση** περιστατικών επικίνδυνης οδηγικής συμπεριφοράς στα αστικά οδικά δίκτυα με χρήση νέων τεχνολογιών μέσω smartphone.

Το πρώτο βήμα της μεθόδου είναι η ανίχνευση του βαθμού στον οποίο υπάρχει χωρική εξάρτηση σε ένα συγκεκριμένο φαινόμενο, μέσω της εκδήλωσής του σε ένα συγκεκριμένο σύνολο δεδομένων. Το πιο ευρέως χρησιμοποιούμενο μέτρο χωρικής αυτοσυσχέτισης είναι ο **συντελεστής Moran's I** που είναι:

$$I = \frac{n}{W} * \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Όπου n είναι ο πληθυσμός, x το χαρακτηριστικό του ενδιαφέροντος του πληθυσμού, x – παύλα η μέση τιμή του χαρακτηριστικού ενδιαφέροντος, i και j δείκτες θέσης, w_{ij} ένας

πίνακας χωρικών βαρών που δίνεται από ένα επιλεγμένο γεωγραφικό κριτήριο με διαγώνια στοιχεία μηδέν. W είναι το άθροισμα όλων των w_{ij} σε ολόκληρη την περιοχή μελέτης έτσι ώστε:

$$W = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$$

Ένα άλλο εργαλείο που χρησιμοποιήθηκε για την ανίχνευση της χωρικής εξάρτησης των παρατηρήσεων είναι το **βαριόγραμμα** (Variogram). Τα διαγράμματα αυτά περιγράφονται κυρίως από τρεις παραμέτρους: Το ψήγμα h , περιγράφοντας τον αρχικό θόρυβο ή τυχαίες αποκλίσεις στις μετρήσεις, το s που είναι η μέγιστη ημιδιακύμανση, περιορίζοντας ασυμπτωτικά τη μεταβλητότητα σε τιμές εκτός απείρου στις μεγαλύτερες αποστάσεις, και το εύρος r , περιγράφοντας την απόσταση στην οποία η ημιδιακύμανση είναι περίπου 95 % της διακύμανσης του s .

Τέλος, χρησιμοποιήθηκε η **Γεωγραφικά Σταθμισμένη Παλινδρόμηση Poisson** (Geographically Weighted Regression – GWR). Η οικογένεια μοντέλων GWR αποτελεί προέκταση των παραδοσιακών μοντέλων γραμμικής παλινδρόμησης σε μια περιοχή μελέτης. Συγκεκριμένα, πρόκειται για τεχνική που χρησιμοποιείται κυρίως για την ένδειξη σημείων ή περιοχών σε έναν χάρτη όταν οι τοπικοί συντελεστές παλινδρόμησης ποικίλλουν από τον παγκόσμιο μέσο όρο τους.

Προκειμένου να επιτευχθεί ο στόχος της μελέτης, διερευνήθηκε η αξιολόγηση της ασφάλειας των αστικών οδικών δικτύων σε βάση οδικού τμήματος με τη χρήση παραμέτρων οδήγησης υψηλής ανάλυσης που συλλέγονται μέσω αισθητήρων smartphone. Τα δεδομένα οδήγησης σχεδιάζονται και αντιστοιχίζονται σε τμήματα του OpenStreetMap, δημιουργώντας σύνολα χωρικών δεδομένων που σχετίζονται με την οδική ασφάλεια με συγκεντρωτικές μετρήσεις συμπεριφοράς που στη συνέχεια αναλύονται σε τμηματική βάση.

Τα **αποτελέσματα** της μελέτης έδειξαν ότι για τους συντελεστές Moran's I υπάρχει μια θετική χωρική αυτοσυσχέτιση τόσο για απότομες επιταχύνσεις όσο και για απότομες πεδήσεις. Η έρευνα έδειξε ότι περίπου 190 μέτρα από κάθε κεντροειδές οδικού τμήματος δεν παρατηρείται χωρική αυτοσυσχέτιση για τις πεδήσεις. Η αντίστοιχη απόσταση για τις επιταχύνσεις είναι 200 μέτρα.

Τα μοντέλα GWR έδειξαν ότι οι απότομες επιταχύνσεις σχετίζονται στενότερα με οδηγούς που επιλέγουν μια πιο επιθετική συμπεριφορά στο οδικό περιβάλλον, ενώ οι απότομες πεδήσεις μπορούν να θεωρηθούν ως αντίδραση για την αποφυγή ατυχήματος.

Το (2011) οι Spoerri et al. πραγματοποίησαν μια μελέτη **χωρικών αναλύσεων** για θανατηφόρα οδικά ατυχήματα στην Ελβετία. Οι μελετητές συνδύασαν κοινωνικούς, δημογραφικούς και γεωγραφικούς παράγοντες με στοιχεία θνησιμότητας για έναν πληθυσμό της τάξης των 5.5 εκατομμυρίων και ένα ηλικιακό εύρος 18 – 94. Τα στοιχεία περιλάμβαναν θανάτους επιβατών ΙΧ αυτοκινήτων, μοτοσικλετιστών, ποδηλατών και πεζών. Σκοπός ήταν ο υπολογισμός των αναλογιών κινδύνου και θνησιμότητας σε όλες τις περιοχές μελέτης.

Οι μεταβλητές ατομικού επιπέδου που χρησιμοποιήθηκαν περιγράφουν την εκπαίδευση, το επάγγελμα, την οικογενειακή κατάσταση, την εθνικότητα και τον τύπο του νοικοκυριού. Παράλληλα, δημιουργήθηκαν 1744 περιοχές με τουλάχιστον 1000 κατοίκους η κάθε μία. Τα δεδομένα σε επίπεδο περιοχής περιλάμβαναν τη γλωσσική περιφέρεια (γαλλικά,

ιταλικά, γερμανικά) και την πυκνότητα του πληθυσμού, η οποία υπολογίστηκε σε κάθε περιοχή ως ο αριθμός των κατοίκων ανά παραγωγική επιφάνεια γης.

Για την ανάλυση των δεδομένων χρησιμοποιήθηκαν 3 **μοντέλα επιβίωσης Weibull** (ελάχιστα προσαρμοσμένο, προσαρμοσμένο ατομικό επίπεδο και πλήρως προσαρμοσμένο) υπολογίζοντας το χρόνο από την ημερομηνία της απογραφής (5 Δεκεμβρίου 2000) έως το νωρίτερο του θανάτου, της μετανάστευσης ή της 31ης Δεκεμβρίου 2005. Τα μοντέλα συγκρίθηκαν χρησιμοποιώντας δοκιμές αναλογίας πιθανότητας και το **κριτήριο πληροφοριών Akaike (AIC)**, το οποίο αντιπροσωπεύει τόσο την προσαρμογή του μοντέλου όσο και την πολυπλοκότητα. Τα αποτελέσματα εκφράστηκαν ως λόγοι κινδύνου με διαστήματα εμπιστοσύνης 95%.

Συμπερασματικά, σε κάθε περίπτωση η θνησιμότητα ήταν υψηλότερη στους άνδρες από ότι στις γυναίκες, με το ποσοστό των ανδρών να κυμαίνεται από 57.1% στους πεζούς έως 95.7% στα ατυχήματα με μοτοσικλέτες. Επίσης, παρατηρήθηκε ότι οι θάνατοι σε επιβάτες μηχανοκίνητων οχημάτων και μοτοσικλετιστές συσχετίστηκαν με νεαρή ηλικία, ενώ οι θάνατοι σε πεζούς ήταν συχνότεροι σε ηλικιωμένους. Υπήρξε τέλος συσχέτιση με χαμηλότερο μορφωτικό επίπεδο, ιδίως για τους θανάτους πεζών.

Μία ακόμα έρευνα πάνω στα **χωρικά μοντέλα συχνότητας ατυχήματος** πραγματοποιήθηκε το (2013) από τον Agüero-Valverde για την Κόστα Ρίκα. Σκοπός του ήταν να επεκτείνει την έννοια των λειτουργιών ασφάλειας που πρέπει να χρησιμοποιούνται σε επιφανειακά μοντέλα συχνότητας ατυχημάτων. Το πολυμεταβλητό χωρικό μοντέλο της μεθόδου συγκρίνεται με το αντίστοιχο μονομεταβλητό, ενώ χρησιμοποιούνται επίσης μια πλήρης Μπεϋζιανή ιεραρχική προσέγγιση και ένα υπό όρους αυτορυθμιζόμενο μοντέλο.

Για κάθε καντόνι της Κόστα Ρίκα ο μελετητής εφάρμοσε την Μπεϋζιανή προσέγγιση, με τα σφάλματα του μοντέλου να κατανέμονται κατά Poisson. Για τη διερεύνηση της χωρικής συσχέτισης χρησιμοποιήθηκε ένα μονομεταβλητό υπό όρους αυτοπαλινδρομικό μοντέλο (CAR), καθώς και ένα πολυμεταβλητό υπό όρους αυτορυθμιζόμενο (MCAR).

Το σύνολο δεδομένων ατυχημάτων που χρησιμοποιείται στην ανάλυση συγκεντρώνεται χωρικά σε επίπεδο καντονιού. Περιλαμβάνονται τρία επίπεδα σοβαρότητας: θανατηφόρα ατυχήματα, τραυματισμοί και υλικές ζημιές μόνο. Τα δεδομένα ατυχημάτων παρασχέθηκαν από τις υπηρεσίες έκτακτης ανάγκης της Κόστα Ρίκα 911 από τα έτη 2002 έως 2006. Οι ανεξάρτητες μεταβλητές που χρησιμοποιούνται στα μοντέλα είναι το μήκος των οδών, το ποσοστό φτώχειας, οι ηλικιακές ομάδες, το ποσοστό των ανδρών και ο αστικός πληθυσμός.

Από την ανάλυση προκύπτει ότι οι μεταβλητές οχήματος – χιλιομέτρων είναι σημαντικές για τα τρία επίπεδα σοβαρότητας τόσο σε μονομεταβλητά όσο και σε πολυμεταβλητά μοντέλα. Η μεταβλητή του μήκους της οδού δεν ήταν σημαντική και στα δύο μοντέλα και στις τρεις περιπτώσεις σοβαρότητας, ενώ το ποσοστό των ατόμων κάτω από το όριο της φτώχειας ήταν σημαντικό και συσχετίστηκε αρνητικά με τη συχνότητα ατυχήματος για τραυματισμούς και σοβαρότητα υλικών ζημιών.

Όσον αφορά τις ηλικιακές ομάδες, η έρευνα **καταλήγει** ότι το ποσοστό των ατόμων μεταξύ 0 και 14 ετών είναι σημαντικό και θετικό για θανατηφόρα ατυχήματα και ατυχήματα τραυματισμών στο πολυμεταβλητό μοντέλο. Το ποσοστό του νεαρού πληθυσμού (ηλικίας 15-24 ετών) βρέθηκε επίσης θετικό και συσχετίστηκε σημαντικά με τη συχνότητα ατυχήματος με τραυματισμό και τη σοβαρότητα των υλικών ζημιών. Το ποσοστό των

ατόμων ηλικίας άνω των 65 ετών είναι επίσης θετικό και σημαντικό για τραυματισμούς και ατυχήματα υλικών ζημιών στο πολυμεταβλητό μοντέλο. Τέλος, το ποσοστό του αστικού πληθυσμού είναι επίσης θετικό και συσχετίζεται σημαντικά με τον κίνδυνο ατυχήματος για όλες τις σοβαρότητες στο πολυμεταβλητό μοντέλο.

2.3 Σύνοψη

Από τη διεθνή βιβλιογραφία προκύπτουν οι ακόλουθες παρατηρήσεις:

- Οι σύγχρονες μελέτες οδικής ασφάλειας έχουν μια τάση προς προηγμένα μοντέλα, Μπεϋζιανές μεθόδους και μεθόδους μηχανικής μάθησης (machine learning).
- Τα υπάρχοντα δεδομένα πάνω σε χωρικές αναλύσεις εξακολουθούν να είναι ελλιπή, ενώ οι ερευνητές βασίζονται κυρίως σε υποθέσεις τις οποίες και προσπαθούν να αποδείξουν.
- Οι σύγχρονες μέθοδοι ανάλυσης τείνουν να μην χρησιμοποιούν τις παραμέτρους της ταχύτητας κυκλοφορίας και της γεωμετρίας του δικτύου όσο οι μελέτες προηγούμενων δεκαετιών.
- Ο κίνδυνος ατυχήματος κατά κανόνα αυξάνεται όσο αυξάνεται το μήκος διαδρομής και η πυκνότητα των οδών, ενώ είναι επίσης αυξημένος στους νέους οδηγούς και τους ηλικιωμένους.
- Ένα μεγάλο ποσοστό των οδικών ατυχημάτων περιλαμβάνει τουλάχιστον έναν ευάλωτο χρήστη, κάτι που συνήθως προκαλείται από τη γεωμετρία των διασταυρώσεων που εκθέτει τους ευάλωτους χρήστες σε κίνδυνο.
- Πόλοι ατυχημάτων που περιλαμβάνουν πεζούς είναι οι σταθμοί δημόσιων συγκοινωνιών και οι κυκλικοί κόμβοι.
- Η κλίμακα της μελέτης έχει ουσιαστική σημασία στις χωρικές αναλύσεις οδικής ασφάλειας.
- Η συνεκτίμηση διάφορων χαρακτηριστικών του οδικού περιβάλλοντος βοηθά στην καλύτερη κατανόηση των αιτιών των οδικών ατυχημάτων.
- Απότομες επιταχύνσεις των οδηγών σχετίζονται με επιθετική συμπεριφορά, ενώ απότομες πεδησεις σχετίζονται με ενέργειες αποφυγής ατυχημάτων.
- Η πλειοψηφία των θυμάτων σε οδικά ατυχήματα είναι άνδρες.
- Θανατηφόρα ατυχήματα με οχήματα σχετίζονται κυρίως με νεαρές ηλικίες, ενώ ατυχήματα πεζών σχετίζονται κυρίως με τους ηλικιωμένους.
- Το μορφωτικό επίπεδο των χρηστών του οδικού περιβάλλοντος επηρεάζει σημαντικά τα ατυχήματα όπου εμπλέκονται πεζοί.

Κεφάλαιο 3: Θεωρητικό Υπόβαθρο

3.1 Εισαγωγή

Στο παρόν κεφάλαιο παρατίθεται η **περιγραφή του θεωρητικού υποβάθρου** που αξιοποιήθηκε για την ανάλυση του αντικειμένου της Διπλωματικής Εργασίας. Αρχικά, γίνεται μια σύντομη περιγραφή ορισμένων βασικών εννοιών στατιστικής που χρησιμοποιούνται σε πολλές εφαρμογές στον τομέα της οδικής ασφάλειας. Στη συνέχεια, ακολουθεί μια συνοπτική επεξήγηση των μαθηματικών προτύπων που χρησιμοποιήθηκαν για την συσχέτιση των οδικών ατυχημάτων με μια σειρά ανεξάρτητων μεταβλητών, καθώς επίσης και της μεθόδου χωρικής ανάλυσης GWR. Επίσης, αναλύεται η διαδικασία ομαδοποίησης δεδομένων που χρησιμοποιήθηκε για την κατάταξη των Δήμων ανάλογα με τα χαρακτηριστικά τους. Τέλος, παρουσιάζονται οι απαραίτητοι στατιστικοί έλεγχοι και τα κριτήρια αποδοχής ή απόρριψης ενός μοντέλου.

3.2 Βασικές Έννοιες Στατιστικής

Ο όρος **πληθυσμός** (population) αναφέρεται στο σύνολο των παρατηρήσεων ενός χαρακτηριστικού που εξετάζεται στη στατιστική έρευνα. Πρόκειται για ένα σύνολο στοιχείων που είναι τελείως καθορισμένα. Ένας πληθυσμός μπορεί να είναι πραγματικός ή θεωρητικός.

Ο όρος **δείγμα** (sample) αναφέρεται σε ένα υποσύνολο του πληθυσμού. Η πλειοψηφία των στατιστικών ερευνών βασίζεται σε δείγματα, καθώς οι ιδιότητες ενός πληθυσμού είναι συνήθως αδύνατο να καταγραφούν. Όλα τα στοιχεία ενός δείγματος ανήκουν υποχρεωτικά και στον πληθυσμό, χωρίς ωστόσο να ισχύει το αντίστροφο. Με την προϋπόθεση ότι το δείγμα είναι αντιπροσωπευτικό του πληθυσμού, τα συμπεράσματα που προκύπτουν κατά τη μελέτη του δείγματος θα ισχύουν με ικανοποιητική ακρίβεια για ολόκληρο τον πληθυσμό. Συνήθως, για το πλήθος των στοιχείων ενός δείγματος χρησιμοποιείται ο συμβολισμός “*n*”.

Ως **μεταβλητές** (variables) ορίζονται τα χαρακτηριστικά που πρέπει να μετρηθούν και να καταγραφούν σε ένα δείγμα. Οι μεταβλητές μπορούν να διακριθούν σε δύο κατηγορίες:

- **Ποιοτικές μεταβλητές** (qualitative variables)
Στην περίπτωση αυτή, οι δυνατές τιμές των μεταβλητών είναι κατηγορίες διαφορετικές μεταξύ τους. Η χρήση αριθμών για την αναπαράσταση των τιμών μιας τέτοιας μεταβλητής είναι αποκλειστικά συμβολική και δεν έχει την έννοια της μέτρησης.
- **Ποσοτικές μεταβλητές** (quantitative variables):
Πρόκειται για τις μεταβλητές με τιμές αριθμούς, που έχουν την έννοια της μέτρησης. Οι ποσοτικές μεταβλητές διακρίνονται με τη σειρά τους σε δύο επιμέρους κατηγορίες, τις διακριτές και τις συνεχείς. Στην περίπτωση μιας **διακριτής μεταβλητής**, η μικρότερη μη μηδενική διαφορά δύο τιμών είναι σταθερή ποσότητα.

Αντίθετα, σε μια **συνεχή μεταβλητή**, δύο τιμές μπορούν να διαφέρουν κατά οποιαδήποτε μικρή ποσότητα. Στην πράξη, συνεχή μεταβλητή θεωρείται αυτή που μπορεί να πάρει όλες τις τιμές σε ένα διάστημα. Σε αντίθετη περίπτωση, θεωρείται διακριτή.

Μέτρα κεντρικής τάσης (measures of central tendency):

Στην περίπτωση ανάλυσης ενός δείγματος x_1, x_2, \dots, x_n , η μέση τιμή υπολογίζεται από την παρακάτω 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):

Αναφέρονται στην περίπτωση όπου τα δεδομένα αποτελούν ένα δείγμα. Η διακύμανση συμβολίζεται με s^2 και υπολογίζεται ως εξής:

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

όπου ο τελευταίος όρος είναι ο δειγματικός μέσος, δηλαδή η μέση τιμή των παρατηρήσεων στο δείγμα.

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

$$s = (s^2)^{1/2} = \left[\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \right]^{1/2}$$

Στην περίπτωση συμμετρικά κατανομημένου δείγματος, σύμφωνα με εμπειρικό κανόνα προκύπτει ότι το διάστημα:

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

Η συνδιακύμανση (covariance) δύο μεταβλητών αποτελεί ένα μέτρο της σχέσης μεταξύ δύο περιοχών δεδομένων, και υπολογίζεται ως εξής:

$$Cov(X, Y) = \left[\frac{1}{n-1} \right] \sum_{i=1}^n [(X_i - \bar{X})(Y_i - \bar{Y})]$$

Μέτρα αξιοπιστίας

- **Επίπεδο εμπιστοσύνης:** η αναλογία των περιπτώσεων που μια εκτίμηση είναι σωστή.
- **Επίπεδο σημαντικότητας:** η αναλογία των περιπτώσεων που ένα συμπέρασμα είναι λανθασμένο.

3.3 Συσχέτιση μεταβλητών & συντελεστής συσχέτισης

Έστω δύο τυχαίες μεταβλητές X, Y . Ο **συντελεστής συσχέτισης** (correlation coefficient) “ ρ ” καθορίζει το βαθμό γραμμικής συσχέτισης των δύο μεταβλητών X και Y , με διασπορά σ_X^2 και σ_Y^2 αντίστοιχα, και συνδιακύμανση $\sigma_{XY} = \text{Cov}(X, Y)$.

$$\rho = \left(\frac{\sigma_{XY}}{\sigma_X} \right) \left(\frac{1}{\sigma_Y} \right)$$

Ο συντελεστής συσχέτισης εκφράζει το βαθμό και τον τρόπο με τον οποίο οι μεταβλητές συσχετίζονται. Λαμβάνει τιμές στο διάστημα $[-1, 1]$ και είναι ανεξάρτητος από τη μονάδα μέτρησης των μεταβλητών. Τιμές κοντά στο 1 δηλώνουν ισχυρή θετική συσχέτιση, τιμές κοντά στο -1 δηλώνουν ισχυρή αρνητική συσχέτιση, ενώ τιμές κοντά στο 0 δηλώνουν γραμμική ανεξαρτησία των X και Y .

3.4 Βασικές Στατιστικές Κατανομές

Σύμφωνα με τη θεωρία της στατιστικής, προκειμένου να μελετηθούν διάφορα μεγέθη θα πρέπει να είναι γνωστή η **μορφή της κατανομής** που ακολουθούν. Παρακάτω παρουσιάζονται οι κυριότερες στατιστικές κατανομές που χρησιμοποιούνται στις αναλύσεις οδικής ασφάλειας.

3.4.1 Κανονική Κατανομή (Gauss)

Μια από τις πλέον σημαντικότερες κατανομές πιθανότητας που αφορά συνεχείς μεταβλητές. Μια συνεχής τυχαία μεταβλητή X ακολουθεί την κανονική κατανομή με παραμέτρους μ και σ ($-\infty < \mu < +\infty$, $\sigma > 0$), όταν έχει συνάρτηση πυκνότητας πιθανότητας την παρακάτω:

$$f(x) = \left(\frac{1}{(2\pi\sigma^2)^{1/2}} \right) e^{[-(x-\mu)^2/2\sigma^2]}$$

όπου μ και σ , σταθερές ίσες με την μέση τιμή και τυπική απόκλιση αντίστοιχα.

3.4.2 Κατανομή Poisson

Η κατανομή αυτή συνήθως εφαρμόζεται σε διακριτές μεταβλητές και αφορά τον αριθμό συμβάντων σε ορισμένο χρονικό ή χωρικό διάστημα. Μια τυχαία μεταβλητή X ακολουθεί την κατανομή Poisson με παράμετρο λ ($\lambda > 0$), όταν έχει συνάρτηση μάζας πυκνότητας πιθανότητας:

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

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

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

3.4.3 Αρνητική Διωνυμική Κατανομή (Negative Binomial)

Η χρήση αυτής της κατανομής ενδείκνυται σε περιπτώσεις όπου η διακύμανση των στοιχείων του δείγματος είναι μεγαλύτερη από τον μέσο όρο. Αυτό μπορεί να παρατηρηθεί σε φαινόμενα που παρουσιάζουν περιοδικές μεταβολές, όπως για παράδειγμα αριθμός αφίξεων οχημάτων σε μικρά χρονικά διαστήματα.

Μια τυχαία μεταβλητή X ακολουθεί την αρνητική διωνυμική κατανομή με παραμέτρους k και p (k θετικός ακέραιος, $0 < p < 1$), όταν έχει συνάρτηση μάζας πιθανότητας:

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

όπου $x = 0, 1, 2, 3, \dots$

3.5 Βασικά Μαθηματικά Πρότυπα

3.5.1 Γραμμική Παλινδρόμηση (Linear Regression)

Ως **ανάλυση παλινδρόμησης** (regression analysis) ορίζεται ο κλάδος της στατιστικής επιστήμης που εξετάζει τη σχέση μεταξύ δύο ή περισσότερων μεταβλητών, με σκοπό την πρόβλεψη της μίας από τις υπόλοιπες. Με τον όρο **εξαρτημένη μεταβλητή** εννοείται η μεταβλητή της οποίας η τιμή είναι η ζητούμενη, ενώ **ανεξάρτητη μεταβλητή** θεωρείται εκείνη που χρησιμοποιείται για την πρόβλεψη της εξαρτημένης. Οι ανεξάρτητες μεταβλητές παίρνουν καθορισμένες τιμές, ενώ η εξαρτημένη είναι τυχαία.

Ένα μαθηματικό πρότυπο αναπτύσσεται μέσω μιας στατιστικής διαδικασίας από όπου προκύπτουν εξισώσεις οι οποίες περιγράφουν τη σχέση μεταξύ της εξαρτημένης μεταβλητής και των ανεξάρτητων. Η κατάλληλη μέθοδος για την ανάπτυξη ενός μοντέλου επιλέγεται με κριτήριο το είδος της εξαρτημένης μεταβλητής (συνεχής ή διακριτή).

Εάν η εξαρτημένη μεταβλητή είναι συνεχής και ακολουθεί την κατανομή Gauss, η γραμμική παλινδρόμηση είναι από τις πιο διαδεδομένες στατιστικές τεχνικές. Στην πιο απλή μορφή της ονομάζεται **απλή γραμμική παλινδρόμηση** (simple linear regression). Σε μια τέτοια ανάλυση, υπάρχει μόνο μία ανεξάρτητη μεταβλητή X και μία εξαρτημένη Y , που προσεγγίζεται ως γραμμική συνάρτηση του X . Για κάθε τιμή x_i της X , η αντίστοιχη τιμή y_i της Y υπολογίζεται ως εξής:

$$y_i = \alpha + \beta * x_i + \varepsilon_i$$

Ο στόχος σε μία γραμμική παλινδρόμηση είναι ο υπολογισμός των παραμέτρων α και β , ώστε να εκφράζεται καλύτερα η γραμμική εξάρτηση των μεταβλητών. Κάθε ζεύγος τιμών (α , β) αντιστοιχεί και σε μία γραμμική σχέση που εκφράζεται γεωμετρικά ως ευθεία γραμμή.

Η παράμετρος α είναι ο **σταθερός όρος** (intercept) και υπολογίζεται ως η τιμή του Y για $X = 0$. Ο συντελεστής β του x είναι η **κλίση** (slope) της ευθείας ή αλλιώς **συντελεστής**

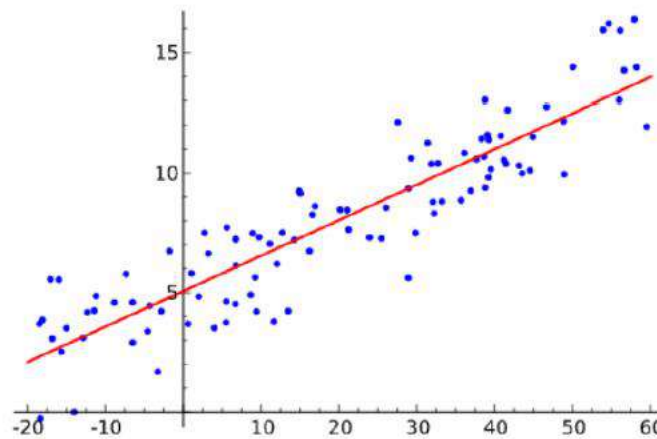
παλινδρόμησης (regression coefficient), και εκφράζει τη μεταβολή του Y για μια μεταβολή του X κατά μία μονάδα. Ο όρος ε_i ονομάζεται **σφάλμα παλινδρόμησης** (regression error).

Σε περίπτωση όπου το Y εξαρτάται γραμμικά από περισσότερες της μίας ανεξάρτητες μεταβλητές (X_1, X_2, X_3, \dots) γίνεται λόγος για **πολλαπλή γραμμική παλινδρόμηση** (multiple linear regression). Στην περίπτωση αυτή, η εξίσωση παίρνει τη μορφή:

$$y_i = \beta_0 + \beta_1 * x_{1i} + \beta_2 * x_{2i} + \beta_3 * x_{3i} + \dots + \beta_k * x_{ki} + \varepsilon_i$$

Ένας σημαντικός έλεγχος στην πολλαπλή γραμμική παλινδρόμηση είναι η αναγκαιότητα ή μη να συμπεριληφθούν όλες οι ανεξάρτητες μεταβλητές στο μοντέλο. Βασική προϋπόθεση είναι η **μηδενική συσχέτιση** μεταξύ των ανεξάρτητων μεταβλητών. Πρέπει, δηλαδή, ο συντελεστής συσχέτισης ρ δύο ανεξάρτητων μεταβλητών να τείνει στο 0, για κάθε τυχαίο ζεύγος x_i, x_j με $i \neq j$.

Σε ένα πρότυπο πολλαπλής γραμμικής παλινδρόμησης, χρησιμοποιείται η **μέθοδος των ελάχιστων τετραγώνων** (least squares method). Ο προσδιορισμός των παραμέτρων β_i σχηματίζει μια προσεγγιστική ευθεία, που ονομάζεται ευθεία παλινδρόμησης του Y πάνω στο X. Παρακάτω φαίνεται ένα τυπικό διάγραμμα μιας τέτοιας ευθείας.



Διάγραμμα 3.1: Ευθεία ελάχιστων τετραγώνων

Σκοπός της μεθόδου είναι η ελαχιστοποίηση του αθροίσματος των τετραγώνων των κατακόρυφων αποστάσεων των σημείων (X, Y) από την ευθεία.

3.5.2 Λογαριθμοκανονική Παλινδρόμηση (Log-normal Regression)

Το πρότυπο αυτό συσχετίζει δύο ή περισσότερες ανεξάρτητες μεταβλητές με τον δεκαδικό λογάριθμο της εξαρτημένης μεταβλητής. Η εξίσωση είναι και σε αυτή την περίπτωση γραμμική.

$$\log y_i = \beta_0 + \beta_1 * x_{1i} + \beta_2 * x_{2i} + \beta_3 * x_{3i} + \dots + \beta_k * x_{ki} + \varepsilon_i$$

Στην λογαριθμοκανονική παλινδρόμηση οι συντελεστές των μεταβλητών είναι οι συντελεστές της γραμμικής παλινδρόμησης και υπολογίζονται με βάση την αρχή των ελάχιστων τετραγώνων.

3.6 Γεωγραφικά Σταθμισμένη Παλινδρόμηση (GWR)

Στο υποκεφάλαιο αυτό παρουσιάζονται ορισμένες βασικές αρχές της μεθόδου GWR που χρησιμοποιήθηκε στην χωρική ανάλυση των δεδομένων, καθώς και στοιχεία λειτουργίας του **στατιστικού λογισμικού R** με το οποίο έγιναν οι αναλύσεις.

Ως τεχνική, η **γεωγραφικά σταθμισμένη παλινδρόμηση** εξετάζει φαινόμενα μη σταθερότητας σε έναν χάρτη, περιοχές δηλαδή όπου οι τοπικά σταθμισμένοι συντελεστές παλινδρόμησης απομακρύνονται από τις καθολικές τους τιμές.

Στο λογισμικό R, αρχικά απαιτείται η φόρτωση των πακέτων 'library(sf)' και 'library(GWmodel)'.

```
#convert dataframe into spatial
library(sf)
library(GWmodel)
Municipal_sp <- st_as_sf(Municipal, coords = c("Long", "Lat"), crs = 4326)
```

Εικόνα 3.1: Κώδικας R έναρξης χωρικής μεθόδου

Στον παραπάνω κώδικα, το αρχείο δεδομένων της Διπλωματικής Εργασίας μετατρέπεται σε μορφή επεξεργάσιμη από χωρικά μοντέλα ανάλυσης. Στην τελευταία γραμμή του κώδικα, η μετατροπή χρησιμοποιεί ως συντεταγμένες το γεωγραφικό μήκος και πλάτος του κέντρου κάθε Δήμου (εδώ 'Long' και 'Lat'). Στη συνέχεια, χρειάζεται να καθοριστεί το **εύρος ζώνης**, καθώς και οι μεταβλητές που θα χρησιμοποιηθούν στο μοντέλο. Τέλος, ως κριτήριο για το χωρικό μοντέλο επιλέγεται ο δείκτης AIC. Ο αντίστοιχος κώδικας στο λογισμικό R παρουσιάζεται παρακάτω.

```
# determine kernel bandwidth
class(Municipal_sp)

# convert to sp
Municipal.sp = as(Municipal_sp, "Spatial")
bw <- bw.gwr(logFtpAr ~ Uni_grad_perc + Fem_perc + Work_perc + hearth_cars_own + GDP2019,
             approach = "AIC",
             adaptive = T,
             data=Municipal.sp)

# fit GWR model
m.gwr <- gwr.basic(logFtpAr ~ Uni_grad_perc + Fem_perc + work_perc + hearth_cars_own + GDP2019,
                  adaptive = T,
                  data = Municipal.sp,
                  bw = bw)

m.gwr
```

Εικόνα 3.2: Κώδικας R μεθόδου GWR

Ο δείκτης AIC είναι το **Κριτήριο Πληροφοριών Akaike** (Akaike Information Criterion) που αναπτύχθηκε την δεκαετία του 70' από τον Ιάπωνα στατιστικό Hirotsugu Akaike (1974), και αποτελεί έναν εκτιμητή της προγνωστικής ακρίβειας ενός στατιστικού μοντέλου. Συχνά, χρησιμοποιείται η διορθωμένη τιμή του δείκτη, που συμβολίζεται ως AICc (corrected AIC), όπως και στην παρούσα Διπλωματική Εργασία. Το λογισμικό R εκτελεί διαδοχικές δοκιμές με διάφορες τιμές εύρους ζώνης μέχρι να επιτύχει την ελαχιστοποίηση του δείκτη AICc. Στη συνέχεια, το εύρος ζώνης που υπολογίστηκε χρησιμοποιείται για την προσαρμογή του χωρικού μοντέλου.

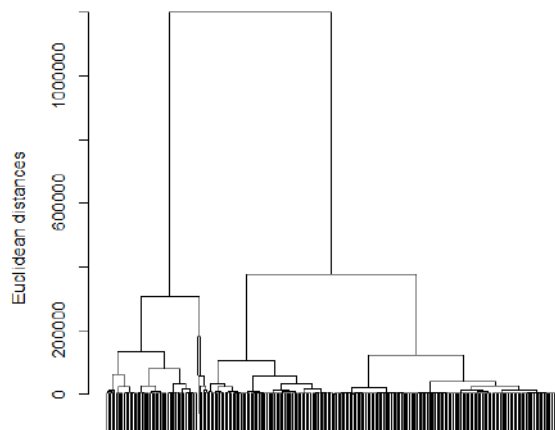
3.7 Ομαδοποίηση δεδομένων (clustering)

Όπως είναι προφανές, ένας αλγόριθμος ομαδοποίησης δημιουργεί ομάδες ή **συστάδες δεδομένων** (clusters) που διατηρούν εσωτερική ομοιομορφία αλλά παραμένουν διακριτές σε σχέση με άλλες συστάδες.

Ένα βασικό χαρακτηριστικό της ομαδοποίησης είναι η **ευκλείδεια απόσταση** μεταξύ δύο ομάδων δεδομένων. Γενικά, όσο μεγαλύτερη αυτή η απόσταση τόσο καλύτερη είναι η ανάλυση.

Από τους πιο συνηθισμένους αλγόριθμους ομαδοποίησης είναι αυτός της **ιεραρχικής ομαδοποίησης** (hierarchical clustering). Η κεντρική αρχή του αλγορίθμου είναι η ενοποίηση των δύο πλησιέστερων ομάδων σε μία ευρύτερη. Η διαδικασία αρχίζει με τον υπολογισμό της ευκλείδειας απόστασης μεταξύ δεδομένων ανά ζεύγη και ο αλγόριθμος ξεκινά να δημιουργεί τις συστάδες. Έπειτα, συγχωνεύει τις πλησιέστερες και υπολογίζει την απόσταση μεταξύ των νέων συστάδων που προέκυψαν, και η διαδικασία επαναλαμβάνεται.

Η ιεραρχική ομαδοποίηση κατατάσσει τα δεδομένα σε ένα διάγραμμα που ονομάζεται **δενδρόγραμμα** (dendrogram) και είναι της μορφής που φαίνεται στην παρακάτω εικόνα.



Διάγραμμα 3.2 : Δενδρόγραμμα ιεραρχικής ομαδοποίησης

Ο κατακόρυφος άξονας (y) του δενδρογράμματος αντιπροσωπεύει την απόσταση συγχώνευσης ή ευκλείδεια απόσταση.

Ένα ουσιαστικό ζήτημα είναι η απόφαση ως προς τον αριθμό των συστάδων. Γενικά, ένας γρήγορος τρόπος είναι μία οριζόντια τομή στο δενδρόγραμμα. Το επιθυμητό είναι η τομή να μπορεί να διανύσει όσο μεγαλύτερη απόσταση γίνεται από πάνω προς τα κάτω, χωρίς να τέμνει κάποια διακλάδωση.

3.8 Στατιστικοί έλεγχοι & κριτήρια επιλογής

Ένα στατιστικό μοντέλο απαιτεί μια διαδικασία αξιολόγησης προκειμένου να καθορισθεί ο βαθμός στον οποίο εξυπηρετεί την ανάλυση. Χρειάζεται, επομένως, να εκτιμηθεί η επιρροή των ανεξάρτητων μεταβλητών στην εξαρτημένη, η μεταξύ τους συσχέτιση, καθώς και η απόδοση του μοντέλου με βάση ορισμένους δείκτες.

Παρακάτω παρουσιάζονται οι συνηθέστεροι έλεγχοι που πραγματοποιούνται στο πλαίσιο ενός μοντέλου παλινδρόμησης.

➤ **Συντελεστής R^2**

Ο πλέον δημοφιλέστερος στατιστικός δείκτης. Ορίζεται ως το **ποσοστό της συνολικής μεταβλητότητας** που εξηγεί το μοντέλο. Παίρνει τιμές από το 0 μέχρι το 1, με τιμές κοντά στο 0 να αντιστοιχούν σε ανεπαρκές μοντέλο, και τιμές κοντά στο 1 να δείχνουν ένα μοντέλο που προσαρμόζεται τέλεια στα δεδομένα. Ωστόσο, στις περισσότερες περιπτώσεις ένα R^2 κοντά στο 0.5 θεωρείται πλήρως αποδεκτό, καθώς στα πραγματικά προβλήματα τα δεδομένα μπορεί να έχουν σημαντικές διακυμάνσεις.

➤ **Συντελεστές παλινδρόμησης (regression coefficients)**

Σε κάθε μοντέλο υπάρχουν οι συντελεστές της **τομής** (intercept) και της **κλίσης** (slope) που συμβολίζονται στην εξίσωση της ευθείας ως “a” και “b” αντίστοιχα. Ένας γρήγορος τρόπος αξιολόγησης του μοντέλου είναι να εξεταστεί εάν τα πρόσημα των τιμών της κλίσης εξηγούνται λογικά.

Για παράδειγμα, έστω ότι ένα μοντέλο παλινδρόμησης εξετάζει την μεταβολή των οδικών ατυχημάτων σε σχέση με τον αριθμό των οδηγών. Προφανώς, όσοι περισσότεροι είναι οι οδηγοί, τόσο περισσότερα αναμένεται να είναι και τα οδικά ατυχήματα. Υπάρχει δηλαδή μια θετική συσχέτιση μεταξύ των μεταβλητών. Στα αποτελέσματα του μοντέλου αναμένεται η τιμή της κλίσης για τον αριθμό των οδηγών να έχει θετικό πρόσημο.

➤ **Υπολείμματα (residuals)**

Τα υπολείμματα του μοντέλου εκφράζουν τις διαφορές ανάμεσα στις προβλεπόμενες τιμές και τις πραγματικές. Ιδανικά, το άθροισμα των υπολειμμάτων πρέπει να είναι κοντά στο 0.

➤ **Η τιμή p (p-value)**

Ο δείκτης αυτός προκύπτει στα αποτελέσματα του μοντέλου για κάθε ανεξάρτητη μεταβλητή, και εκφράζει το κατά πόσο αυτή έχει νόημα σαν προγνωστικός παράγοντας. Έτσι, υποδεικνύει εάν η υπόθεση πρέπει να απορριφθεί ή όχι. Το επιθυμητό είναι η τιμή p να είναι μικρότερη από 0.05. Όσο χαμηλότερη είναι, τόσο πιο κατάλληλη είναι η μεταβλητή για το μοντέλο.

➤ **Κριτήριο Πληροφοριών Akaike (AIC)**

Επιθυμητό είναι ο δείκτης αυτός να είναι όσο μικρότερος γίνεται. Η τιμή του AIC χρησιμοποιείται ως ένα γρήγορο και αποτελεσματικό μέτρο σύγκρισης της προγνωστικής ακρίβειας μεταξύ στατιστικών μοντέλων. Στο πλαίσιο της παρούσας Διπλωματικής Εργασίας, ο δείκτης AIC, καθώς και η διορθωμένη του τιμή (AICc), χρησιμοποιούνται για την αξιολόγηση των μοντέλων χωρικής παλινδρόμησης σε σχέση με τα κλασικά μοντέλα.

➤ **Συντελεστής Πληθωρισμού Διακύμανσης (Variance Inflation Factor – VIF)**

Σε μία ανάλυση παλινδρόμησης με πολλαπλές μεταβλητές, είναι πιθανό να υπάρχει υψηλή **συσχέτιση μεταξύ ορισμένων μεταβλητών**, κάτι που δεν είναι επιθυμητό. Ο Συντελεστής Πληθωρισμού Διακύμανσης είναι ένας εύκολος τρόπος να εντοπιστεί αυτό το πρόβλημα.

Ο συντελεστής δέχεται τιμές από το 1 μέχρι το θετικό άπειρο, και γενικά για τιμές μικρότερες του 5 θεωρείται ότι δεν υπάρχει πρόβλημα στο μοντέλο.

➤ **Συντελεστής εμπιστοσύνης t**

Βασίζεται στην **κατανομή του Student** και προσδιορίζει την στατιστική σημαντικότητα των ανεξάρτητων μεταβλητών. Ο συντελεστής t ορίζεται από την παρακάτω εξίσωση:

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

Όπου s. e. το τυπικό σφάλμα (standard error).

Ο συντελεστής προφανώς αυξάνεται όσο το τυπικό σφάλμα μειώνεται. Όσο μεγαλύτερη είναι η απόλυτη τιμή του t_{stat} , τόσο μεγαλύτερη επιρροή θα έχει η συγκεκριμένη μεταβλητή στο αποτέλεσμα. Παρακάτω φαίνονται οι κρίσιμες τιμές του συντελεστή ανά επίπεδο εμπιστοσύνης.

Βαθμοί Ελευθερίας	Επίπεδο Εμπιστοσύνης				
	0,900	0,950	0,975	0,990	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

Πίνακας 3.1: Κρίσιμες τιμές συντελεστή t_{stat}

Το πλήθος των βαθμών ελευθερίας (degrees of freedom) ορίζεται ως το πλήθος δείγματος μείον ένα.

➤ **Ελαστικότητα (elasticity)**

Η ελαστικότητα εκφράζει την **ευαισθησία της εξαρτημένης μεταβλητής** στη μεταβολή μίας ή περισσότερων από τις ανεξάρτητες μεταβλητές. Η ελαστικότητα είναι ένα αδιάστατο μέγεθος και δεν εξαρτάται από τις μονάδες μέτρησης των μεταβλητών. Ουσιαστικά, πρόκειται για ποσοστιαία μεταβολή της εξαρτημένης

μεταβλητής που οφείλεται σε μεταβολή της ανεξάρτητης κατά 1%. Σε περίπτωση γραμμικού μοντέλου, η ελαστικότητα προκύπτει από την παρακάτω σχέση:

$$e_i = \left(\frac{\Delta Y_i}{\Delta X_i} \right) * \left(\frac{X_i}{Y_i} \right) = \beta_i * \left(\frac{X_i}{Y_i} \right)$$

Χρειάζεται να σημειωθεί ότι η παραπάνω εξίσωση εφαρμόζεται μόνο για συνεχείς μεταβλητές, ενώ στην περίπτωση των διακριτών μεταβλητών χρησιμοποιείται η ψευδοελαστικότητα.

Ο **βαθμός επιρροής** των ανεξάρτητων μεταβλητών στις εξαρτημένες σε ένα μαθηματικό πρότυπο εκφράζεται μέσω του μεγέθους της **σχετικής επιρροής**, για τον υπολογισμό του οποίου χρησιμοποιείται η θεωρία της ελαστικότητας. Αρχικά, υπολογίζεται η τιμή της εξαρτημένης μεταβλητής όπως προκύπτει από την μαθηματική εξίσωση του μοντέλου. Η ελαστικότητα (e_i) υπολογίζεται ως το γινόμενο της τιμής της ανεξάρτητης μεταβλητής με τον συντελεστή της (με το πρόσημό του), και διαιρεμένο με την αντίστοιχη τιμή της εξαρτημένης μεταβλητής.

Ο μέσος όρος της ελαστικότητας όλων των περιπτώσεων (των Δήμων στην παρούσα Διπλωματική Εργασία) είναι και η τελική της τιμή για την συγκεκριμένη ανεξάρτητη μεταβλητή. Για τον υπολογισμό της σχετικής επιρροής διαιρούνται όλες οι τιμές e_i με την μικρότερη κατά απόλυτη τιμή ελαστικότητα.

Κεφάλαιο 4: Συλλογή & Επεξεργασία Στοιχείων

4.1 Εισαγωγή

Όπως προαναφέρθηκε στο Κεφάλαιο 1, σκοπός της παρούσας Διπλωματικής Εργασίας είναι η **χωρική ανάλυση οδικών ατυχημάτων** όπως αυτά καταγράφηκαν στους Δήμους της Ελλάδας. Σε αρχικό στάδιο, επιδιώκεται η διερεύνηση της επιρροής ορισμένων παραγόντων κοινωνικοοικονομικής και συγκοινωνιακής φύσης στα ατυχήματα. Πραγματοποιήθηκε λοιπόν βιβλιογραφική ανασκόπηση συναφών μελετών οδικής ασφάλειας, και έπειτα θεμελιώθηκε το θεωρητικό υπόβαθρο που βασίζεται στην στατιστική ανάλυση δεδομένων. Στο παρόν κεφάλαιο παρουσιάζεται η διαδικασία συλλογής των απαραίτητων στοιχείων, καθώς και η κατάλληλη επεξεργασία τους.

Αρχικά, εξηγούνται τα είδη των δεδομένων που χρησιμοποιήθηκαν, και πως αυτά συμβάλλουν στην επίτευξη του στόχου της Διπλωματικής Εργασίας. Στη συνέχεια, ακολουθεί η πρώτη επεξεργασία των στοιχείων με το Microsoft Excel, και η κατάλληλη διαμόρφωσή τους ώστε να μπορούν να υποστούν περαιτέρω επεξεργασία με το στατιστικό λογισμικό R.

4.2 Συλλογή Δεδομένων

Η συλλογή των στοιχείων αφορά το σύνολο των 305 δήμων που αποτελούν την Ελλάδα. Τα δεδομένα συγκεκριμένα είναι τα ακόλουθα:

- Αριθμός θανάτων σε οδικά ατυχήματα κατά την περίοδο 2016 – 2020
- Πληθυσμός (απογραφή 2011)
- Φύλο (ποσοστό γυναικών, ποσοστό ανδρών)
- Αριθμός και ποσοστό ηλικιωμένων
- Αριθμός και ποσοστό απόφοιτων λυκείου
- Αριθμός και ποσοστό τελειόφοιτων πανεπιστημίου
- Αριθμός και ποσοστό απασχολούμενων σε εργασία
- Αριθμός νοικοκυριών με ιδιοκτησία ΙΧ
- Στόλος οχημάτων 2020 (ΙΧ, φορτηγά, μοτοσικλέτες, λεωφορεία)
- Ακαθάριστο Εθνικό Προϊόν 2019
- Γεωγραφικές συντεταγμένες κέντρου δήμου
- Έκταση Δήμου

Ο αριθμός των θανάτων σε οδικά ατυχήματα είναι το βασικό στοιχείο πάνω στο οποίο βασίζεται η Διπλωματική Εργασία. Τα πληθυσμιακά στοιχεία, δηλαδή ο συνολικός πληθυσμός, το φύλο και οι ηλικιωμένοι αποτελούν ουσιαστικούς παράγοντες σε οποιαδήποτε στατιστική ανάλυση. Κοινωνικά και οικονομικά χαρακτηριστικά όπως το επίπεδο εκπαίδευσης, ο αριθμός εργαζόμενων, η ιδιοκτησία ΙΧ και το ΑΕΠ επηρεάζουν σημαντικά τα οδικά ατυχήματα, όπως προφανώς και ο αριθμός και ο τύπος οχημάτων. Τέλος, οι γεωγραφικές συντεταγμένες είναι ένα απαραίτητο δεδομένο για την παρούσα Διπλωματική Εργασία, καθώς χρησιμοποιούνται για την ανάπτυξη των χωρικών μοντέλων της μεθόδου GWR, όπως αναφέρθηκε στο Κεφάλαιο 3.

Για την εύρεση των γεωγραφικών συντεταγμένων των Δήμων χρησιμοποιήθηκαν οι χάρτες Google Maps. Η συλλογή όλων των υπόλοιπων δεδομένων έγινε από την Ελληνική Στατιστική Αρχή (ΕΛΣΤΑΤ).

Η ΕΛΣΤΑΤ αποτελεί μια ανεξάρτητη διοικητική αρχή της οποίας η λειτουργία υπάγεται στον έλεγχο της Βουλής των Ελλήνων. Η πρώτη υπηρεσία απογραφών στην Ελλάδα υπήρξε το Γραφείο Δημοσίου Οικονομίας που ιδρύθηκε το 1834, στο πλαίσιο του Υπουργείου Εσωτερικών. Σταδιακά, η στατιστική ανάλυση ανεξαρτητοποιήθηκε ως υπηρεσία, ενώ το 1986 συστάθηκε η Εθνική Στατιστική Υπηρεσία της Ελλάδος (ΕΣΥΕ), που υπαγόταν στο Υπουργείο Εθνικής Οικονομίας. Η ΕΛΣΤΑΤ στην σημερινή μορφή της ιδρύθηκε το 2010.

Σκοπός της υπηρεσίας είναι η πραγματοποίηση στατιστικών ερευνών και η παραγωγή στατιστικών αποτελεσμάτων για το σύνολο της Ελλάδας, σε όλους τους τομείς δημόσιας και ιδιωτικής δραστηριότητας. Τα δεδομένα της ΕΛΣΤΑΤ χρησιμοποιούνται και εκτός Ελλάδας από οργανισμούς όπως η Ευρωπαϊκή Στατιστική Υπηρεσία (Eurostat), η UNESCO και ο Οργανισμός Ηνωμένων Εθνών (UN).

4.3 Βάση δεδομένων Διπλωματικής Εργασίας

Αρχικά, προκειμένου τα στοιχεία να επεξεργαστούν, απαιτήθηκε η δημιουργία μιας ενιαίας βάσης δεδομένων σε ένα αρχείο του προγράμματος Excel. Ο πίνακας του αρχείου αυτού περιλαμβάνει στην μία διάσταση τους Δήμους της χώρας, ενώ στην άλλη διάσταση όλα τα προαναφερθέντα δεδομένα, καθώς και ορισμένα που προέκυψαν από πρωτογενή επεξεργασία.

Municipality	Fatal2016	Fatal2017	Fatal2018	Fatal2019	Fatal2020	Tot. Fatal	Pop2011	Fatal/Pop	logRP	Fem. Pop2011	Fem. perc	Mal. Pop2011	Mal. perc	elder2011	elder. perc	Uni. Grad2011	Uni. grad. perc	HS. Grad2011	HS. grad. perc
ΔΗΜΟΣ ΚΟΜΟΤΗΝΗΣ	5	5	2	7	2	21	66919	3.14	0.4667	34595	48.3	32354	51.7	9381	85.2	9754	85.5	15503	75.8
ΔΗΜΟΣ ΑΡΦΙΜΟΝΙΩΝ	1	0	1	0	0	2	16577	1.21	0.0815	8364	49.4	8183	50.6	2448	85.2	382	87.6	688	95.9
ΔΗΜΟΣ ΙΑΣΜΟΥ	2	0	0	3	1	6	13810	4.34	0.6380	7079	48.7	6731	51.3	2030	85.2	439	86.6	1183	91.6
ΔΗΜΟΣ ΜΑΡΙΝΕΙΑΣ - ΙΑΦΙΩΝ	3	0	0	1	0	4	14733	2.71	0.4338	7446	49.5	7287	50.5	2176	85.2	675	85.4	1516	80.7
ΔΗΜΟΣ ΔΡΑΜΑΣ	3	3	3	3	2	14	88944	2.38	0.3757	30640	48.0	28304	52.0	11879	78.8	9130	84.5	12634	78.6
ΔΗΜΟΣ ΔΟΞΑΤΟΥ	0	1	2	0	1	4	14516	2.78	0.4402	7531	48.1	6985	51.9	2326	78.8	1145	82.8	2129	85.3
ΔΗΜΟΣ ΚΑΤΟ ΜΕΥΡΟΚΟΠΟΥ	1	2	0	1	0	4	7860	5.00	0.7066	3867	50.4	3993	49.6	1584	78.8	282	86.4	1202	84.7
ΔΗΜΟΣ ΠΑΡΑΝΕΣΤΙΟΥ	0	0	0	0	0	0	3001	0.00	0.0000	1930	50.5	1071	49.5	766	78.8	234	84.0	687	82.4
ΔΗΜΟΣ ΠΡΟΙΩΤΣΙΑΝΗΣ	1	0	0	0	1	2	13096	1.53	0.1849	8747	48.4	4349	51.6	2833	78.8	725	84.5	1977	84.9
ΔΗΜΟΣ ΑΓΕΛΑΜΟΥΡΙΩΤΩΝ	5	6	5	1	6	23	72959	3.15	0.4688	37184	48.0	35775	51.0	13066	80.1	12511	85.9	17257	75.3
ΔΗΜΟΣ ΔΑΦΝΙΟΤΕΜΟΥ	2	0	2	0	1	5	19493	2.57	0.4091	9353	52.0	10140	48.0	3491	80.1	1654	81.5	3468	82.2
ΔΗΜΟΣ ΟΡΕΣΤΙΑΔΑΣ	5	3	1	2	1	12	37695	3.18	0.5029	18141	51.9	19554	48.1	6751	80.1	3918	80.4	7442	80.3
ΔΗΜΟΣ ΣΑΜΟΘΡΑΚΗΣ	0	0	1	0	1	2	2859	7.00	0.9448	1263	88.6	1596	43.8	512	86.1	343	88.0	688	78.7
ΔΗΜΟΣ ΣΟΥΦΛΙΟΥ	1	2	1	0	2	6	14941	4.02	0.6038	7321	51.0	7620	49.0	2678	80.1	1026	83.1	2356	84.2
ΔΗΜΟΣ ΘΕΣΣΟΝ	2	1	0	1	1	5	13770	3.63	0.5900	6965	49.3	6765	50.7	292	87.9	1127	81.8	2332	83.8
ΔΗΜΟΣ ΚΙΒΔΙΑΣ	4	1	0	4	1	10	70501	1.42	0.1518	36499	48.2	34002	51.8	12608	82.1	12033	82.9	17057	75.8
ΔΗΜΟΣ ΝΕΣΤΟΥ	1	0	2	1	5	15	22331	6.72	0.8272	11288	49.5	11045	50.5	3993	80.1	1856	81.7	3451	84.6
ΔΗΜΟΣ ΠΑΓΓΑΚΟΥ	4	5	14	3	1	27	32085	8.42	0.9251	16278	49.3	15807	50.7	5738	80.1	2501	80.2	5065	84.1
ΔΗΜΟΣ ΞΑΝΘΗΣ	1	2	2	4	1	10	65133	1.54	0.1882	20804	49.6	32329	50.4	7465	86.5	10862	83.3	14207	78.2
ΔΗΜΟΣ ΑΒΘΗΡΩΝ	3	4	2	3	1	13	19005	6.84	0.8351	9558	49.2	9347	50.8	2176	86.5	685	85.5	2436	87.2
ΔΗΜΟΣ ΜΥΚΗΣ	4	2	1	0	0	4	5540	2.57	0.4106	7853	49.5	7687	50.5	1781	86.5	332	87.9	740	95.2
ΔΗΜΟΣ ΤΥΠΕΡΟΥ	1	0	1	5	0	10	11544	8.66	0.8376	5705	50.6	5839	49.4	1323	86.5	376	86.6	1157	80.0
ΔΗΜΟΣ ΘΕΣΣΑΛΟΝΙΚΗΣ	9	11	8	6	2	36	325182	1.11	0.0442	170712	45.7	148470	54.3	41719	87.2	81912	74.6	66250	96.5
ΔΗΜΟΣ ΑΜΠΕΛΟΚΗΤΩΝ - ΜΕΜΕΛΕΝΗΣ	1	1	4	2	0	8	52127	1.53	0.1880	27435	47.4	24721	52.6	6888	87.2	6750	87.1	12939	75.2
ΔΗΜΟΣ ΒΙΘΑΣ	7	7	4	3	2	23	23478	9.80	0.9911	11736	50.0	11742	50.0	3912	87.2	1527	85.5	4088	82.7
ΔΗΜΟΣ ΔΕΛΤΑ	7	5	2	7	3	24	45836	5.24	0.7190	23010	49.8	22829	50.2	5881	87.2	3808	81.5	9196	80.0
ΔΗΜΟΣ ΘΕΡΜΑΚΟΥ	1	6	0	0	5	12	50394	2.39	0.3719	25487	49.3	24777	50.7	5446	87.2	8546	83.0	12688	74.8
ΔΗΜΟΣ ΘΕΡΜΗΣ	8	3	3	10	6	30	53201	5.64	0.7812	27884	49.1	25327	50.9	6626	87.2	11639	77.8	12911	77.4
ΔΗΜΟΣ ΚΑΛΑΜΑΡΙΑΣ	0	0	0	3	1	4	61516	0.44	-0.3594	48349	47.2	43180	52.8	11741	80.1	2837	70.9	23159	74.7
ΔΗΜΟΣ ΚΟΡΦΕΣΙΩΝ - ΕΥΟΣΜΟΥ	1	6	0	2	3	12	101753	1.18	0.0718	52140	48.8	49613	51.2	13654	87.2	14005	86.2	26799	73.8
ΔΗΜΟΣ ΛΑΓΚΑΔΑΣ	4	3	4	5	4	20	41103	4.87	0.8872	20745	49.5	20358	50.5	6273	87.2	5496	81.5	7800	81.4
ΔΗΜΟΣ ΝΕΑΡΧΩΝ - ΣΥΜΕΩΝ	2	0	1	1	1	4	64741	0.47	-0.3250	44335	47.7	40400	52.3	10872	87.2	14941	80.4	21048	74.5
ΔΗΜΟΣ ΠΑΥΛΟΥ ΜΕΛΑ	1	7	1	5	2	17	99245	1.71	0.2337	50909	48.7	48337	51.3	12733	87.2	13335	86.6	23540	74.5
ΔΗΜΟΣ ΠΥΛΑΙΑΣ - ΧΟΡΤΙΑΤΗ	2	3	2	5	6	18	70110	2.57	0.4095	35814	49.8	34296	51.1	8065	87.2	21689	86.1	16395	76.7
ΔΗΜΟΣ ΣΑΚΚΗΔΙΩΝ	6	3	4	3	3	19	33873	5.64	0.7515	16918	49.8	16756	50.2	4320	87.2	3023	81.0	5668	82.2
ΔΗΜΟΣ ΣΦΑΚΙΑΚΤΡΟΥ	2	0	3	7	2	14	38317	3.65	0.5627	18825	50.6	19492	49.1	4916	87.2	1745	78.8	3043	75.9
ΔΗΜΟΣ ΒΕΡΟΙΑΣ	6	3	4	5	3	23	66547	3.46	0.5386	34030	48.9	32517	51.1	10703	83.0	9886	85.0	12488	81.2
ΔΗΜΟΣ ΑΓΕΡΙΝΩΡΕΩΝ	2	3	4	1	3	13	41570	3.13	0.4952	20810	49.6	20750	50.1	6886	83.9	2941	82.9	8412	84.6
ΔΗΜΟΣ ΝΙΦΩΤΑΣ	3	5	2	1	2	13	32484	4.00	0.8021	16330	49.7	16144	50.3	5226	85.9	3435	86.4	5782	82.2
ΔΗΜΟΣ ΚΙΛΙΚΙΑΣ	2	0	3	2	7	14	51828	2.70	0.4307	26178	49.6	25746	50.4	5944	78.8	5526	89.1	10524	79.7
ΔΗΜΟΣ ΠΙΣΑΡΙΔΑΣ	6	2	8	1	4	21	26493	7.57	0.8875	14882	50.5	14403	49.5	6035	78.8	2307	81.9	5146	81.9
ΔΗΜΟΣ ΕΛΕΥΤΕΡΑΣ	4	1	1	0	2	8	28814	2.76	0.4435	14631	49.2	14183	50.8	5056	82.5	3917	86.4	3816	79.8
ΔΗΜΟΣ ΑΜΑΛΙΑΣ	3	0	0	3	1	7	27556	2.54	0.4049	13814	49.9	13742	50.1	4835	82.5	1804	83.1	4074	85.2
ΔΗΜΟΣ ΠΕΛΑΙΑΣ	10	3	3	2	5	23	63122	3.84	0.5919	31939	49.4	31183	50.6	11075	83.0	6301	80.0	11082	82.5
ΔΗΜΟΣ ΣΚΥΔΡΑΣ	2	2	0	0	2	6	20188	2.97	0.4731	10214	49.4	9974	50.6	3542	82.5	1473	82.7	3365	83.4
ΔΗΜΟΣ ΚΑΤΕΡΙΝΗΣ	4	1	2	6	3	16	85851	1.86	0.2704	44282	48.4	41569	51.6	13683	84.1	12213	85.8	16779	80.5
ΔΗΜΟΣ ΔΟΥ - ΟΥΛΙΠΟΥ	3	4	2	3	2	14	25606	5.45	0.7367	12340	51.6	13322	48.1	4091	84.1	2474	80.4	4521	82.4

Πίνακας 4.1: Βάση δεδομένων (1)

Παραπάνω παρουσιάζεται ένα τυπικό απόσπασμα της βάσης δεδομένων της Διπλωματικής Εργασίας. Στην στήλη A καταγράφονται οι Δήμοι, στις στήλες B έως και F ο αριθμός θανάτων σε οδικά ατυχήματα για την αντίστοιχη χρονιά, ενώ στην στήλη G η μεταβλητή “Tot_Fatal” αποτελεί το άθροισμα των στηλών B έως και F. Στην στήλη H καταγράφεται ο πληθυσμός της απογραφής του 2011.

Στην στήλη I, η μεταβλητή “FatalpPop” (fatalities per population) εκφράζει τους θανάτους ως προς τον πληθυσμό του κάθε Δήμου, και προκύπτει συνεπώς από διαίρεση των στηλών G και H. Για διευκόλυνση, έγινε στρογγυλοποίηση του αποτελέσματος στο δεύτερο δεκαδικό ψηφίο. Στην στήλη J, και με ακρίβεια τεσσάρων δεκαδικών ψηφίων, παρουσιάζεται η μεταβλητή “logFpP” που εκφράζει τον δεκαδικό λογάριθμο των θανάτων ανά πληθυσμό.

Στις στήλες K και L οι μεταβλητές “Fem_Pop2011” (female population) και “Fem_perc” (female percentage), αποτελούν τον γυναικείο πληθυσμό του 2011, και το ποσοστό των γυναικών σε σχέση με τον συνολικό, αντίστοιχα. Οι αντίστοιχες μεταβλητές για τους άνδρες παρουσιάζονται στις στήλες M και N. Παρόμοια, οι στήλες O και P περιέχουν τον πληθυσμό και το ποσοστό των ηλικιωμένων.

Η στήλη Q περιέχει την μεταβλητή “Uni_Grad2011” (university graduates) που αποτελεί τον αριθμό των τελειόφοιτων ανώτερης εκπαίδευσης με βάση την απογραφή του 2011. Το αντίστοιχο ποσοστό βρίσκεται στην στήλη R. Παρόμοια, η μεταβλητή “HS_Grad2011” (high school graduates) της στήλης S αντιστοιχεί στους απόφοιτους λυκείου, με το ποσοστό τους να βρίσκεται στην στήλη T.

	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
	Work2011	Work perc	Lat	Long	Area	FtpAr	logFtpAr	hearth cars own	cars2020	trucks2020	motor2020	buses2020	GDP2019
1	20631	69.2	41,11854	25,402465	644.9	0.033	-1,4873	17,220	19298	8587	6331	97	6282
2	6658	59.8	41,08134	25,9954693	771.2	0.003	-2,5861	3,529	4781	2127	1568	24	1556
3	4505	67.4	41,12794	25,1849283	485.3	0.012	-1,9079	3,028	3983	1772	1306	20	1296
4	4004	72.8	40,90522	25,5173564	641.8	0.006	-2,2053	3,313	4249	1891	1394	21	1383
5	18021	69.4	41,1501	24,1464002	840	0.017	-1,7782	15446	17218	7216	5925	179	6262
6	3810	73.8	41,0959	24,2310597	243.4	0.016	-1,7843	3572	4240	1777	1459	44	1542
7	2150	72.6	41,34336	23,8670027	873.6	0.005	-2,3393	1462	2296	962	790	24	835
8	808	79.3	41,26681	24,5009114	1029.4	0.000	0,0000	782	1139	478	392	12	414
9	3231	75.3	41,17977	23,9702785	481.8	0.004	-2,3818	3008	3817	1600	1313	40	1388
10	25044	65.7	40,84578	25,8735251	1217	0.019	-1,7236	22018	27879	11119	7286	90	6086
11	5334	72.6	41,34772	26,494498	565.4	0.009	-2,0534	4563	7395	2971	1947	24	1626
12	12657	66.4	41,50348	26,5281306	955.6	0.013	-1,9011	10231	14301	5745	3764	47	3144
13	937	67.2	40,47447	25,5244964	178	0.011	-1,9494	817	1085	438	285	4	238
14	4471	70.1	41,19523	26,2985671	1325.7	0.005	-2,3443	3426	5668	2277	1492	18	1246
15	4478	67.5	40,77695	24,7084721	380.1	0.013	-1,8809	3730	6660	2467	2180	39	1565
16	22033	68.7	40,93803	24,4113451	351.4	0.028	-1,5458	18809	34098	12631	11161	200	8015
17	7060	68.4	40,98585	24,7002435	678.8	0.022	-1,6557	5896	10801	4001	3535	63	2539
18	9849	69.3	40,91549	24,2549471	701.4	0.038	-1,4146	8130	15518	5749	5079	91	3648
19	19989	69.3	41,13657	24,8869732	495.1	0.020	-1,6947	16930	23362	6913	10873	94	5741
20	5014	73.6	40,98177	24,9511587	352	0.037	-1,4326	4373	6817	2017	3173	27	1675
21	3616	76.7	41,2444	24,921751	633.3	0.006	-2,1995	2499	5574	1649	2594	22	1370
22	3006	74.0	40,96311	24,8060305	312.5	0.032	-1,4949	2423	4141	1225	1927	17	1017
23	103243	68.3	40,64012	22,9442478	19.3	1.865	0,2707	81274	158600	29281	39755	706	4274
24	15498	70.3	40,85635	22,9053191	9.8	0.816	-0,0881	12165	25424	4694	6373	113	685
25	6005	74.4	40,82655	23,4410355	783	0.029	-1,5320	5642	11451	2114	2870	51	309
26	14203	68.9	40,82715	22,7291064	311.1	0.077	-1,1127	11300	22357	4128	5604	100	602
27	16435	67.3	40,43065	22,9290361	133.4	0.090	-1,0460	14113	24515	4526	6145	109	661
28	19743	82.9	40,5485	23,0188789	382.1	0.079	-1,1051	14956	25948	4790	6504	115	699
29	34420	62.4	40,58284	22,9527298	6.4	0.625	-0,2041	27919	44636	8241	11188	199	1203
30	36421	64.2	40,67054	22,9040932	13.4	0.896	-0,0479	27893	49028	9162	12440	221	1337
31	11589	71.8	40,75096	23,0687491	1222.7	0.016	-1,7863	10157	20047	3701	5025	89	540
32	28301	66.6	40,65096	22,9478489	12.9	0.310	-0,5085	21928	41330	7630	10360	184	1114
33	32621	67.1	40,6778	22,9388488	23.8	0.714	-0,1461	25947	48404	8936	12133	215	1304
34	27334	61.0	40,60114	22,9895668	155.6	0.116	-0,9367	20696	34195	6313	8571	152	921
35	10162	69.8	40,7324	22,5979106	391.5	0.049	-1,3140	8533	16423	3032	4117	73	443
36	13357	65.1	40,72882	22,9178804	217.9	0.064	-1,1921	10200	18888	3450	4684	83	504
37	21275	68.0	40,51965	22,2053102	796.5	0.029	-1,5395	17090	26584	12308	4934	102	5353
38	12054	71.0	40,82863	22,443452	478.8	0.027	-1,5662	10159	16906	7689	3082	64	3344
39	9774	69.9	40,82991	22,0686048	424.9	0.031	-1,5143	8154	12981	6010	2409	50	2614
40	15439	70.3	40,99433	22,8743837	1599.6	0.009	-2,0579	13836	17307	7352	2637	106	7403
41	7547	73.5	40,99721	22,5692594	919.3	0.023	-1,6412	7018	9497	4034	1447	58	4062
42	9191	68.1	40,80206	22,0434583	611.2	0.013	-1,8831	7339	8645	8052	2395	47	2565
43	8505	69.1	40,97347	22,0609874	985.8	0.007	-2,1487	6664	8268	5788	2290	45	2453
44	19917	68.4	40,76333	22,5262732	669.2	0.034	-1,4638	15577	18939	13259	5246	102	5618
45	6190	69.3	40,76771	22,1539007	239.5	0.025	-1,6012	4959	6057	4241	1678	33	1797
46	27504	68.0	40,27051	22,5117929	681.9	0.023	-1,6296	21724	28392	12828	7538	133	7678
47	7558	70.6	40,17119	22,4630871	495.3	0.028	-1,5487	6355	8489	3776	2254	40	2295

Πίνακας 4.2: Βάση δεδομένων (2)

Ο Πίνακας 4.2 αποτελεί οριζόντια συνέχεια του Πίνακα 4.1 και ολοκληρώνει την βάση δεδομένων.

Στις πρώτες δύο στήλες του Πίνακα 4.2, παρουσιάζονται αντίστοιχα ο αριθμός και το ποσοστό των απασχολούμενων σε εργασία, λαμβάνοντας ξανά υπόψη την απογραφή του 2011. Οι στήλες με τις μεταβλητές “Lat” και “Long” αποτελούν αντίστοιχα το γεωγραφικό πλάτος (latitude) και το γεωγραφικό μήκος (longitude) του κέντρου κάθε Δήμου, όπως αυτά προέκυψαν από τους ψηφιακούς χάρτες της Google. Στην στήλη Y καταγράφεται η έκταση του Δήμου σε τετραγωνικά χιλιόμετρα.

Στην στήλη Z περιέχεται η μεταβλητή “FtpAr” (fatalities per area) που προκύπτει ως το πηλίκο των θανάτων σε οδικά ατυχήματα προς την συνολική έκταση του Δήμου. Στην επόμενη στήλη, έχει υπολογιστεί ο δεκαδικός λογάριθμος των θανάτων ανά τετραγωνικό χιλιόμετρο και εκφράζεται μέσω της μεταβλητής “logFtpAr”.

Στην επόμενη στήλη του πίνακα, η μεταβλητή “hearth_cars_own” αντιστοιχεί στον αριθμό νοικοκυριών με ιδιοκτησία ΙΧ, ενώ στις στήλες AC έως και AF καταγράφονται οι στόλοι οχημάτων για αυτοκίνητα ΙΧ, φορτηγά, μοτοσικλέτες και λεωφορεία αντίστοιχα, σύμφωνα με δεδομένα του 2020.

Τέλος, στην στήλη AG του πίνακα βρίσκεται το Ακαθάριστο Εθνικό Προϊόν (GDP) κάθε δήμου για το έτος 2019.

4.4 Ομαδοποίηση δεδομένων (clustering)

Προκειμένου η στατιστική επεξεργασία των δεδομένων να είναι πιο εύκολη, καθώς επίσης και να προκύψουν πιο πρακτικά και κατανοητά συμπεράσματα, κρίθηκε αναγκαία η ομαδοποίηση των στοιχείων της βάσης δεδομένων. Ο αλγόριθμος που χρησιμοποιήθηκε στην παραπάνω διαδικασία είναι αυτός της **ιεραρχικής ομαδοποίησης** (hierarchical clustering) που παρουσιάστηκε αναλυτικά στο Κεφάλαιο 3.

Αρχικά, χρειάζεται η βάση δεδομένων της Διπλωματικής Εργασίας να μεταφερθεί από το Microsoft Excel σε περιβάλλον προγραμματισμού R, όπως φαίνεται παρακάτω.

```
#START
rm(list=ls())
Sys.sleep(0.1)
options(scipen = 7)

library(readxl)
#import dataset
Municipal <- read_excel("Municipalities_data_final.xlsx")
summary(Municipal)
str(Municipal)
```

Εικόνα 4.1: Μετατροπή βάσης δεδομένων σε περιβάλλον λογισμικού R

Στον παραπάνω κώδικα, αρχικά χρησιμοποιείται η βιβλιοθήκη “library(readxl)” που επιτρέπει στο πρόγραμμα να διαβάσει αρχεία excel. Η βάση δεδομένων παραμένει ως έχει, όμως είναι πλέον στο περιβάλλον της R και μπορεί να υποστεί προχωρημένη στατιστική επεξεργασία.

Ένα τυπικό απόσπασμα της βάσης δεδομένων όπως αυτή φαίνεται στο περιβάλλον της R είναι το ακόλουθο.

Municipality	Fatal2016	Fatal2017	Fatal2018	Fatal2019	Fatal2020	Tot_Fatal	Pop2011	FatalpPop	logFP	Fem_Pop2011	Fem_perc	Mal_Pop2011	Mal
1 ΔΗΜΟΣ ΚΟΜΟΤΗΝΗΣ	5	5	2	7	2	21	65919	3.1361222	0.486669652	34565	48.34800	32354	
2 ΔΗΜΟΣ ΑΡΡΑΝΩΝ	1	0	1	0	0	2	16577	1.2064909	0.081524058	8394	49.36358	8183	
3 ΔΗΜΟΣ ΙΑΣΜΟΥ	2	0	0	3	1	6	13810	4.3446778	0.637957572	7079	48.74004	6731	
4 ΔΗΜΟΣ ΜΑΡΟΝΙΑΣ - ΣΑΠΩΝ	3	0	0	1	0	4	14733	2.7149936	0.433768002	7446	49.46040	7287	
5 ΔΗΜΟΣ ΔΡΑΜΑΣ	3	3	3	3	2	14	58944	2.3751357	0.375688431	30640	48.01846	28304	
6 ΔΗΜΟΣ ΔΟΣΣΑΤΟΥ	0	1	2	0	1	4	14516	2.7555800	0.440213032	7531	48.11932	6985	
7 ΔΗΜΟΣ ΚΑΤΩ ΝΕΥΡΟΚΟΠΙΟΥ	1	2	0	1	0	4	7660	5.0890565	0.706637445	3897	50.41985	3963	
8 ΔΗΜΟΣ ΠΑΡΑΝΕΣΤΙΟΥ	0	0	0	0	0	0	3901	0.0000000	0.000000000	1930	50.52551	1971	
9 ΔΗΜΟΣ ΠΡΟΣΟΤΣΑΝΗΣ	1	0	0	0	1	2	13066	1.5306903	0.184887342	6747	48.36216	6319	
10 ΔΗΜΟΣ ΑΓΓΕΛΑΡΟΥΠΟΛΗΣ	5	6	5	1	6	23	72929	3.1524555	0.490540963	37104	49.03439	35775	
11 ΔΗΜΟΣ ΔΙΔΥΜΟΤΕΙΧΟΥ	2	0	2	0	1	5	19493	2.5650233	0.409091322	9353	52.01867	10140	
12 ΔΗΜΟΣ ΟΡΕΣΤΙΑΔΑΣ	5	3	1	2	1	12	37695	3.1834461	0.502897498	18141	51.87425	19554	
13 ΔΗΜΟΣ ΣΑΜΟΘΡΑΚΗΣ	0	0	1	0	1	2	2839	6.9954530	0.844815940	1243	56.52326	1616	
14 ΔΗΜΟΣ ΣΟΥΦΛΙΟΥ	1	2	1	0	2	6	14941	4.0157955	0.603771585	7521	51.00060	7620	
15 ΔΗΜΟΣ ΘΑΣΟΥ	2	1	0	1	1	5	13770	3.6310821	0.560036064	6985	49.27378	6785	
16 ΔΗΜΟΣ ΚΑΒΑΛΑΣ	4	1	0	4	1	10	70501	1.4104196	0.151004723	36499	48.22910	34002	
17 ΔΗΜΟΣ ΝΕΣΤΟΥ	1	6	2	1	5	15	22331	6.7171197	0.827183087	11286	49.46039	11049	
18 ΔΗΜΟΣ ΠΑΓΤΑΙΟΥ	4	5	14	3	1	27	32085	8.4151473	0.925061721	16278	49.26601	15807	
19 ΔΗΜΟΣ ΞΑΝΘΗΣ	1	2	2	4	1	10	65133	1.5353200	0.106190918	32804	49.63536	32329	
20 ΔΗΜΟΣ ΑΒΔΗΡΩΝ	3	4	2	3	1	13	19005	6.8403052	0.835075478	9658	49.18179	9347	
21 ΔΗΜΟΣ ΜΥΚΗΣ	1	2	1	0	0	4	15540	2.5740026	0.410608877	7853	49.46589	7687	
22 ΔΗΜΟΣ ΤΟΠΕΡΟΥ	4	0	1	5	0	10	11544	8.6625067	0.937643682	5705	50.50039	5839	
23 ΔΗΜΟΣ ΘΕΣΣΑΛΟΝΙΚΗΣ	9	11	8	6	2	36	325182	1.1070723	0.044176003	176712	45.65751	148470	
24 ΔΗΜΟΣ ΑΜΠΕΛΟΚΗΠΩΝ - ΜΕΝΕΜΕΝΗΣ	1	1	4	2	0	8	52127	1.5347133	0.186027256	27406	47.42456	24721	
25 ΔΗΜΟΣ ΒΟΛΒΗΣ	7	7	4	3	2	23	23476	9.7964951	0.991066738	11736	50.01276	11742	
26 ΔΗΜΟΣ ΔΕΛΤΑ	7	5	2	7	3	24	45839	5.2357163	0.718976107	23010	49.80257	22829	

Εικόνα 4.2: Η βάση δεδομένων στο λογισμικό R

Όπως έχει ήδη αναφερθεί στο Θεωρητικό Υπόβαθρο της Διπλωματικής Εργασίας, ο αλγόριθμος της ιεραρχικής ομαδοποίησης εκτιμά την ευκλείδεια απόσταση μεταξύ δύο συστάδων δεδομένων (clusters) και ενοποιεί τις πλησιέστερες σε μεγαλύτερες. Προκειμένου η διαδικασία της ομαδοποίησης να έχει το καλύτερο δυνατό αποτέλεσμα, πραγματοποιήθηκαν δοκιμές με τις μεταβλητές έτσι ώστε να επιλεγούν αυτές που θα δώσουν το πιο αποδεκτό δένδρογραμμα. Τελικά, οι τρεις μεταβλητές που χρησιμοποιήθηκαν στην διαδικασία είναι ο αριθμός θανάτων ανά πληθυσμό, το ΑΕΠ και η ιδιοκτησία ΙΧ ανά νοικοκυριό.

Ο κώδικας της διαδικασίας ομαδοποίησης στο λογισμικό R φαίνεται στην παρακάτω εικόνα.

```
#HIERARCHICAL CLUSTERING - Combination 6

#creating subset
keep_vars = c("FatalpPop", 'GDP2019', 'hearth_cars_own')
dataset = subset(Municipal, select = c(keep_vars))

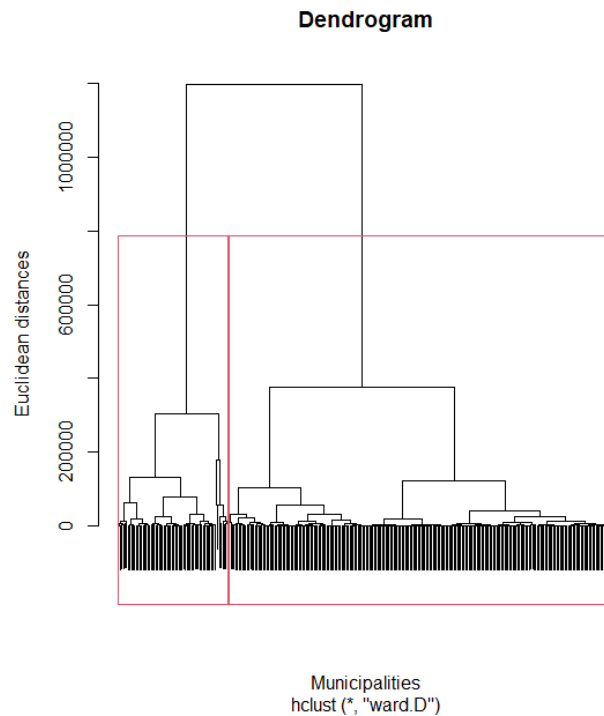
#dendrogram
dendrogram = hclust(d = dist(dataset, method = 'euclidean'), method = 'ward.D')
plot(dendrogram,
     main = paste('Dendrogram'),
     xlab = 'Municipalities',
     ylab = 'Euclidean distances', labels = F)

rect.hclust(dendrogram, k = 2)

hc = hclust(d = dist(dataset, method = 'euclidean'), method = 'ward.D')
y_hc = cutree(hc, 2)
y_hc <- as.factor(y_hc)
summary(y_hc)
Municipal$y_hc <- y_hc
```

Εικόνα 4.3: Ιεραρχική ομαδοποίηση

Στον παραπάνω κώδικα αρχικά δημιουργείται ένα υποσύνολο της βάσης δεδομένων, όπου οι μεταβλητές που χρησιμοποιούνται είναι οι τρεις που αναφέρθηκαν προηγουμένως. Στη συνέχεια, ο αλγόριθμος της ομαδοποίησης παράγει το δένδρογραμμα με βάση τις ευκλείδειες αποστάσεις. Τέλος, εξετάζοντας τη μορφή του δένδρογραμματος κρίνεται ότι πρέπει να διαχωριστεί σε δύο συστάδες δεδομένων, όπως φαίνεται παρακάτω.



Εικόνα 4.4: Τελικό διαχωρισμένο δένδρογραμμα ομαδοποίησης

Από την ομαδοποίηση των δεδομένων προέκυψαν συνεπώς δύο συστάδες δήμων. Η πρώτη συστάδα περιέχει τους λιγότερους σε αριθμό δήμους, συγκεκριμένα εξήντα οχτώ. Η δεύτερη περιέχει τους υπόλοιπους διακόσιους τριάντα επτά δήμους. Στο εξής, για λόγους ευκολίας και συμβατότητας με τα αρχεία επεξεργασίας της Διπλωματικής Εργασίας, θα αναφερόμαστε σε αυτές ως Cluster-1 και Cluster-2 αντίστοιχα.

Σε πρώτο στάδιο πραγματοποιήθηκε μια πρωταρχική στατιστική επεξεργασία με το λογισμικό R για κάθε συστάδα, προκειμένου να παρατηρηθούν ορισμένα βασικά χαρακτηριστικά τους. Συγκεκριμένα, υπολογίστηκε ο μέσος όρος της κάθε μίας από τις τρεις μεταβλητές. Παρακάτω φαίνεται το αποτέλεσμα της R, ανά μεταβλητή και ανά συστάδα.

```
> #aggregate statistics
> aggregate(Municipal$FatalPop, list(Municipal$y_hc), FUN = mean)
  Group.1      x
1      1  2.469142
2      2  4.680783
> aggregate(Municipal$GDP2019, list(Municipal$y_hc), FUN = mean)
  Group.1      x
1      1 3854.110
2      2 1829.068
> aggregate(Municipal$hearth_cars_own, list(Municipal$y_hc), FUN = mean)
  Group.1      x
1      1 25495.338
2      2  4792.363
```

Εικόνα 4.5: Μέσος όρος μεταβλητών ανά συστάδα δεδομένων

Πολύ εύκολα παρατηρούμε ότι η συστάδα Cluster-1 περιέχει τους μεγαλύτερους πληθυσμιακά δήμους. Περιλαμβάνει πολλούς από τους δήμους της Μητροπολιτικής Περιοχής της Αθήνας, καθώς και δήμους άλλων πόλεων όπως Θεσσαλονίκης, Καβάλας, Βόλου, Λαμιέων, Κορινθίων, Πατρέων και Καλαμάτας. Επιπλέον, στη συστάδα αυτή κατατάσσονται και οι δήμοι των μεγαλύτερων νησιών της χώρας όπως Ηρακλείου (Κρήτης), Ρεθύμνης, Χανίων, Κερκύρας, Ρόδου, Χίου και Λέσβου.

Οι υπόλοιποι δήμοι ανήκουν στην συστάδα Cluster-2 και πρόκειται κυρίως για επαρχιακές περιοχές, καθώς και δήμους περιφερειακά των μεγάλων πόλεων – για παράδειγμα οι δήμοι Σαλαμίνας και Ωρωπού. Στην συστάδα αυτή παρατηρούνται και ορισμένοι αστικοί δήμοι που ωστόσο έχουν χαμηλότερο πληθυσμό, όπως Αλίμου, Φιλοθέης – Ψυχικού και Αγ. Βαρβάρας.

Είναι λοιπόν προφανές ότι η διαδικασία της ομαδοποίησης ήταν επιτυχής.

4.5 Περιγραφική Στατιστική

Στο υποκεφάλαιο αυτό γίνεται μια βασική στατιστική επεξεργασία των πλέον ομαδοποιημένων δεδομένων, από την οποία μπορούν να προκύψουν ορισμένα πρωταρχικά συμπεράσματα και να διαμορφωθεί μια γενικότερη εικόνα σχετικά με τα οδικά ατυχήματα στην Ελλάδα.

4.5.1 Στατιστικοί υπολογισμοί & γραφήματα

Αρχικά, για κάθε συστάδα χωριστά και ανά μεταβλητή, υπολογίζονται τέσσερα θεμελιώδη μεγέθη της στατιστικής:

- Ελάχιστο (minimum): η χαμηλότερη τιμή της μεταβλητής στο σύνολο της συστάδας.
- Μέσος όρος (average): το άθροισμα όλων των τιμών της μεταβλητής διαιρεμένο με τον αριθμό των δήμων στην συστάδα.
- Διάμεσος (median): η ενδιάμεση τιμή του συνόλου τιμών της μεταβλητής.
- Μέγιστο (maximum): η υψηλότερη τιμή της μεταβλητής στο σύνολο της συστάδας.

Πρέπει να σημειωθεί ότι δεν χρησιμοποιούνται όλες οι μεταβλητές της βάσης δεδομένων σε αυτό το στάδιο της ανάλυσης. Περιλαμβάνονται μόνο αυτές για τις οποίες τα παραπάνω μεγέθη έχουν πρακτικό νόημα και μπορούν να δώσουν ουσιαστικά συμπεράσματα. Συγκεκριμένα, η μεταβλητή “FtpAr” (αριθμός θανάτων ανά τετραγωνικό χιλιόμετρο) δεν περιλαμβάνεται καθώς έχει γενικά πολύ χαμηλές τιμές και δεν θα προσέφερε ουσιαστικά στις γραφικές απεικονίσεις.

Οι μεταβλητές που χρησιμοποιήθηκαν στην ανάλυση αυτή είναι συγκεκριμένα: αριθμός θανάτων ανά πληθυσμό, γυναικείος πληθυσμός, ανδρικός πληθυσμός, τελειόφοιτοι πανεπιστημίου, απόφοιτοι λυκείου, εργαζόμενοι, έκταση δήμου, αριθμός νοικοκυριών με

ιδιοκτησία ΙΧ, στόλος αυτοκινήτων ΙΧ, στόλος φορτηγών, στόλος μοτοσικλετών, πληθυσμός ηλικιωμένων, Ακαθάριστο Εθνικό Προϊόν, στόλος λεωφορείων.

Με αυτή την σειρά παρατίθενται οι μεταβλητές στους πίνακες και στα γραφήματα.

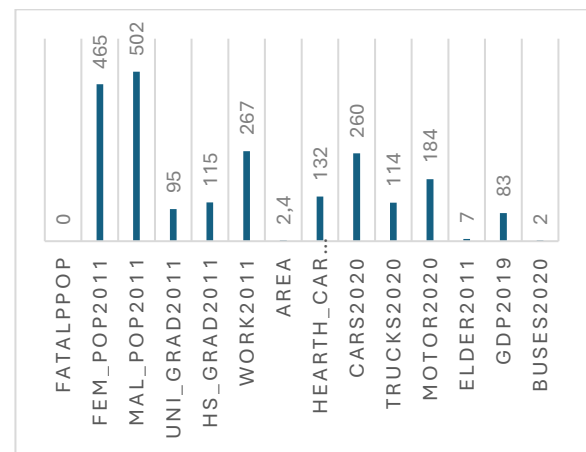
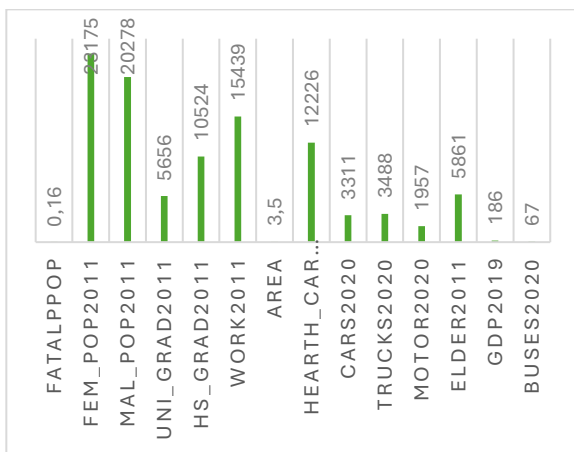
Cluster-1	Stats			
Variables	min	avg	med	max
FatalpPop	0,16	2,47	2,38	6,84
Fem_Pop2011	23175	48710,4	37884	348836
Mal_Pop2011	20278	45528,74	34644	315210
Uni_Grad2011	5656	18877,04	13311	161997
HS_Grad2011	10524	24275,74	17827	196705
Work2011	15439	33728,84	26436	260658
Area	3,5	346,04	73,45	1677,1
hearth_cars_own	12226	25495,34	20283	158873
cars2020	3311	54261,11	47842	536984
trucks2020	3488	10679,26	8248	51996
motor2020	1957	15345,19	11931	126297
elder2011	5861	13339,36	10933	87391
GDP2019	186	3854,11	1254	15510
buses2020	67	242,48	200	2168

Πίνακας 4.5: Στατιστικά Cluster-1

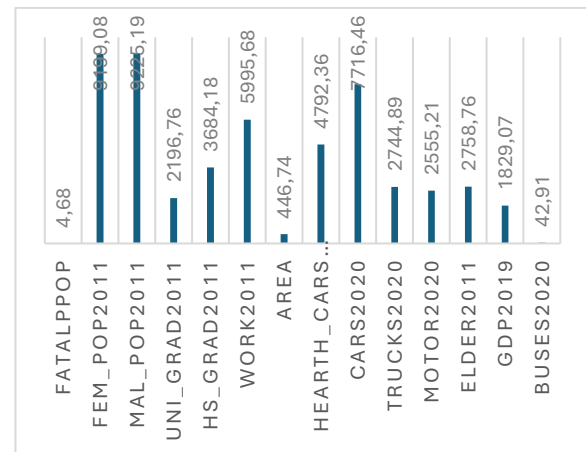
Cluster-2	Stats			
Variables	min	avg	med	max
FatalpPop	0	4,68	4,45	15,43
Fem_Pop2011	465	9199,08	8204	27406
Mal_Pop2011	502	9225,19	8183	24721
Uni_Grad2011	95	2196,76	1301	13628
HS_Grad2011	115	3684,18	2682	12939
Work2011	267	5995,68	4853	17722
Area	2,4	446,74	400,5	1859
hearth_cars_own	132	4792,36	3872	13347
cars2020	260	7716,46	4536	37924
trucks2020	114	2744,89	2341	9512
motor2020	184	2555,21	1738	10673
elder2011	7	2758,76	2625	9184
GDP2019	83	1829,07	1191	19250
buses2020	2	42,91	33	153

Πίνακας 4.6: Στατιστικά Cluster-2

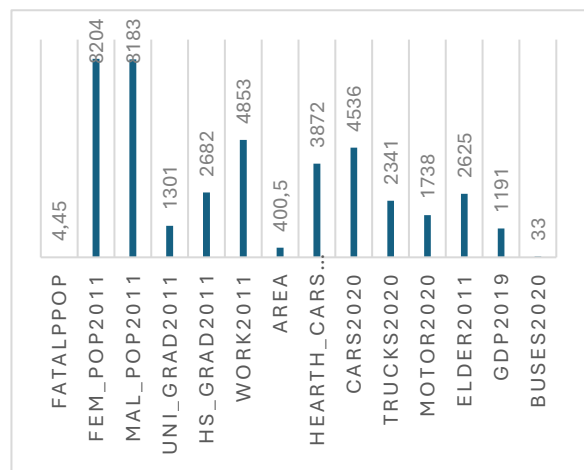
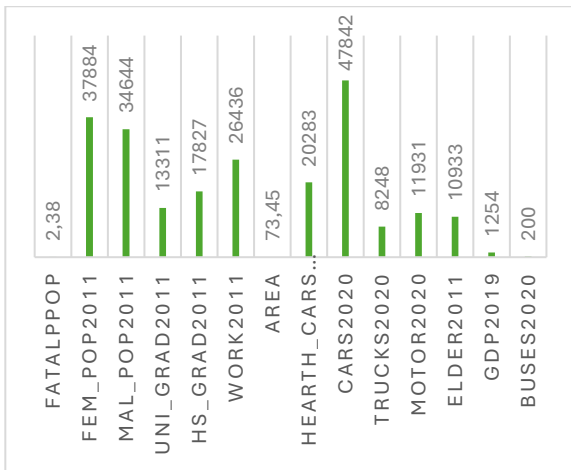
Ακολουθούν τα διαγράμματα σύγκρισης των στατιστικών μεγεθών για τις μεταβλητές. Σε όλες τις περιπτώσεις, με πράσινο φαίνεται η συστάδα Cluster-1, ενώ με μπλε η Cluster-2.



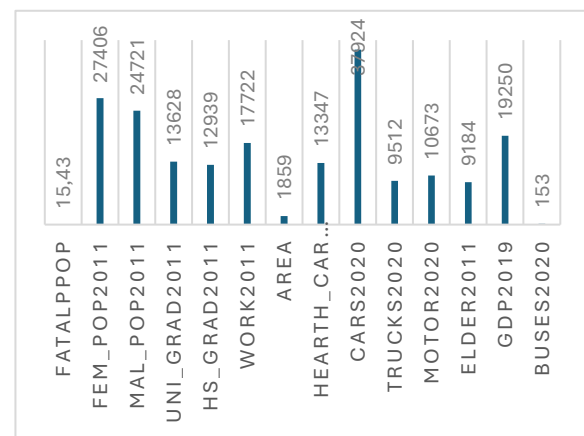
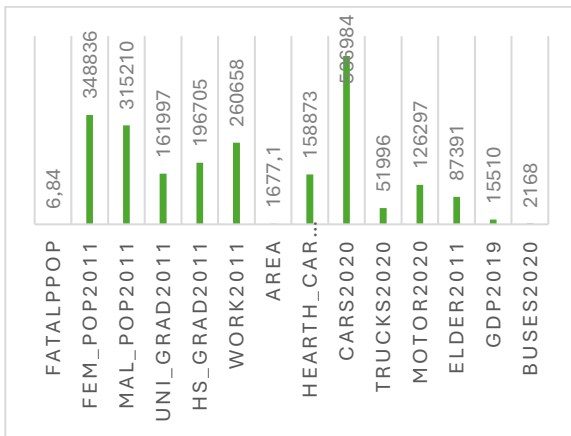
Διαγράμματα 4.1 – 4.2: Σύγκριση ελαχίστων (min)



Διαγράμματα 4.3 – 4.4: Σύγκριση μέσων όρων (avg)

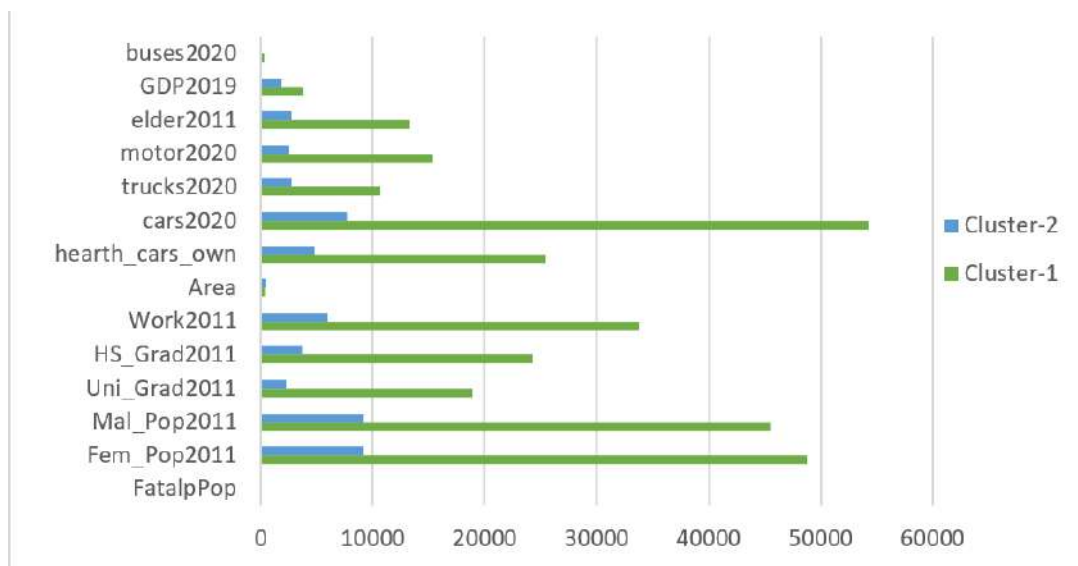


Διαγράμματα 4.5 – 4.6: Σύγκριση διαμέσων (med)

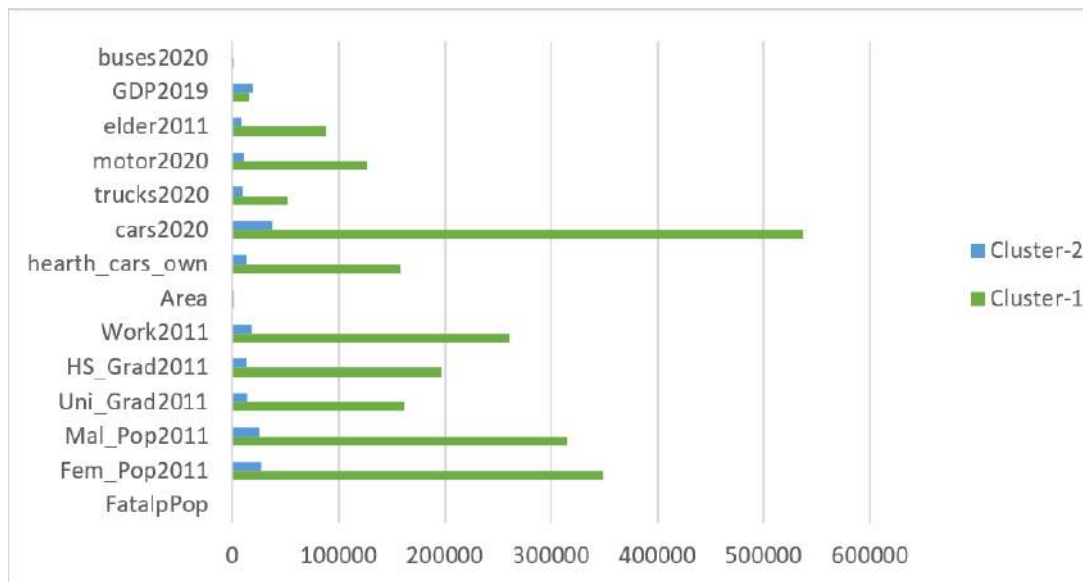


Διαγράμματα 4.7 – 4.8: Σύγκριση μεγίστων (max)

Στη συνέχεια αυτής της βασικής στατιστικής επεξεργασίας, παρατίθενται ορισμένα ακόμα χρήσιμα διαγράμματα.

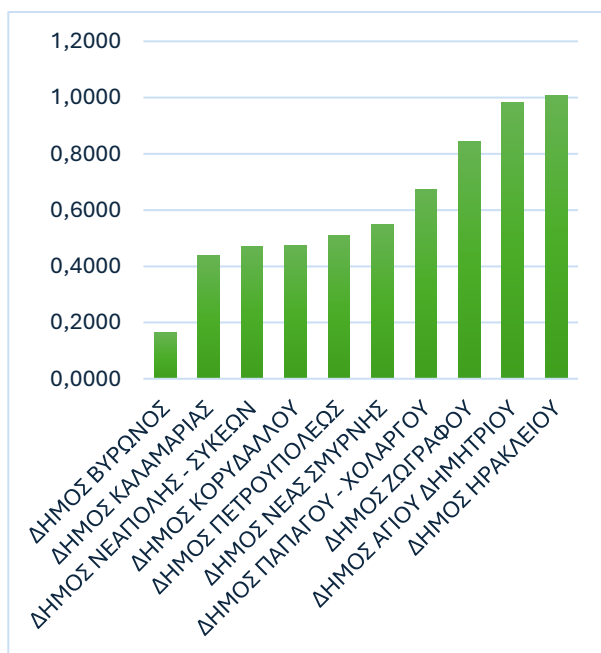


Διάγραμμα 4.9: Σύγκριση μέσων όρων μεταξύ των συστάδων

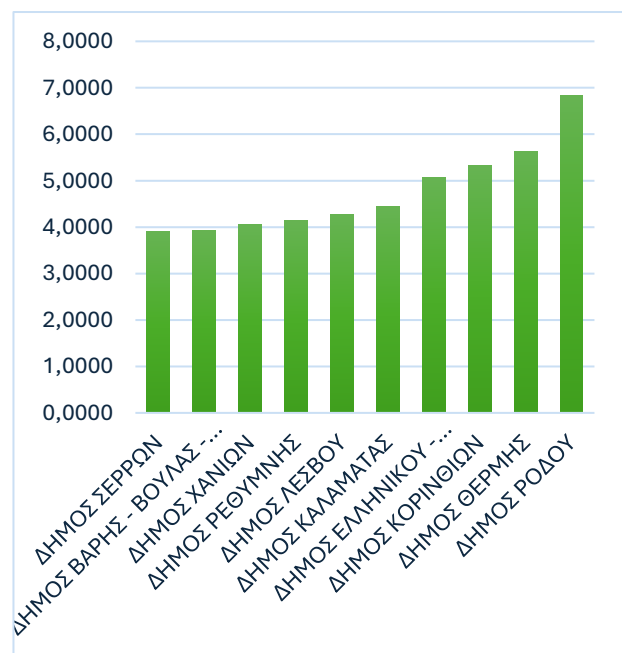


Διάγραμμα 4.10: Σύγκριση μεγίστων μεταξύ των συστάδων

Τέλος, οι δήμοι κάθε συστάδας κατατάχθηκαν με βάση τον αριθμό θανάτων σε οδικά ατυχήματα ανά πληθυσμό (μεταβλητή “FatalpPop”). Από την κατάταξη προκύπτουν τα γραφήματα σύγκρισης των δέκα “καλύτερων” και των δέκα “χειρότερων” δήμων, ως προς την παραπάνω μεταβλητή.

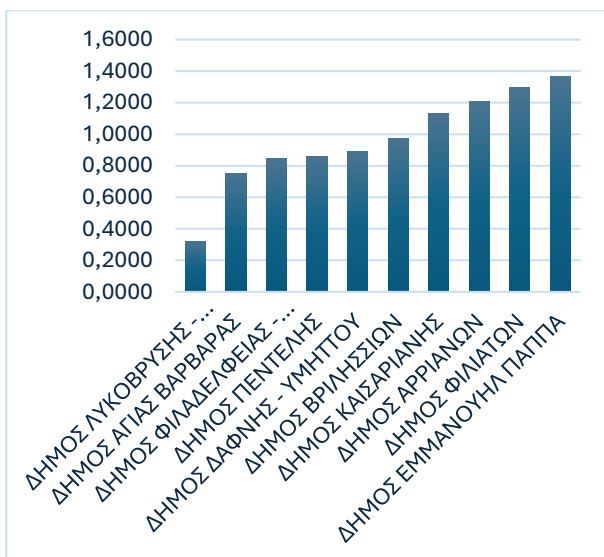


Διάγραμμα 4.11: Καλύτεροι δήμοι Cluster-1



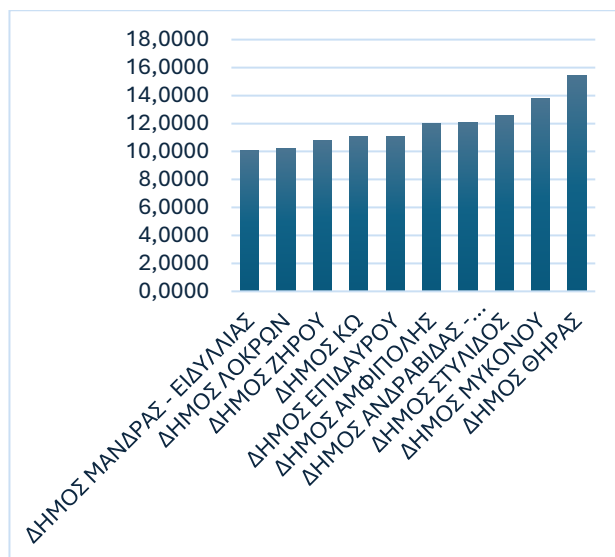
Διάγραμμα 4.12: Χειρότεροι δήμοι Cluster-1

Χρειάζεται να διευκρινιστεί ότι στο παραπάνω Διάγραμμα 4.11, ο δήμος Ηρακλείου αναφέρεται στο Ηράκλειο Αττικής και όχι στην πρωτεύουσα της Κρήτης.



Διάγραμμα 4.13: Καλύτεροι δήμοι

Cluster-2



Διάγραμμα 4.14: Χειρότεροι δήμοι

Cluster-2

Πρέπει να σημειωθεί ότι στην συστάδα Cluster-2 υπήρξαν και δήμοι με μηδενικό αριθμό θανάτων ανά πληθυσμό, οι οποίοι παρουσιάζονται παρακάτω. Στην πράξη λοιπόν, αυτοί είναι οι δήμοι της Cluster-2 με τις καλύτερες επιδόσεις ως προς τα οδικά ατυχήματα.

CLUSTER-2	FatalpPop
ΔΗΜΟΣ ΠΑΡΑΝΕΣΤΙΟΥ	0,0000
ΔΗΜΟΣ ΦΟΥΡΝΩΝ ΚΟΡΣΕΩΝ	0,0000
ΔΗΜΟΣ ΑΝΑΦΗΣ	0,0000
ΔΗΜΟΣ ΣΙΚΙΝΟΥ	0,0000
ΔΗΜΟΣ ΦΟΛΕΓΑΝΔΡΟΥ	0,0000
ΔΗΜΟΣ ΛΕΙΨΩΝ	0,0000
ΔΗΜΟΣ ΚΑΣΟΥ	0,0000
ΔΗΜΟΣ ΚΕΑΣ	0,0000
ΔΗΜΟΣ ΚΥΘΝΟΥ	0,0000
ΔΗΜΟΣ ΝΙΣΥΡΟΥ	0,0000
ΔΗΜΟΣ ΑΜΟΡΓΟΥ	0,0000
ΔΗΜΟΣ ΤΗΛΟΥ	0,0000
ΔΗΜΟΣ ΟΡΟΠΕΔΙΟΥ ΛΑΣΙΘΙΟΥ	0,0000

Πίνακας 4.7: Δήμοι μηδενικών θανάτων ανά πληθυσμό Cluster-2

4.5.2 Παρατηρήσεις

Ακολουθούν ορισμένες χρήσιμες παρατηρήσεις που προέκυψαν από την περιγραφική στατιστική επεξεργασία.

Αρχικά, φαίνεται άμεσα ότι ο πληθυσμός στην συστάδα Cluster-1 είναι υψηλότερος από αυτόν στην Cluster-2, κάτι που είναι λογικό εφόσον η πρώτη περιλαμβάνει τους πιο αστικοποιημένους δήμους της χώρας. Επίσης, παρατηρούμε ότι στην πρώτη συστάδα ο αριθμός απόφοιτων λυκείου και πανεπιστημιακών σχολών είναι υψηλότερος. Το ίδιο ισχύει

και για τους εργαζόμενους, την ιδιοκτησία ΙΧ και τους στόλους οχημάτων. Επιβεβαιώνεται λοιπόν ότι η ομαδοποίηση των δεδομένων έδωσε λογικά αποτελέσματα.

Επιπλέον, ο αριθμός θανάτων λόγω οδικών ατυχημάτων ανά πληθυσμό βρίσκεται κατά μέσο όρο μεγαλύτερος στην συστάδα Cluster-2, ενώ το μέγιστό της είναι σημαντικά υψηλότερο σε σχέση με την Cluster-1. Μια λογική εξήγηση είναι το γεγονός ότι ο πληθυσμός στις περιφερειακές και επαρχιακές περιοχές είναι πολύ χαμηλότερος από τον αντίστοιχο των πόλεων, συνεπώς το πηλίκο (αριθμός θανάτων / πληθυσμός) αυξάνεται.

Κεφάλαιο 5: Εφαρμογή Μεθοδολογίας

5.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζεται η αναλυτική **περιγραφή της μεθοδολογίας** και των αποτελεσμάτων που προέκυψαν από αυτήν. Στη συνέχεια της συλλογής και πρωτογενούς επεξεργασίας των στοιχείων, που αναλύθηκαν στο Κεφάλαιο 4, ακολούθησε η **στατιστική επεξεργασία** με μοντέλα παλινδρόμησης στο λογισμικό R. Αρχικά, περιγράφεται η διαδικασία ανάπτυξης των μοντέλων και οι στατιστικοί έλεγχοι που καθορίζουν την αποδοχή ή απόρριψή τους. Στη συνέχεια, εφαρμόζεται η ειδική μέθοδος χωρικής παλινδρόμησης GWR.

Τέλος, γίνεται **παρουσίαση των αποτελεσμάτων** που προέκυψαν από τα μοντέλα, τα οποία έπειτα επεξηγούνται και ερμηνεύονται με βάση τη λογική. Συγκεκριμένα στην περίπτωση της Γεωγραφικά Σταθμισμένης Παλινδρόμησης, γίνεται σύγκριση με τα παραδοσιακά μοντέλα, και εξετάζεται το κατά πόσον η χωρική ανάλυση βελτιώνει τα αποτελέσματα.

Κατά την εφαρμογή της μεθοδολογίας αναπτύχθηκαν αρκετά μοντέλα παλινδρόμησης πραγματοποιώντας **δοκιμές με διάφορους συνδυασμούς μεταβλητών**, οι οποίες σε συνδυασμό με τους απαραίτητους στατιστικούς ελέγχους οδήγησαν στην επιλογή των καταλληλότερων μοντέλων. Χρειάζεται να σημειωθεί ότι, ανεξάρτητα από τις ονομασίες των μαθηματικών μοντέλων κατά την στατιστική επεξεργασία, στο εξής θα αναφέρονται απλά με τις περιγραφές τους.

5.2 Συνολική επεξεργασία

5.2.1 Συσχέτιση μεταβλητών

Όπως έχει ήδη αναφερθεί στο Κεφάλαιο 4, τα δεδομένα της Διπλωματικής Εργασίας κατατάχθηκαν σε μία ενιαία βάση δεδομένων στο πρόγραμμα Excel, η οποία ύστερα εντάχθηκε στο λογισμικό R, όπου έγινε η στατιστική επεξεργασία.

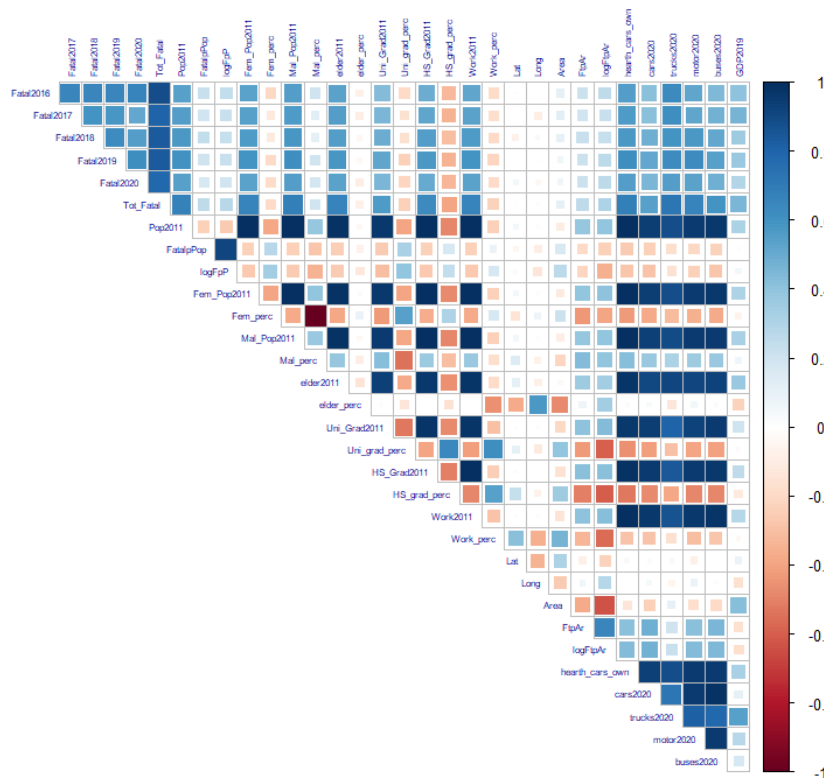
Ένα ουσιαστικό βήμα είναι η διερεύνηση της **συσχέτισης** των μεταβλητών. Πρακτικά, αυτό που επιδιώκεται είναι να υπάρχει η μέγιστη δυνατή συσχέτιση μεταξύ εξαρτημένης και ανεξάρτητων μεταβλητών, ενώ χαμηλή συσχέτιση μεταξύ των ανεξάρτητων μεταβλητών. Όπως φαίνεται στον παρακάτω κώδικα, το λογισμικό R χρησιμοποιεί μια βιβλιοθήκη που του επιτρέπει να δημιουργεί γραφήματα συσχέτισης, και με την βάση δεδομένων ως αρχικό στοιχείο εισόδου, απομονώνονται οι στήλες που περιέχουν μόνο αριθμούς. Με τον τρόπο αυτό, αγνοείται η στήλη με τις ονομασίες των δήμων και προκύπτει ένας πίνακας αριθμών. Στη συνέχεια, με κριτήριο τον συντελεστή **Pearson** κατασκευάζεται ο πίνακας συσχέτισης των μεταβλητών.


```
#correlation estimation for numeric data - Grab only numeric columns
library(corrplot)
Municipal_numeric<-Municipal[,sapply(Municipal, is.numeric)]
CorMatrix <- cor(Municipal_numeric,method = c("pearson"))
CorMatrix

#visualize correlation matrix
corrplot::corrplot(CorMatrix, method = 'square', type = 'upper', diag = FALSE, tl.cex = 0.5, tl.col = "darkblue")
```

Εικόνα 5.1: Δημιουργία πίνακα συσχέτισης μεταβλητών

Τέλος, το λογισμικό R κατασκευάζει το **γράφημα του πίνακα συσχέτισης** όπως φαίνεται παρακάτω. Με σκούρο μπλε παρουσιάζεται η πλήρης θετική συσχέτιση, ενώ με σκούρο κόκκινο η πλήρης αρνητική. Παρατηρείται ότι οι δύο άξονες του γραφήματος δεν ξεκινούν και τελειώνουν στις ίδιες μεταβλητές, καθώς δεν θα είχε νόημα η διερεύνηση συσχέτισης μιας μεταβλητής με τον εαυτό της.



Εικόνα 5.2: Πίνακας συσχέτισης μεταβλητών

Εκ πρώτης όψεως, φαίνεται ότι υπάρχει υψηλή θετική συσχέτιση μεταξύ των μεταβλητών που αναπαριστούν πληθυσμιακά δεδομένα, όπως πληθυσμός γυναικών και ανδρών, πληθυσμός ηλικιωμένων, απόφοιτων και εργαζόμενων. Κατά συνέπεια, αυτές οι μεταβλητές δεν θα ήταν επιθυμητό να χρησιμοποιηθούν μαζί σε ένα μοντέλο παλινδρόμησης. Παρατηρείται όμως ότι δεν ισχύει το ίδιο και για τα αντίστοιχα ποσοστά. Τέλος, φαίνεται ότι μια πολύ ισορροπημένη μεταβλητή είναι το Ακαθάριστο Εθνικό Προϊόν (GDP), και συνεπώς θα είναι χρήσιμη στην στατιστική ανάλυση.

5.2.2 Ανάπτυξη μαθηματικού προτύπου για το σύνολο των Δήμων

Στην ενότητα αυτή, παρουσιάζεται το τελικό μαθηματικό μοντέλο που αναπτύχθηκε για το **σύνολο της βάσης δεδομένων**, πριν από την διαδικασία της ομαδοποίησης. Μετά από δοκιμές, προέκυψε το μοντέλο **λογαριθμοκανονικής παλινδρόμησης** με εξαρτημένη μεταβλητή τον δεκαδικό λογάριθμο των θανάτων ανά τετραγωνικό χιλιόμετρο (“logFtpAr”), και ανεξάρτητες τις παρακάτω μεταβλητές:

- “Uni_grad_perc”: ποσοστό (%) των απόφοιτων πανεπιστημιακών σχολών
- “Fem_perc”: ποσοστό (%) του γυναικείου πληθυσμού σε σχέση με τον συνολικό
- “Work_perc”: ποσοστό (%) των απασχολούμενων σε εργασία
- “hearth_cars_own”: αριθμός νοικοκυριών με ιδιοκτησία τουλάχιστον ενός ΙΧ
- “GDP2019”: Ακαθάριστο Εθνικό Προϊόν 2019

Ο **λογάριθμος** προτιμήθηκε ως εξαρτημένη μεταβλητή καθώς μετά από δοκιμές φάνηκε ότι παράγει τα καλύτερα αποτελέσματα σε σχέση με την κανονική μεταβλητή (θάνατοι ανά τετραγωνικό χιλιόμετρο). Είναι προφανές ότι πραγματοποιήθηκαν αρκετές δοκιμές προκειμένου να προκύψει ο παραπάνω συνδυασμός μεταβλητών που οδηγεί σε ένα ορθό και λογικά αποδεκτό μαθηματικό μοντέλο. Λόγω του μεγάλου αριθμού των δοκιμών, δεν θα ήταν δυνατό να παρουσιαστούν αυτές στο πλαίσιο της Διπλωματικής Εργασίας, και έτσι εμφανίζεται μόνο το τελικό μοντέλο.

Στον παρακάτω κώδικα φαίνεται η προσαρμογή του προτύπου στο λογισμικό R, ενώ ακολουθούν τα αποτελέσματα επεξεργασμένα στο Microsoft Excel.

```

28 library(car)
29
30 #Linear regression logarithmic (Area)
31 model2 <- lm(logFtpAr ~ uni_grad_perc + Fem_perc + work_perc + hearth_cars_own + GDP2019 , data = Municipal)
32 summary(model2)
33 vif(model2)
34 #final total analysis model
35

```

Εικόνα 5.3: Προσαρμογή στο λογισμικό R – Συνολικό πρότυπο

Coefficients				
	Estimate	Standard Error	p - value	t - value
(Intercept)	7.428	1.170	< 0.001	6.348
Uni_grad_perc	-0.019	0.007	0.005	-2.806
Fem_perc	-0.065	0.024	0.007	-2.709
Work_perc	-0.057	0.008	< 0.001	-6.903
hearth_cars_own	$2 \cdot 10^{-5}$	0	< 0.001	6.124
GDP2019	$(-8) \cdot 10^{-5}$	0	< 0.001	-6.955

Πίνακας 5.1: Αποτελέσματα παλινδρόμησης – Συνολικό πρότυπο

Residuals				
Min	1Q	Median	3Q	Max
-1.642	-0.326	-0.071	0.241	2.221

Πίνακας 5.2: Υπολείμματα – Συνολικό πρότυπο

Model Summary	
Residual std. error	0.564 on 299 degrees of freedom
Multiple R-squared	0.537
Adjusted R-squared	0.529
p - value	< 0.001

Πίνακας 5.3: Περίληψη αποτελεσμάτων – Συνολικό πρότυπο

Variance Inflation Factor (VIF)				
Uni_grad_perc	Fem_perc	Work_perc	hearth_cars_own	GDP2019
2.475	1.598	1.707	1.517	1.139

Πίνακας 5.4: Δείκτης VIF – Συνολικό πρότυπο

Συνεπώς, η μαθηματική σχέση που διέπει αυτό το πρότυπο είναι η ακόλουθη:

$$\log F_{tpAr} = 7.428 - 0.019*(Uni_grad_perc) - 0.065*(Fem_perc) - 0.057*(Work_perc) + 2*10^{(-5)}*(hearth_cars_own) - 8*10^{(-5)}*(GDP2019)$$

5.2.3 Ποιότητα προτύπου & σχολιασμός

Σύμφωνα με τους στατιστικούς ελέγχους που αναφέρονται εκτενώς στο Κεφάλαιο 3, παρατηρούνται τα παρακάτω:

- Ο **προσαρμοσμένος συντελεστής R²** ισούται με 0.529 και θεωρείται επαρκής.
- Τα **υπολείμματα** της παλινδρόμησης είναι κοντά στο 0 και συνεπώς θεωρούνται αμελητέα.
- Το **σφάλμα** της παλινδρόμησης για κάθε μεταβλητή είναι πολύ χαμηλό.
- Οι τιμές **p – values** είναι πολύ χαμηλές για όλες τις μεταβλητές, και σαφώς μικρότερες από 0.05.
- Για κάθε μεταβλητή, ο **δείκτης VIF** είναι σαφώς μικρότερος του 5 και κοντά στο 1, επομένως δεν τίθεται ζήτημα πολυσυγγραμμικότητας στο μοντέλο.

Για να είναι το πρότυπο πλήρως αποδεκτό, θα πρέπει τα **πρόσημα** των μεταβλητών που προκύπτουν να εξηγούνται λογικά. Θετικό πρόσημο δηλώνει θετική συσχέτιση της εξαρτημένης μεταβλητής με την εκάστοτε ανεξάρτητη, ενώ αρνητικό πρόσημο δηλώνει αρνητική συσχέτιση.

Το **ποσοστό των απόφοιτων πανεπιστημίου** (Uni_grad_perc) συσχετίζεται αρνητικά με την εξαρτημένη μεταβλητή, δηλαδή όσο περισσότεροι οι απόφοιτοι πανεπιστημίου, τόσο λιγότεροι θάνατοι σε οδικά ατυχήματα. Με την υπόθεση ότι οι τελειόφοιτοι σχολών βρίσκονται σε μια ηλικία άνω των 24 ετών και έχουν σταδιακά αναπτύξει πιο ώριμη οδική συμπεριφορά, το πρόσημο εξηγείται λογικά.

Το **ποσοστό των γυναικών** συσχετίζεται αρνητικά με την εξαρτημένη μεταβλητή, κάτι που είναι αναμενόμενο καθώς έχει αποδειχθεί ότι οι γυναίκες οδηγούν πιο υπεύθυνα και συνεπώς εμπλέκονται σε λιγότερα οδικά ατυχήματα από τους άνδρες.

Το **ποσοστό των εργαζομένων** επίσης συσχετίζεται αρνητικά με την εξαρτημένη μεταβλητή. Όσο αυξάνονται οι απασχολούμενοι σε εργασία τόσο θα βελτιώνεται το βιοτικό επίπεδο της κοινωνίας, ενώ οι οδηγοί θα είναι πιο ενημερωμένοι και υπεύθυνοι, επομένως οι θάνατοι σε οδικά ατυχήματα ανά τετραγωνικό χιλιόμετρο θα μειώνονται.

Ο **αριθμός νοικοκυριών με ιδιοκτησία ΙΧ** (hearth_cars_own) συσχετίζεται θετικά με την εξαρτημένη μεταβλητή, κάτι που είναι απολύτως λογικό. Όσα περισσότερα νοικοκυριά έχουν ΙΧ αυτοκίνητο, τόσο αυξάνεται ο στόλος οχημάτων, κατ' επέκταση και η έκθεση στον κίνδυνο, και συνεπώς και οι θάνατοι εξαιτίας οδικών ατυχημάτων.

Το **Ακαθάριστο Εθνικό Προϊόν** συσχετίζεται αρνητικά με την εξαρτημένη μεταβλητή, κάτι που επίσης εξηγείται λογικά. Όσο το ΑΕΠ μιας χώρας αυξάνεται, θα αυξάνεται και το βιοτικό και μορφωτικό επίπεδο. Οι οδηγοί θα είναι πιο ενημερωμένοι, και το κράτος θα επενδύει περισσότερο σε υποδομές και λοιπά μέτρα βελτίωσης του οδικού περιβάλλοντος. Συνεπώς, οι θάνατοι σε οδικά ατυχήματα θα μειώνονται.

Συμπερασματικά, το πρότυπο που προέκυψε είναι στατιστικά ορθό και τα αποτελέσματά του εξηγούνται με βάση τη λογική. Είναι συνεπώς ένα επαρκές και αποδεκτό μοντέλο πρόβλεψης για το σύνολο της βάσης δεδομένων.

5.2.4 Σχετική επιρροή μεταβλητών

Σε αυτό το σημείο υπολογίζεται η σχετική επιρροή των ανεξάρτητων μεταβλητών του παραπάνω προτύπου στην εξαρτημένη, δηλαδή τον λογάριθμο των θανάτων σε οδικά ατυχήματα ανά τετραγωνικό χιλιόμετρο ("logFtpAr").

Με βάση την θεωρία της ελαστικότητας, όπως έχει αναφερθεί στο Κεφάλαιο 3, και μετά από τους κατάλληλους υπολογισμούς στο Microsoft Excel, οι τιμές της ελαστικότητας και της σχετικής επιρροής για τις ανεξάρτητες μεταβλητές είναι οι ακόλουθες.

ΣΥΝΟΛΟ ΔΗΜΩΝ			
Ανεξάρτητες μεταβλητές	ΚΛΑΣΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ		
	Τιμές συντελεστών	Σχετική επιρροή	
		ei* (σχετική επιρροή)	ei (ελαστικότητα)
Σταθερός όρος	7.428	-	-
Απόφοιτοι σχολών (%)	-0.019	1.00	-0.0023
Γυναίκες (%)	-0.065	-1.91	-0.0044
Εργαζόμενοι (%)	-0.057	-2.26	-0.0052
Ιδιοκτησία ΙΧ	$2 \cdot 10^{-5}$	10.83	0.0249
GDP	$(-8) \cdot 10^{-5}$	-11.13	-0.0256

Πίνακας 5.5: Σχετική επιρροή ανεξάρτητων μεταβλητών – Συνολικό πρότυπο

Από τον παραπάνω πίνακα προκύπτουν οι εξής παρατηρήσεις:

- Το **ποσοστό των απόφοιτων πανεπιστημίου** είναι η μεταβλητή που εμφανίζει την χαμηλότερη επιρροή στο πρότυπο.
- Η ανεξάρτητη μεταβλητή με την μεγαλύτερη επιρροή είναι το **Ακαθάριστο Εθνικό Προϊόν** (GDP). Αυτό σημαίνει ότι οι μεταβολές στο ΑΕΠ επηρεάζουν σημαντικά

τους θανάτους σε οδικά ατυχήματα. Συγκεκριμένα, σε σχέση με την μεταβλητή χαμηλότερης επιρροής, το ΑΕΠ επηρεάζει την εξαρτημένη μεταβλητή 11.13 φορές περισσότερο.

- Το **ποσοστό των γυναικών** και το **ποσοστό των εργαζομένων** έχουν χαμηλή σχετική επιρροή στην εξαρτημένη μεταβλητή, με απόλυτες τιμές e_i^* ίσες με 1.91 και 2.26, αντίστοιχα.
- Ο **αριθμός νοικοκυριών με ιδιοκτησία ΙΧ** φαίνεται να έχει αρκετά υψηλή επιρροή στην εξαρτημένη μεταβλητή, με τιμή e_i^* ίση με 10.83, που βρίσκεται πολύ κοντά στην μεγαλύτερη (αυτήν του ΑΕΠ).

5.3 Επεξεργασία ανά συστάδα

Όπως έχει εκτενώς αναφερθεί σε προηγούμενα κεφάλαια, τα δεδομένα ομαδοποιήθηκαν με αλγόριθμο ιεραρχικής ομαδοποίησης, από όπου προέκυψαν δύο συστάδες δεδομένων. Για κάθε μία ξεχωριστά, αναπτύχθηκαν τα μοντέλα παλινδρόμησης όπως παρουσιάζονται παρακάτω.

5.3.1 Πρότυπο 1^{ης} Συστάδας Δήμων

5.3.1.1 Ανάπτυξη μαθηματικού προτύπου

Για την συστάδα **Cluster-1**, έπειτα από πολλές δοκιμές αναπτύχθηκε ένα μοντέλο **λογαριθμοκανονικής παλινδρόμησης** όπου ως εξαρτημένη μεταβλητή χρησιμοποιείται ο δεκαδικός λογάριθμος των θανάτων σε οδικά ατυχήματα ανά τετραγωνικό χιλιόμετρο (“logFtpAr”). Οι ανεξάρτητες μεταβλητές περιλαμβάνουν αυτές από το συνολικό πρότυπο, με την προσθήκη μίας ακόμα. Συγκεκριμένα, είναι:

- “Uni_grad_perc”: ποσοστό (%) των απόφοιτων πανεπιστημιακών σχολών
- “Fem_perc”: ποσοστό (%) του γυναικείου πληθυσμού σε σχέση με τον συνολικό
- “Work_perc”: ποσοστό (%) των απασχολούμενων σε εργασία
- “hearth_cars_own”: αριθμός νοικοκυριών με ιδιοκτησία τουλάχιστον ενός ΙΧ
- “GDP2019”: Ακαθάριστο Εθνικό Προϊόν 2019
- “elder_perc”: ποσοστό (%) του πληθυσμού ηλικιωμένων σε σχέση με τον συνολικό

```
86 #Linear regression logarithmic (Area)
87 model6 <- lm(logFtpAr ~ uni_grad_perc + Fem_perc + work_perc + hearth_cars_own + GDP2019 + elder_perc , data = cluster1)
88 summary(model6)
89 vif(model6)
90 #Final cluster 1 model
91
```

Εικόνα 5.4: Προσαρμογή στο λογισμικό R – Πρότυπο 1^{ης} συστάδας

Πρέπει να σημειωθεί ότι η μεταβλητή “elder_perc” προστέθηκε μετά από πολλές δοκιμές, επειδή στα ομαδοποιημένα δεδομένα ο συνδυασμός των μεταβλητών του συνολικού προτύπου δεν έδινε ορθά αποτελέσματα. Στον παραπάνω κώδικα λογισμικού R φαίνεται η προσαρμογή του προτύπου για την πρώτη συστάδα. Τα αποτελέσματα της διαδικασίας είναι τα ακόλουθα.

Coefficients				
	Estimate	Standard Error	p - value	t - value
(Intercept)	10.109	3.440	0.005	2.939
Uni_grad_perc	0.039	0.009	< 0.001	3.981
Fem_perc	-0.239	0.053	< 0.001	-4.488
Work_perc	-0.087	0.019	< 0.001	-4.448
hearth_cars_own	$9.3 \cdot 10^{-6}$	0	< 0.001	3.877
GDP2019	$(-1.1) \cdot 10^{-4}$	0	< 0.001	-7.563
elder_perc	0.040	0.021	0.058	1.925

Πίνακας 5.6: Αποτελέσματα παλινδρόμησης – Πρότυπο 1^{ης} συστάδας

Residuals				
Min	1Q	Median	3Q	Max
-1.052	-0.220	0.010	0.273	0.618

Πίνακας 5.7: Υπολείμματα – Πρότυπο 1^{ης} συστάδας

Model Summary	
Residual std. error	0.352 on 61 degrees of freedom
Multiple R-squared	0.828
Adjusted R-squared	0.811
p - value	< 0.001

Πίνακας 5.8: Περίληψη αποτελεσμάτων – Πρότυπο 1^{ης} συστάδας

Variance Inflation Factor (VIF)					
Uni_grad_perc	Fem_perc	Work_perc	hearth_cars_own	GDP2019	elder_perc
3.034	2.357	2.551	1.221	1.753	1.740

Πίνακας 5.9: Δείκτης VIF – Πρότυπο 1^{ης} συστάδας

Επομένως, η μαθηματική σχέση του προτύπου θα είναι:

$$\log FtpAr = 10.109 + 0.039 \cdot (\text{Uni_grad_perc}) - 0.239 \cdot (\text{Fem_perc}) - 0.087 \cdot (\text{Work_perc}) + 9.3 \cdot 10^{-6} \cdot (\text{hearth_cars_own}) - 1.1 \cdot 10^{-4} \cdot (\text{GDP2019}) + 0.04 \cdot (\text{elder_perc})$$

5.3.1.2 Ποιότητα προτύπου & σχολιασμός

Αντίστοιχα με το προηγούμενο πρότυπο προκύπτουν τα εξής:

- Ο προσαρμοσμένος συντελεστής R^2 ισούται με 0.811 και θεωρείται εξαιρετικός.

- Τα **υπολείμματα** της παλινδρόμησης είναι κοντά στο 0 και συνεπώς θεωρούνται αμελητέα.
- Το **σφάλμα** της παλινδρόμησης για κάθε μεταβλητή είναι πολύ χαμηλό.
- Η τιμή **p – value** για το ποσοστό των ηλικιωμένων (elder_perc) βρίσκεται στο όριο του 0.05 και θα θεωρηθεί αποδεκτή. Όλες οι υπόλοιπες μεταβλητές έχουν τιμές p – value πολύ χαμηλότερες.
- Όλες οι μεταβλητές έχουν **δείκτη VIF** χαμηλότερο από 5.

Στην συγκεκριμένη περίπτωση – αντίθετα από το συνολικό πρότυπο – το **ποσοστό των τελειόφοιτων πανεπιστημίου** συσχετίζεται θετικά με τους θανάτους σε οδικά ατυχήματα. Όπως προέκυψε από την διαδικασία της ομαδοποίησης στο Κεφάλαιο 4, η συστάδα Cluster-1 περιέχει τους μεγαλύτερους πληθυσμιακά Δήμους της χώρας, κυρίως μεγάλες πόλεις. Μία λογική εξήγηση για το πρόσημο θα ήταν η αύξηση των μετακινήσεων στους Δήμους αυτούς, κάτι που συνεπάγεται την αυξημένη έκθεση στον κίνδυνο, και τελικά την αύξηση των θανάτων σε οδικά ατυχήματα.

Τα πρόσημα για τις μεταβλητές “ποσοστό γυναικών”, “ποσοστό εργαζομένων”, “νοικοκυριά με ιδιοκτησία ΙΧ” και “Ακαθάριστο Εθνικό Προϊόν” είναι τα ίδια με αυτά που προέκυψαν στο συνολικό πρότυπο, συνεπώς εξηγούνται λογικά για τους λόγους που αναφέρθηκαν παραπάνω.

Τέλος, το **ποσοστό των ηλικιωμένων** συσχετίζεται θετικά με την εξαρτημένη μεταβλητή. Πρόκειται για λογικό αποτέλεσμα, καθώς οι ηλικιωμένοι ανήκουν στους ευάλωτους χρήστες του οδικού περιβάλλοντος, και λόγω αδυναμίας και μειωμένων αντανακλαστικών εμπλέκονται περισσότερο σε οδικά ατυχήματα.

Συνεπώς, το πρότυπο είναι αποδεκτό και περιγράφει ικανοποιητικά την συσχέτιση των οδικών ατυχημάτων με τις προαναφερθείσες ανεξάρτητες μεταβλητές, στο πλαίσιο της συστάδας Cluster-1.

5.3.1.3 Σχετική επιρροή μεταβλητών

Αντίστοιχα με το συνολικό πρότυπο, προέκυψαν οι τιμές ελαστικότητας και σχετικής επιρροής για τις ανεξάρτητες μεταβλητές, οι οποίες παρουσιάζονται παρακάτω.

ΣΥΣΤΑΔΑ Cluster 1			
Ανεξάρτητες μεταβλητές	ΚΛΑΣΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ		
	Τιμές συντελεστών	Σχετική επιρροή	
		ei*	ei (ελαστικότητα)
		(σχετική επιρροή)	
Σταθερός όρος	10.109	-	-
Απόφοιτοι σχολών (%)	0.039	1.00	0.0032
Γυναίκες (%)	-0.239	-3.69	-0.0118
Εργαζόμενοι (%)	-0.087	-1.78	-0.0057
Ιδιοκτησία ΙΧ	$9.3 \cdot 10^{-6}$	7.50	0.0240
GDP	$(-1.1) \cdot 10^{-4}$	-14.09	-0.0451
Ηλικιωμένοι (%)	0.04	1.09	0.0035

Πίνακας 5.10: Σχετική επιρροή ανεξάρτητων μεταβλητών – 1^η συστάδα

Παρατηρώντας τον παραπάνω πίνακα, προκύπτουν τα εξής συμπεράσματα:

- Την μικρότερη επιρροή στην εξαρτημένη μεταβλητή έχει ξανά το **ποσοστό απόφοιτων πανεπιστημίου**.
- Όπως και στο συνολικό πρότυπο, το **Ακαθάριστο Εθνικό Προϊόν** έχει την μεγαλύτερη επιρροή στην εξαρτημένη μεταβλητή, συγκεκριμένα 14.09 φορές μεγαλύτερη από την χαμηλότερη.
- Πολύ χαμηλή σχετική επιρροή εμφανίζει το **ποσοστό των ηλικιωμένων** με τιμή e_i^* ίση με 1.09, που βρίσκεται πολύ κοντά στην χαμηλότερη τιμή. Επίσης, το **ποσοστό των εργαζόμενων** παρουσιάζει χαμηλή σχετική επιρροή (απόλυτη $e_i^* = 1.78$).
- Το **ποσοστό των γυναικών** παρουσιάζει χαμηλή επιρροή, με απόλυτη τιμή e_i^* ίση με 3.69, που ωστόσο είναι αυξημένη σε σχέση με το συνολικό πρότυπο.
- Τέλος, ο **αριθμός νοικοκυριών με ιδιοκτησία ΙΧ** εμφανίζει ενδιάμεση σχετική επιρροή στο πρότυπο, καθώς η τιμή e_i^* της εν λόγω μεταβλητής ισούται με 7.50, σχεδόν μισή από αυτήν του ΑΕΠ, που είναι και η μεγαλύτερη.

5.3.2 Πρότυπο 2^{ης} Συστάδας Δήμων

5.3.2.1 Ανάπτυξη μαθηματικού προτύπου

Για την συστάδα **Cluster-2** αναπτύχθηκε ένα μοντέλο **λογαριθμοκανονικής παλινδρόμησης** όπου όλες οι μεταβλητές είναι ίδιες με το προηγούμενο πρότυπο για την 1^η συστάδα. Παρακάτω φαίνεται ο κώδικας στο λογισμικό R, καθώς και τα αποτελέσματα του προτύπου της 2^{ης} συστάδας.

```
110 model19.1 <- lm(logFtpAR ~ Uni_grad_perc + Fem_perc + work_perc + hearth_cars_own + GDP2019 + elder_perc , data = cluster2)
111 summary(model19.1)
112 vif(model19.1)
113 #final cluster 2 model
114
```

Εικόνα 5.5: Προσαρμογή στο λογισμικό R – Πρότυπο 2^{ης} συστάδας

Coefficients				
	Estimate	Standard Error	p - value	t - value
(Intercept)	-0.628	1.747	0.719	-0.360
Uni_grad_perc	-0.021	0.008	0.017	-2.404
Fem_perc	-0.017	0.025	0.494	-0.685
Work_perc	-0.021	0.009	0.032	-2.152
hearth_cars_own	$6.5 \cdot 10^{-5}$	0	< 0.001	4.442
GDP2019	$(-6.6) \cdot 10^{-5}$	0	< 0.001	-3.917
elder_perc	0.036	0.007	< 0.001	4.651

Πίνακας 5.11: Αποτελέσματα παλινδρόμησης – Πρότυπο 2^{ης} συστάδας

Residuals				
Min	1Q	Median	3Q	Max
-0.973	-0.345	-0.025	0.207	2.286

Πίνακας 5.12: Υπολείμματα – Πρότυπο 2^{ης} συστάδας

Model Summary	
Residual std. error	0.548 on 230 degrees of freedom
Multiple R-squared	0.411
Adjusted R-squared	0.396
p - value	< 0.001

Πίνακας 5.13: Περίληψη αποτελεσμάτων – Πρότυπο 2^{ης} συστάδας

Variance Inflation Factor (VIF)					
Uni_grad_perc	Fem_perc	Work_perc	hearth_cars_own	GDP2019	elder_perc
2.022	1.257	2.076	1.839	1.245	1.576

Πίνακας 5.14: Δείκτης vif – Πρότυπο 2^{ης} συστάδας

Η μαθηματική εξίσωση που περιγράφει το πρότυπο αυτό είναι:

$$\log FtpAr = -0.628 - 0.021*(Uni_grad_perc) - 0.017*(Fem_perc) - 0.021*(Work_perc) + 6.5*10^{(-5)}*(hearth_cars_own) - 6.6*10^{(-5)}*(GDP2019) + 0.036*(elder_perc)$$

5.3.2.2 Ποιότητα προτύπου & σχολιασμός

Όμοια με τα προηγούμενα πρότυπα προκύπτουν τα παρακάτω.

- Ο **προσαρμοσμένος συντελεστής R²** ισούται με 0.396, και θεωρείται οριακά αποδεκτός.
- Τα **υπολείμματα** της παλινδρόμησης είναι κοντά στο 0 και συνεπώς θεωρούνται αμελητέα.
- Το **σφάλμα** της παλινδρόμησης για κάθε μεταβλητή είναι πολύ χαμηλό.
- Η τιμή **p – value** για το ποσοστό των γυναικών είναι υψηλή (μεγαλύτερη από 0.05). Επίσης, η τιμή του συντελεστή t_{stat} (t – value) είναι κατά απόλυτη τιμή ιδιαίτερα χαμηλή. Συνεπώς, η συγκεκριμένη ανεξάρτητη μεταβλητή δεν είναι στατιστικά σημαντική. Η τιμή p – value για τις υπόλοιπες μεταβλητές είναι αποδεκτή, και άρα αυτές θα είναι στατιστικά σημαντικές.
- Όλες οι μεταβλητές έχουν **δείκτη VIF** χαμηλότερο από 5.

Το **ποσοστό των απόφοιτων πανεπιστημίου** συσχετίζεται αρνητικά με την εξαρτημένη μεταβλητή. Χρησιμοποιώντας την λογική που αναπτύχθηκε στο πρότυπο της 1^{ης} συστάδας, και λαμβάνοντας υπόψη ότι η συστάδα Cluster-2 περιλαμβάνει τους πιο επαρχιακούς Δήμους, το πρόσημο της μεταβλητής εξηγείται λογικά. Τα πρόσημο των υπόλοιπων μεταβλητών είναι ίδια με αυτά στο προηγούμενο πρότυπο, και συνεπώς εξηγούνται λογικά.

Επομένως, το πρότυπο είναι επαρκές για την πρόβλεψη στην συστάδα Cluster-2.

5.3.2.3 Σχετική επιρροή μεταβλητών

Παρόμοια με τα προηγούμενα πρότυπα, υπολογίστηκαν οι τιμές της ελαστικότητας και της αντίστοιχης σχετικής επιρροής για τις ανεξάρτητες μεταβλητές. Τα αποτελέσματα παρουσιάζονται στον παρακάτω πίνακα.

ΣΥΣΤΑΔΑ Cluster 2			
Ανεξάρτητες μεταβλητές	ΚΛΑΣΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ		
	Τιμές συντελεστών	Σχετική επιρροή	
		e_i^* (σχετική επιρροή)	e_i (ελαστικότητα)
Σταθερός όρος	-0.628	-	-
Απόφοιτοι σχολών (%)	-0.021	2.89	0.0182
Γυναίκες (%)	-0.017	1.00	0.0063
Εργαζόμενοι (%)	-0.021	1.92	0.0121
Ιδιοκτησία ΙΧ	$6.5 \cdot 10^{-5}$	57.54	0.3625
GDP	$(-6.6) \cdot 10^{-5}$	38.32	0.2414
Ηλικιωμένοι (%)	0.036	-2.84	-0.0179

Πίνακας 5.15: Σχετική επιρροή ανεξάρτητων μεταβλητών – 2^η συστάδα

Από τον παραπάνω πίνακα προκύπτουν οι εξής παρατηρήσεις:

- Η ανεξάρτητη μεταβλητή με την χαμηλότερη σχετική επιρροή σε αυτό το πρότυπο είναι το **ποσοστό των γυναικών**.
- Την μεγαλύτερη σχετική επιρροή στο πρότυπο παρουσιάζει ο **αριθμός νοικοκυριών με ιδιοκτησία ΙΧ**, με τιμή e_i^* ίση με 57.54. Επίσης, μεγάλη σχετική επιρροή εμφανίζει το **Ακαθάριστο Εθνικό Προϊόν** ($e_i^* = 38.32$), το οποίο είχε την μεγαλύτερη επιρροή στα δύο προηγούμενα πρότυπα.
- Μεταβλητές που παρουσιάζουν χαμηλή σχετική επιρροή είναι το **ποσοστό ηλικιωμένων** (απόλυτη $e_i^* = 2.84$), το **ποσοστό απόφοιτων σχολών** ($e_i^* = 2.89$), και κυρίως το **ποσοστό εργαζομένων** ($e_i^* = 1.92$).

5.4 Χωρική Ανάλυση

Στην ενότητα αυτή παρουσιάζονται τα μοντέλα **Γεωγραφικά Σταθμισμένης Παλινδρόμησης** (GWR) που αναπτύχθηκαν στο πλαίσιο της Διπλωματικής Εργασίας. Ταυτόχρονα, πραγματοποιείται σύγκριση των αποτελεσμάτων με αυτά των κλασικών μοντέλων παλινδρόμησης που παρουσιάστηκαν παραπάνω. Επισημαίνεται ότι σε αυτή την ενότητα, δεν πραγματοποιήθηκε ανάλυση ελαστικότητας / σχετικής επιρροής, καθώς τα χωρικά πρότυπα περιλαμβάνουν ακριβώς τις ίδιες μεταβλητές με τα αντίστοιχα κλασικά.

Εδώ δίνεται μεγαλύτερη έμφαση στην γεωγραφική παράμετρο της ανάλυσης, που άλλωστε αποτελεί και το κύριο αντικείμενο της παρούσας Διπλωματικής Εργασίας.

5.4.1 Ανάπτυξη χωρικού μοντέλου για το σύνολο των Δήμων

Όπως έχει αναφερθεί στο Κεφάλαιο 3, η μέθοδος **GWR** ξεκινά με την μετατροπή της βάσης δεδομένων σε μορφή συμβατή με χωρική παλινδρόμηση, χρησιμοποιώντας τις γεωγραφικές συντεταγμένες του κέντρου κάθε Δήμου.

Στον παρακάτω κώδικα λογισμικού R παρουσιάζεται η προσέγγιση με χωρική μέθοδο για το σύνολο της βάσης δεδομένων.

```
120
121 #Spatial Approach Total
122
123 #convert dataframe into spatial
124 library(sf)
125 library(gwmodel)
126 Municipal_sp <- st_as_sf(Municipal, coords = c("Long", "Lat"), crs = 4326)
127
128 # determine kernel bandwidth
129 class(Municipal_sp)
130
131 # convert to sp
132 Municipal_sp = as(Municipal_sp, "Spatial")
133 bw <- bw.gwr(logFtpr ~ Uni_grad_perc + Fem_perc + Work_perc + hearth_cars_own + GDP2019,
134             approach = "AIC",
135             adaptive = T,
136             data=Municipal.sp)
137
138 # fit GWR model
139 m.gwr <- gwr.basic(logFtpr ~ Uni_grad_perc + Fem_perc + Work_perc + hearth_cars_own + GDP2019,
140                  adaptive = T,
141                  data = Municipal.sp,
142                  bw = bw)
143 m.gwr
144
```

Εικόνα 5.6: Προσαρμογή στο λογισμικό R – Συνολικό χωρικό μοντέλο

Στη συνέχεια, το πρόγραμμα εκτελεί δοκιμές για διάφορες τιμές του εύρους ζώνης (bandwidth) προκειμένου να ελαχιστοποιηθεί ο δείκτης AICc, όπως φαίνεται παρακάτω.

```
Adaptive bandwidth (number of nearest neighbours): 196 AICc value: 463.3405
Adaptive bandwidth (number of nearest neighbours): 129 AICc value: 434.9047
Adaptive bandwidth (number of nearest neighbours): 87 AICc value: 435.7789
Adaptive bandwidth (number of nearest neighbours): 154 AICc value: 445.2526
Adaptive bandwidth (number of nearest neighbours): 112 AICc value: 434.0177
Adaptive bandwidth (number of nearest neighbours): 103 AICc value: 434.4058
Adaptive bandwidth (number of nearest neighbours): 119 AICc value: 433.3281
Adaptive bandwidth (number of nearest neighbours): 122 AICc value: 433.6634
Adaptive bandwidth (number of nearest neighbours): 115 AICc value: 433.0481
Adaptive bandwidth (number of nearest neighbours): 115 AICc value: 433.0481
>
```

Εικόνα 5.7: Διαδικασία καθορισμού εύρους ζώνης – Συνολικό πρότυπο

Στην παραπάνω εικόνα, έπειτα από έναν αριθμό δοκιμών, το πρόγραμμα κατέληξε σε τιμή εύρους ζώνης 115. Αυτό σημαίνει ότι η χωρική παλινδρόμηση θα εξετάσει 115 γειτονικούς Δήμους για την ανάλυσή της. Για το συγκεκριμένο χωρικό μοντέλο, χρησιμοποιούνται ακριβώς οι ίδιες μεταβλητές με το απλό μοντέλο παλινδρόμησης για το σύνολο των Δήμων, έτσι ώστε να συγκριθούν πιο ουσιαστικά τα αποτελέσματα.

Το λογισμικό εκτελεί ξανά την απλή παλινδρόμηση (global regression), της οποίας τα αποτελέσματα παρουσιάστηκαν στην ενότητα 5.2.2, και ύστερα χρησιμοποιώντας μικρο – παλινδρομήσεις εξάγει τα τελικά αποτελέσματα της χωρικής μεθόδου, τα οποία είναι τα παρακάτω.

Summary of GWR coefficient estimates					
	Min	1Q	Median	3Q	Max
(Intercept)	-7.090	1.349	5.353	11.106	16.105
Uni_grad_perc	-0.067	-0.017	0.003	0.024	0.032
Fem_perc	-0.250	-0.165	-0.081	-0.023	0.146
Work_perc	-0.094	-0.084	-0.053	-0.029	0.045
hearth_cars_own	$(-1.3) \cdot 10^{-6}$	$1.5 \cdot 10^{-5}$	$2.4 \cdot 10^{-5}$	$2.8 \cdot 10^{-5}$	0
GDP2019	$(-2.7) \cdot 10^{-4}$	$(-1.6) \cdot 10^{-4}$	$(-9.9) \cdot 10^{-5}$	$(-6.9) \cdot 10^{-5}$	0

Πίνακας 5.13: Εκτιμήσεις συντελεστών GWR – Συνολικό πρότυπο

Calibration & Diagnostic Information	
Adaptive bandwidth	115
Effective degrees of freedom	256.981
AIC	381.628
R-squared	0.732
Adjusted R-squared	0.681

Πίνακας 5.14: Περίληψη αποτελεσμάτων GWR – Συνολικό πρότυπο

5.4.2 Ποιότητα & σχολιασμός συνολικού χωρικού προτύπου

- Ο **προσαρμοσμένος συντελεστής R^2** προκύπτει ίσος με 0.681 και κρίνεται ικανοποιητικός. Επιπλέον, ο αντίστοιχος συντελεστής στο απλό μοντέλο παλινδρόμησης για το σύνολο των Δήμων προέκυψε 0.529. Συνεπώς, η εφαρμογή της χωρικής μεθόδου GWR βελτίωσε την προσαρμογή στα δεδομένα.
- Το **Κριτήριο Πληροφοριών Akaike (AIC)** ισούται με 381.628 και είναι χαμηλότερο σε σχέση με αυτό που προέκυψε στην απλή παλινδρόμηση (524.403). Όπως έχει αναφερθεί στο Κεφάλαιο 3, επιθυμητό είναι ο δείκτης αυτός να είναι όσο μικρότερος γίνεται. Επομένως, η χωρική μέθοδος φαίνεται ότι έχει καλύτερη προγνωστική ακρίβεια.
- Για την σύγκριση των εκτιμήσεων των **συντελεστών παλινδρόμησης** ανάμεσα στο απλό και το χωρικό πρότυπο, θα χρησιμοποιηθεί η διάμεσος των εκτιμήσεων (median) του Πίνακα 5.13.

Παρατηρείται ότι τα **πρόσημα των συντελεστών** ταυτίζονται με αυτά του απλού μοντέλου, με μόνη εξαίρεση την μεταβλητή “Uni_grad_perc”, δηλαδή το ποσοστό των απόφοιτων πανεπιστημίου. Ωστόσο, όπως φαίνεται στον Πίνακα 5.13, ο συντελεστής της συγκεκριμένης μεταβλητής μπορεί να πάρει και αρνητικές τιμές. Στην περίπτωση αυτή, όλα τα πρόσημα είναι λογικά, όπως έχει εξηγηθεί στην ενότητα 5.2.2. Επιπλέον, το πρότυπο της μεθόδου GWR υπολογίζει μεγάλο εύρος τιμών για τους συντελεστές παλινδρόμησης, κάτι που αυξάνει την ακρίβεια του προτύπου.

5.4.3 Ανάπτυξη χωρικού μοντέλου για τους Δήμους της 1^{ης} συστάδας

Για τα δεδομένα της συστάδας **Cluster-1** χρησιμοποιήθηκαν οι ίδιες μεταβλητές με το απλό μοντέλο παλινδρόμησης. Παρακάτω φαίνεται ο κώδικας του προτύπου στο λογισμικό R.

```

146 #spatial approach Cluster 1
147
148 Cluster1_sp <- st_as_sf(Cluster1, coords = c("Long", "Lat"), crs = 4326)
149 class(Cluster1_sp)
150
151 Cluster1.sp = as(Cluster1_sp, "Spatial")
152 bw <- bw.gwr(logFtpAr ~ Uni_grad_perc + Fem_perc + work_perc + hearth_cars_own + GDP2019 + elder_perc,
153             approach = "AIC",
154             adaptive = T,
155             data = Cluster1.sp)
156
157 m.gwr <- gwr.basic(logFtpAr ~ Uni_grad_perc + Fem_perc + work_perc + hearth_cars_own + GDP2019 + elder_perc,
158                  adaptive = T,
159                  data = Cluster1.sp,
160                  bw = bw)
161 m.gwr
162

```

Εικόνα 5.8: Προσαρμογή στο λογισμικό R – Χωρικό μοντέλο 1^{ης} συστάδας

Μετά από δοκιμές, το πρόγραμμα κατέληξε σε ένα εύρος ζώνης (πλήθος κοντινότερων γειτονικών περιοχών) ίσο με 66, τιμή για την οποία οριστικοποιείται ο ελάχιστος δείκτης AICc. Οι διαδοχικές δοκιμές φαίνονται στην παρακάτω εικόνα.

```

Adaptive bandwidth (number of nearest neighbours): 49 AICC value: 70.87385
Adaptive bandwidth (number of nearest neighbours): 38 AICC value: 82.0796
Adaptive bandwidth (number of nearest neighbours): 56 AICC value: 65.92764
Adaptive bandwidth (number of nearest neighbours): 60 AICC value: 58.35884
Adaptive bandwidth (number of nearest neighbours): 63 AICC value: 55.16018
Adaptive bandwidth (number of nearest neighbours): 64 AICC value: 55.12557
Adaptive bandwidth (number of nearest neighbours): 66 AICC value: 54.98926
Adaptive bandwidth (number of nearest neighbours): 66 AICC value: 54.98926
>

```

Εικόνα 5.9: Διαδικασία καθορισμού εύρους ζώνης – 1^η συστάδα

Τα αποτελέσματα του μοντέλου χωρικής παλινδρόμησης είναι τα ακόλουθα.

Summary of GWR coefficient estimates					
	Min	1Q	Median	3Q	Max
(Intercept)	0.123	0.729	3.141	10.090	10.613
Uni_grad_perc	0.023	0.032	0.033	0.047	0.048
Fem_perc	-0.277	-0.271	-0.189	-0.186	-0.147
Work_perc	-0.105	-0.102	-0.052	-0.043	-0.018
hearth_cars_own	6.5*10 ^{^-6}	6.6*10 ^{^-6}	7.1*10 ^{^-6}	9*10 ^{^-6}	0
GDP2019	(-1.4)*10 ^{^-4}	(-1.1)*10 ^{^-4}	(-1.1)*10 ^{^-4}	(-9.3)*10 ^{^-5}	(-10) ^{^-4}
elder_perc	0.019	0.061	0.086	0.097	0.099

Πίνακας 5.15: Εκτιμήσεις συντελεστών GWR – 1^η συστάδα

Calibration & Diagnostic Information	
Adaptive bandwidth	66
Effective degrees of freedom	53.939
AIC	34.461
R-squared	0.874
Adjusted R-squared	0.841

Πίνακας 5.16: Περίληψη αποτελεσμάτων GWR – 1^η συστάδα

5.4.4 Ποιότητα & σχολιασμός χωρικού προτύπου 1^{ης} συστάδας

- Ο **προσαρμοσμένος συντελεστής R²** της χωρικής παλινδρόμησης ισούται με 0.841, και είναι πολύ ικανοποιητικός. Ταυτόχρονα, είναι υψηλότερος από τον αντίστοιχο της απλής παλινδρόμησης που προέκυψε ίσως με 0.811. Επομένως, η χωρική μέθοδος βελτιώνει το μαθηματικό μοντέλο.
- Ο **δείκτης AIC** προέκυψε 34.461 και είναι χαμηλότερος από την τιμή του στην απλή παλινδρόμηση που είναι 59.663, συνεπώς η ακρίβεια του μοντέλου βελτιώθηκε με την εφαρμογή χωρικής μεθόδου.
- Χρησιμοποιώντας τον Πίνακα 5.15 και με κριτήριο την διάμεσο των εκτιμήσεων (median), προκύπτει ότι τα **πρόσημα των συντελεστών** της χωρικής παλινδρόμησης ταυτίζονται πλήρως με αυτά του κλασικού προτύπου, και επομένως εξηγούνται λογικά.

5.4.5 Ανάπτυξη χωρικού μοντέλου για τους Δήμους της 2^{ης} συστάδας

Για το πρότυπο χωρικής παλινδρόμησης για την συστάδα **Cluster-2** χρησιμοποιήθηκαν ξανά οι ίδιες μεταβλητές με το αντίστοιχο κλασικό μοντέλο. Στην παρακάτω εικόνα παρουσιάζεται η προσαρμογή στο λογισμικό R του χωρικού προτύπου της 2^{ης} συστάδας.

```

164 #Spatial approach Cluster 2
165
166 Cluster2_sp <- st_as_sf(Cluster2, coords = c("Long", "Lat"), crs = 4326)
167 class(Cluster2_sp)
168
169 Cluster2.sp = as(Cluster2_sp, "Spatial")
170 bw <- bw.gwr(logFtpAr ~ Uni_grad_perc + Fem_perc + Work_perc + hearth_cars_own + GDP2019 + elder_perc,
171             approach = "AIC",
172             adaptive = T,
173             data = Cluster2.sp)
174
175 m.gwr <- gwr.basic(logFtpAr ~ Uni_grad_perc + Fem_perc + Work_perc + hearth_cars_own + GDP2019 + elder_perc,
176                  adaptive = T,
177                  data = Cluster2.sp,
178                  bw = bw)
179 m.gwr

```

Εικόνα 5.10: Προσαρμογή στο λογισμικό R – Χωρικό μοντέλο 2^{ης} συστάδας

Το εύρος ζώνης καθορίστηκε στην τιμή 126, με κριτήριο την ελαχιστοποίηση του δείκτη AICc. Η διαδικασία καθορισμού και τα αποτελέσματα φαίνονται παρακάτω.

```

Adaptive bandwidth (number of nearest neighbours): 154 AICc value: 354.7924
Adaptive bandwidth (number of nearest neighbours): 103 AICc value: 357.3202
Adaptive bandwidth (number of nearest neighbours): 186 AICc value: 357.2479
Adaptive bandwidth (number of nearest neighbours): 134 AICc value: 353.6248
Adaptive bandwidth (number of nearest neighbours): 122 AICc value: 353.1918
Adaptive bandwidth (number of nearest neighbours): 114 AICc value: 353.9965
Adaptive bandwidth (number of nearest neighbours): 126 AICc value: 352.5221
Adaptive bandwidth (number of nearest neighbours): 129 AICc value: 352.6903
Adaptive bandwidth (number of nearest neighbours): 124 AICc value: 353.0055
Adaptive bandwidth (number of nearest neighbours): 127 AICc value: 352.6649
Adaptive bandwidth (number of nearest neighbours): 125 AICc value: 352.6752
Adaptive bandwidth (number of nearest neighbours): 126 AICc value: 352.5221

```

Εικόνα 5.11: Διαδικασία καθορισμού εύρους ζώνης – 2^η συστάδα

Summary of GWR coefficient estimates					
	Min	1Q	Median	3Q	Max
(Intercept)	-3.567	-1.221	0.439	2.778	4.584
Uni_grad_perc	-0.055	-0.013	0.008	0.024	0.033
Fem_perc	-0.109	-0.091	-0.039	0.003	0.058
Work_perc	-0.071	-0.052	-0.040	-0.014	0.041
hearth_cars_own	$2.8 \cdot 10^{-5}$	$6.8 \cdot 10^{-5}$	$8.4 \cdot 10^{-5}$	$9.4 \cdot 10^{-5}$	10^{-4}
GDP2019	$(-2.2) \cdot 10^{-4}$	$(-1.7) \cdot 10^{-4}$	$(-10)^{-4}$	$(-4.9) \cdot 10^{-5}$	0
elder_perc	-0.009	0.006	0.013	0.026	0.061

Πίνακας 5.17: Εκτιμήσεις συντελεστών GWR – 2^η συστάδα

Calibration & Diagnostic Information	
Adaptive bandwidth	126
Effective degrees of freedom	199.356
AIC	311.699
R-squared	0.611
Adjusted R-squared	0.537

Πίνακας 5.18: Περίληψη αποτελεσμάτων GWR – 2^η συστάδα

5.4.6 Ποιότητα & σχολιασμός χωρικού προτύπου 2^{ης} συστάδας

- Ο προσαρμοσμένος συντελεστής R^2 προέκυψε 0.537 και ξεπερνά την αντίστοιχη τιμή της κλασικής παλινδρόμησης (0.396), επομένως η χωρική μέθοδος βελτίωσε την ανάλυση.
- Ο δείκτης **AIC** της χωρικής μεθόδου ισούται με 311.699, ενώ ο αντίστοιχος της κλασικής μεθόδου είναι 396.276 που είναι υψηλότερος, οπότε η ακρίβεια του προτύπου βελτιώνεται με την εφαρμογή χωρικής ανάλυσης.
- Τα **πρόσημα των συντελεστών** παλινδρόμησης – με κριτήριο την διάμεσο – δεν ταυτίζονται πλήρως ανάμεσα στο κλασικό και στο χωρικό πρότυπο. Ωστόσο, όπως φαίνεται στον Πίνακα 5.17, οι συντελεστές δέχονται ένα εύρος τιμών που περιλαμβάνει και αρνητικούς αριθμούς, στην οποία περίπτωση τα πρόσημα θα ταυτίζονταν πλήρως.

Κεφάλαιο 6: Συμπεράσματα

6.1 Σύνοψη αποτελεσμάτων

Αντικείμενο της παρούσας Διπλωματικής Εργασίας είναι η **ανάλυση των οδικών ατυχημάτων στους Δήμους της Ελλάδας με χωρικά στατιστικά μοντέλα**. Πιο συγκεκριμένα, επιδιώκεται η σύγκριση των αποτελεσμάτων των χωρικών μοντέλων με αυτά των πιο κλασικών μεθόδων παλινδρόμησης. Σε κάθε περίπτωση, πραγματοποιείται διερεύνηση σχετικά με τον συνδυασμό παραγόντων που επηρεάζουν τα οδικά ατυχήματα. Στη συνέχεια, με τον στόχο της Διπλωματικής Εργασίας καθορισμένο, ακολούθησε η **βιβλιογραφική ανασκόπηση** προς αναζήτηση ερευνών με παρόμοιο αντικείμενο.

Ακολούθως, χρησιμοποιώντας τις βάσεις δεδομένων της Ελληνικής Στατιστικής Αρχής, πραγματοποιήθηκε η **συλλογή των απαραίτητων στοιχείων** για την στατιστική ανάλυση. Η βάση δεδομένων της Διπλωματικής Εργασίας ξεκινά με την καταγραφή του αριθμού θανάτων σε οδικά ατυχήματα κατά την περίοδο 2016 – 2020, ανά Δήμο. Τα υπόλοιπα δεδομένα περιλαμβάνουν οικονομικά, κοινωνικά και δημογραφικά χαρακτηριστικά για τους Δήμους της Ελλάδας, καθώς και τις γεωγραφικές συντεταγμένες του κέντρου κάθε Δήμου που προέκυψαν από τους διαδικτυακούς χάρτες της Google.

Σε αρχικό στάδιο, η βάση δεδομένων της Διπλωματικής Εργασίας αναπτύχθηκε στο περιβάλλον του Microsoft Excel και στη συνέχεια εισήχθη στο ειδικό στατιστικό λογισμικό R. Με εφαρμογή ενός **αλγορίθμου ιεραρχικής ομαδοποίησης**, οι Δήμοι κατατάχθηκαν σε δύο συστάδες, οι οποίες υπέστησαν ξεχωριστή επεξεργασία. Για κάθε συστάδα, προέκυψαν ορισμένα βασικά στατιστικά αποτελέσματα καθώς και γραφήματα, με σκοπό την καλύτερη κατανόηση των δεδομένων.

Η ουσιαστική **στατιστική ανάλυση** στο πλαίσιο της Διπλωματικής Εργασίας έγινε με το λογισμικό R στο οποίο αναπτύχθηκαν συνολικά έξι μαθηματικά πρότυπα, τα οποία έχουν όλα την ίδια εξαρτημένη μεταβλητή: τον δεκαδικό λογάριθμο του αριθμού των θανάτων σε οδικά ατυχήματα ανά τετραγωνικό χιλιόμετρο. Συγκεκριμένα, αναπτύχθηκαν τρία κλασικά μοντέλα λογαριθμοκανονικής παλινδρόμησης – ένα για το σύνολο των Δήμων, και ένα για κάθε μία από τις συστάδες Δήμων όπως αυτές προέκυψαν από την ομαδοποίηση.

Επιπλέον, αναπτύχθηκαν τρία μοντέλα **γεωγραφικά σταθμισμένης παλινδρόμησης** που λαμβάνουν υπόψη την χωρική γειτονία των Δήμων για την ανάλυσή τους. Επισημαίνεται ότι τα έξι παραπάνω μοντέλα προέκυψαν μετά από μεγάλο αριθμό δοκιμών για διάφορους συνδυασμούς των ανεξάρτητων μεταβλητών.

Αναφορικά με την **σχετική επιρροή** των ανεξάρτητων μεταβλητών στην εξαρτημένη, οι υπολογισμοί έγιναν μόνο για τα τρία κλασικά μοντέλα παλινδρόμησης. Στα πρότυπα αυτά, η σχετική επιρροή προσδιορίζεται μέσω του μεγέθους της ελαστικότητας, και οδηγεί σε μια καλύτερη αντίληψη της σχέσης μεταξύ του αριθμού θανάτων σε οδικά ατυχήματα ανά τετραγωνικό χιλιόμετρο, και των διάφορων ανεξάρτητων μεταβλητών.

Παρακάτω φαίνεται ο **συγκεντρωτικός πίνακας** των τελικών μοντέλων της Διπλωματικής Εργασίας.

Πρότυπα Λογαριθμοκανονικής Παλινδρόμησης															
Ανεξάρτητες μεταβλητές	Σύνολο Δήμων					Συστάδα 1					Συστάδα 2				
	R ²	AIC	Estimate	t	ei*	R ²	AIC	Estimate	t	ei*	R ²	AIC	Estimate	t	ei*
Ποσοστό απόφοιτων πανεπιστημίου	0.529	524.403	-0.019	-2.806	1.00	0.811	59.663	0.039	3.981	1.00	0.396	396.276	-0.021	-2.404	2.89
Ποσοστό γυναικών			-0.065	-2.709	-1.91			-0.239	-4.488	-3.69			-0.017	-0.685	1.00
Ποσοστό εργαζόμενων			-0.057	-6.903	-2.26			-0.087	-4.448	-1.78			-0.021	-2.152	1.92
Αριθμός νοικοκυριών με ιδιοκτησία ΙΧ			0.00002	6.124	10.83			0.0000093	3.877	7.50			0.000065	4.442	57.54
Ακαθάριστο Εθνικό Προϊόν			-0.00008	-6.955	-11.13			-0.00011	-7.563	-14.09			-0.000066	-3.917	38.32
Ποσοστό ηλικιωμένων							0.04	1.925	1.09			0.036	4.651	-2.84	

Πρότυπα Χωρικής Παλινδρόμησης GWR												
Ανεξάρτητες μεταβλητές	Σύνολο Δήμων				Συστάδα 1				Συστάδα 2			
	R ²	AIC	Estimate (Median)	Εύρος ζώνης	R ²	AIC	Εύρος ζώνης	Estimate (Median)	R ²	AIC	Εύρος ζώνης	Estimate (Median)
Ποσοστό απόφοιτων πανεπιστημίου	0.681	381.628	0.003	115	0.841	34.461	66	0.033	0.537	311.699	126	0.008
Ποσοστό γυναικών			-0.081					-0.189				-0.039
Ποσοστό εργαζόμενων			-0.053					-0.052				-0.04
Αριθμός νοικοκυριών με ιδιοκτησία ΙΧ			0.000024					0.0000071				0.000084
Ακαθάριστο Εθνικό Προϊόν			-0.000099					-0.00011				-0.001
Ποσοστό ηλικιωμένων							0.086					0.013

Πίνακας 6.1: Συγκεντρωτικός πίνακας αποτελεσμάτων μαθηματικών προτύπων

Σημειώνεται ότι όλες οι τιμές του συντελεστή R² αναφέρονται στον προσαρμοσμένο δείκτη (adjusted R²). Επίσης, οι εκτιμήσεις των συντελεστών στα χωρικά πρότυπα είναι οι διάμεσοι (median) του εκάστοτε εύρους τιμών.

6.2 Συνολικά συμπεράσματα

Σε όλα τα στάδια εκπόνησης της παρούσας Διπλωματικής Εργασίας προέκυψε μια σειρά από συμπεράσματα που σχετίζονται με τους στόχους της. Σε αυτό το σημείο παρατίθενται οι απαντήσεις στα διάφορα ερωτήματα που προέκυψαν, χρησιμοποιώντας τα αποτελέσματα των προηγούμενων κεφαλαίων. Συνοπτικά, τα **γενικά συμπεράσματα** της Διπλωματικής Εργασίας είναι:

- Υπάρχει μια σειρά από αρκετούς παράγοντες που μπορούν να επηρεάσουν τον αριθμό των θανάτων σε οδικά ατυχήματα. Ωστόσο, προέκυψε ότι η ιδιοκτησία ΙΧ και το Ακαθάριστο Εθνικό Προϊόν είναι πιο **καθοριστικοί παράγοντες** σε σχέση με δημογραφικά ή εκπαιδευτικά χαρακτηριστικά.
- Για το **σύνολο των Δήμων**, όπως και για τους **Δήμους με τον μεγαλύτερο πληθυσμό**, προέκυψε το αποτέλεσμα ότι η πιο κρίσιμη μεταβλητή είναι το **Ακαθάριστο Εθνικό Προϊόν**. Συγκεκριμένα, εάν αυξηθεί, παρατηρείται μείωση των θανάτων σε οδικά ατυχήματα. Η πιο πιθανή εξήγηση είναι ότι ένα υψηλότερο ΑΕΠ συνδέεται με καλύτερη ποιότητα ζωής, και επιτρέπει την ανάπτυξη πιο αποτελεσματικών μέτρων οδικής ασφάλειας.

3. Παρατηρήθηκε ότι το **ποσοστό απόφοιτων πανεπιστημίου** συσχετίζεται θετικά με την εξαρτημένη μεταβλητή στους Δήμους μεγαλύτερου πληθυσμού, ενώ συσχετίζεται αρνητικά στους περιφερειακούς και επαρχιακούς Δήμους, όπως και στο σύνολο των Δήμων. Αυτή η διαφορά μπορεί να εξηγηθεί ως εξής:

Γενικά, έγινε η παραδοχή ότι οι απόφοιτοι πανεπιστημίου βρίσκονται σε ηλικία άνω των 24 ετών και έχουν αναπτύξει πιο ώριμη οδική συμπεριφορά, κάτι που εξηγεί το αρνητικό πρόσημο του συντελεστή. Ωστόσο, στους μεγάλους αστικούς Δήμους οι μετακινήσεις είναι αυξημένες, κάτι που αυξάνει την έκθεση στον κίνδυνο, και κατά συνέπεια και τους θανάτους σε οδικά ατυχήματα. Επομένως, υπάρχει μια θετική συσχέτιση των απόφοιτων πανεπιστημίου με την εξαρτημένη μεταβλητή σε αυτούς τους Δήμους.

4. Για τους **περιφερειακούς Δήμους**, βρέθηκε ότι ο πιο καθοριστικός παράγοντας είναι ο **αριθμός νοικοκυριών με ιδιοκτησία ΙΧ**. Συγκεκριμένα, η αύξηση της εν λόγω μεταβλητής προκαλεί αύξηση των θανάτων σε οδικά ατυχήματα. Το αποτέλεσμα είναι λογικό καθώς η αυξημένη ιδιοκτησία ΙΧ, οδηγεί σε μεγαλύτερο στόλο οχημάτων και κατά συνέπεια μεγαλύτερη έκθεση στον κίνδυνο.
5. Παρατηρήθηκε ότι το **ποσοστό του γυναικείου πληθυσμού** γενικά δεν έχει μεγάλη επιρροή στον αριθμό των θανάτων σε οδικά ατυχήματα, ενώ στην περίπτωση των περιφερειακών Δήμων είναι η μεταβλητή με την χαμηλότερη επιρροή. Σε όλες τις περιπτώσεις, η συσχέτιση με την εξαρτημένη μεταβλητή είναι αρνητική, κάτι που είναι λογικό καθώς οι γυναίκες οδηγούν πιο υπεύθυνα σε σχέση με τους άνδρες.
6. Το **ποσοστό των εργαζομένων** είναι μια μεταβλητή με χαμηλή επιρροή στην εξαρτημένη, ενώ τόσο στο σύνολο των Δήμων, όσο και χωριστά στους αστικούς και περιφερειακούς Δήμους, η συσχέτιση είναι αρνητική. Η εξήγηση για αυτό το αποτέλεσμα είναι ότι η αύξηση των απασχολούμενων σε εργασία οδηγεί σε καλύτερο βιοτικό επίπεδο, και συνεπώς σε πιο ενημερωμένους και υπεύθυνους οδηγούς, κάτι που προκαλεί μείωση των θανάτων σε οδικά ατυχήματα.
7. Γενικά, φαίνεται ότι σε κάθε περίπτωση η χωρική παλινδρόμηση παράγει ένα πρότυπο με υψηλότερο R^2 και χαμηλότερο AIC από τους αντίστοιχους δείκτες της λογαριθμοκανονικής. Συνεπώς, η **εφαρμογή της χωρικής μεθόδου** στην ανάλυση των οδικών ατυχημάτων βελτίωσε την ακρίβεια των μαθηματικών μοντέλων.
8. Ανάμεσα στο πρότυπο λογαριθμοκανονικής παλινδρόμησης για τους Δήμους με τον μεγαλύτερο πληθυσμό και σε αυτό για τους περιφερειακούς Δήμους, παρατηρείται μια σημαντική απόκλιση στον **συντελεστή R^2** , δηλαδή στην προσαρμογή του προτύπου στα δεδομένα. Μια πιθανή εξήγηση είναι η μεγάλη διαφορά στο πλήθος των Δήμων μεταξύ των δύο περιπτώσεων.

Οι πιο αστικοποιημένοι Δήμοι είναι λιγότεροι, οπότε το μοντέλο προσαρμόζεται πιο εύκολα στα δεδομένα, οδηγώντας σε πολύ ικανοποιητικό δείκτη R^2 . Αντιθέτως, η συστάδα των επαρχιακών και περιφερειακών Δήμων περιέχει πολύ

περισσότερους, που πιθανώς παρουσιάζουν αποκλίσεις στα δεδομένα, κάτι που οδηγεί σε πιο δύσκολη προσαρμογή του προτύπου – δηλαδή χαμηλότερο δείκτη R^2 .

9. Οι στατιστικές μέθοδοι που χρησιμοποιήθηκαν στην ανάλυση είναι η **λογαριθμοκανονική παλινδρόμηση** και η **γεωγραφικά σταθμισμένη παλινδρόμηση**, οι οποίες κρίθηκαν κατάλληλες για την παρούσα Διπλωματική Εργασία. Τα μαθηματικά μοντέλα που προέκυψαν είναι αξιόπιστα και περιγράφουν ικανοποιητικά την εξάρτηση των θανάτων σε οδικά ατυχήματα από τις ανεξάρτητες μεταβλητές.

6.3 Προτάσεις για την μείωση των θανάτων σε οδικά ατυχήματα

Με βάση το σύνολο των αποτελεσμάτων και συμπερασμάτων της Διπλωματικής Εργασίας, παρατίθενται ορισμένες **προτάσεις** που πιθανώς να συμβάλλουν στην μείωση των θανάτων σε οδικά ατυχήματα, και στην γενικότερη **βελτίωση του επιπέδου οδικής ασφάλειας** της Ελλάδας.

1. Στους πιο αστικούς Δήμους της χώρας, προτείνεται η **λήψη μιας σειράς μέτρων διαχείρισης της κυκλοφορίας**, με σκοπό την αντιμετώπιση των αυξημένων μετακινήσεων με αυτοκίνητα ΙΧ, και συνεπώς την μείωση της έκθεσης των μετακινούμενων στον κίνδυνο ατυχήματος. Στο πλαίσιο μιας γενικότερης αλλαγής στην συγκοινωνιακή πολιτική, θα πρέπει να δοθεί έμφαση στην ανάπτυξη των μέσων μαζικής μεταφοράς που θα υποκαταστήσουν σημαντικό μέρος των μετακινήσεων με ΙΧ.
2. Το ζήτημα της οδικής ασφάλειας χρειάζεται να αντιμετωπιστεί πιο ολοκληρωμένα, με **εκστρατείες ενημέρωσης** των πολιτών σχετικά με τα οδικά ατυχήματα και τα αίτιά τους. Πρέπει, επίσης, να ενταχθεί ένα **εκπαιδευτικό πρόγραμμα** οδικής συμπεριφοράς ως υποχρεωτικό μάθημα στα σχολεία, με σκοπό την ευαισθητοποίηση των μελλοντικών οδηγών από μικρή ηλικία.
3. Σημαντική είναι, επίσης, η καθολική **εφαρμογή μεθόδων χωρικής ανάλυσης** στην μελέτη των οδικών ατυχημάτων. Όπως συμπεραίνεται από την παρούσα Διπλωματική Εργασία, οι στατιστικές μέθοδοι που λαμβάνουν υπόψη τον παράγοντα του χώρου στην ανάλυση παράγουν ακριβέστερα και πιο αξιόπιστα αποτελέσματα. Επομένως, η χρήση χωρικών μεθόδων στην οδική ασφάλεια θα συμβάλλει στην δημιουργία βελτιωμένων μοντέλων ανάλυσης των ατυχημάτων, και κατά συνέπεια στην ανάπτυξη καλύτερων μέτρων πρόληψής τους.

6.4 Προτάσεις για περαιτέρω έρευνα

Χρησιμοποιώντας τα αποτελέσματα της παρούσας Διπλωματικής Εργασίας, θα μπορούσε να γίνει επέκταση και **περαιτέρω διερεύνηση** του αντικειμένου της, όπως παρουσιάζεται στις ακόλουθες προτάσεις:

1. Ενδιαφέρον θα παρουσίαζε η ανάλυση με περισσότερες και διαφορετικές ανεξάρτητες μεταβλητές, ώστε να καθοριστούν και **άλλοι σημαντικοί παράγοντες** που επηρεάζουν τα οδικά ατυχήματα. Τέτοιες μεταβλητές θα μπορούσαν να είναι το ύψος βροχής στον κάθε Δήμο, η κατάσταση των οδοστρωμάτων, και χαρακτηριστικά των οχημάτων.
2. Τα δεδομένα θα μπορούσαν να ομαδοποιηθούν με τη χρήση άλλου **αλγόριθμου ομαδοποίησης** (πχ, K-means clustering), ώστε να προκύψουν διαφορετικές συστάδες Δήμων, και να εξεταστεί ο βαθμός που η χωρική μέθοδος εξακολουθεί να παράγει βελτιωμένα αποτελέσματα σε σχέση με την κλασική.
3. Στο πλαίσιο μιας πιο ολοκληρωμένης μελέτης των οδικών ατυχημάτων, θα ήταν ενδιαφέρουσα η **σύγκριση των αποτελεσμάτων** της παρούσας Διπλωματικής Εργασίας με άλλες χωρικές αναλύσεις ανά την Ευρώπη.

Κεφάλαιο 7: Βιβλιογραφικές Αναφορές

1. European Commission, Directorate-General for Mobility and Transport, October 2023, https://transport.ec.europa.eu/news-events/news/road-safety-20640-people-died-road-crash-last-year-progress-remains-too-slow-2023-10-19_en
2. European Commission, Road safety statistics 2022 in more detail, https://transport.ec.europa.eu/background/road-safety-statistics-2022-more-detail_en
3. Ελληνική Στατιστική Αρχή, <https://www.statistics.gr/>
4. National Technical University of Athens, Road Safety Observatory, <https://www.nrso.ntua.gr/>
5. Yannis, G., Papadimitriou, E., & Folla, K. (2014). Effect of GDP changes on road traffic fatalities. *Safety science*, 63, 42-49.
6. Nikolaou, D., Folla, K., & Yannis, G. (2021). Impact of socioeconomic and transport indicators on road safety during the crisis period in Europe. *International journal of injury control and safety promotion*, 28(4), 479-485.
7. Ziakopoulos, A., & Yannis, G. (2020). A review of spatial approaches in road safety. *Accident Analysis & Prevention*, 135, 105323.
8. Aguero-Valverde, J., & Jovanis, P. P. (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis & Prevention*, 38(3), 618-625.
9. Bassani, M., Rossetti, L., & Catani, L. (2020). Spatial analysis of road crashes involving vulnerable road users in support of road safety management strategies. *Transportation research procedia*, 45, 394-401.
10. Noland, R. B., & Quddus, M. A. (2005). Congestion and safety: A spatial analysis of London. *Transportation Research Part A: Policy and Practice*, 39(7-9), 737-754.
11. Gudes, O., Varhol, R., Sun, Q. C., & Meuleners, L. (2017). Investigating articulated heavy-vehicle crashes in Western Australia using a spatial approach. *Accident Analysis & Prevention*, 106, 243-253.
12. Plug, C., Xia, J. C., & Caulfield, C. (2011). Spatial and temporal visualisation techniques for crash analysis. *Accident Analysis & Prevention*, 43(6), 1937-1946.
13. Zeng, Q., & Huang, H. (2014). Bayesian spatial joint modeling of traffic crashes on an urban road network. *Accident Analysis & Prevention*, 67, 105-112.

14. Ziakopoulos, A. (2021). Spatial analysis of harsh driving behavior events in urban networks using high-resolution smartphone and geometric data. *Accident Analysis & Prevention*, 157, 106189.
15. Spoerri, A., Egger, M., & von Elm, E. (2011). Mortality from road traffic accidents in Switzerland: longitudinal and spatial analyses. *Accident Analysis & Prevention*, 43(1), 40-48.
16. Agüero-Valverde, J. (2013). Multivariate spatial models of excess crash frequency at area level: Case of Costa Rica. *Accident Analysis & Prevention*, 59, 365-373.
17. Geographically Weighted Regression, <https://cran.r-project.org/web/packages/spgwr/vignettes/GWR.html>
18. Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.